



**XII INTERNATIONAL
CONFERENCE AND WORKSHOPS
“INFORMATION TECHNOLOGY
AND IMPLEMENTATION”
(SATELLITE)
(IT&I(S)-2025)**

NOVEMBER 21, 2025



TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV
(FACULTY OF INFORMATION TECHNOLOGY,
FACULTY OF COMPUTER SCIENCE AND CYBERNETICS)
NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
“IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE”
VIKTOR GLUSHKOV INSTITUTE OF CYBERNETICS OF THE NAS OF UKRAINE
INSTITUTE FOR INFORMATION RECORDING OF THE NAS OF UKRAINE
INSTITUTE OF SOFTWARE SYSTEMS OF THE NAS OF UKRAINE
THE COUNCIL OF YOUNG SCIENTISTS OF THE FACULTY OF COMPUTER SCIENCE AND
CYBERNETICS AND THE FACULTY OF INFORMATION TECHNOLOGY OF
TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV
PUBLIC ASSOCIATION «COGNITIVE RESEARCH»

XI INTERNATIONAL CONFERENCE

Information Technology and Implementation (Satellite)

21 November, 2025

Conference Proceedings

Kyiv

2025

UDC 004.9
I-68

Volume editor: Vitaliy Snytyuk, Dr.Sc., Prof.

Program Committee: Aldrich Chris, Andreas Pester, Frederic Mallet, Iurii Krak, Nickolas Cogan, Oleksandr Marchenko, Sándor Bozóki, Tom Needham, Vitaliy Tsyganok, Vladimir Vovk

Organizing Committee: Anatoly Anisimov, Vitaliy Snytyuk, Oleksii Bychkov, Oleh Ilarionov, Yurii Kravchenko, Iurii Krak, Oleksandr Marchenko, Viktor Morozov, Taras Panchenko, Valentyna Pleskach, Yurii Samokhvalov, Oleksii Tkachenko, Serhii Toliupa.

Conference Secretary: Hryhorii Hnatiienko.

I-68 Information Technology and Implementation (Satellite): Conference Proceedings, November 21, 2025, Kyiv, Ukraine / Ministry of Education and Science of Ukraine, Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publishing House «Caravela», 2025. 333 p.

ISBN 978-960-801-928-7

This book includes abstracts of the 12th International Conference "Information Technology and Implementation (Satellite) – 2025". Philosophical, theoretical and applied aspects which describe the results, problems and prospects of the creation and use of intelligent computing methods and creating of information systems and technology on their basis are reviewing.

Main tracks of the conference are: Artificial Intelligence Technologies and Data Science; Cyberspace Protection Technologies; E-government, E-commerce and E-learning Technologies; Network, Internet and Smart Technologies; Software Engineering, Data Structures and Computing.

До цієї збірки увійшли тези доповідей 12-ї Міжнародної конференції «Інформаційні технології та впровадження (сателітної) – 2025». Розглядаються філософські, теоретичні та прикладні аспекти, які описують результати, проблеми та перспективи створення і використання інтелектуальних обчислювальних методів та створення на їх основі інформаційних систем і технологій.

Основні напрямки конференції: Технології штучного інтелекту та Наука про дані, Технології захисту кіберпростору, Електронне урядування, Електронна комерція та Технології електронного навчання, Мережеві та Інтернет-технології, Розумні технології, Розробка програмного забезпечення, Структури даних та Обчислення.

UDC 004.9

ISBN 978-960-801-928-7

© Authors of abstracts, 2025

CONTENTS

<i>TRACK 1: ARTIFICIAL INTELLIGENCE TECHNOLOGIES AND DATA SCIENCE</i>	12
<i>Andrii Askerov, Lybov Borkovska. AI IN SPORT: AN ARTIFICIAL PERSONAL TRAINER FOR DYNAMIC CYCLICAL PERIODIZATION BASED ON MACHINE LEARNING</i>	13
<i>Nikita Avdeev, Iryna Yurchuk. USING HYBRID NEURAL-NOISE METHODS TO DETECT ARTIFICIALLY GENERATED AND MANIPULATED IMAGES</i>	16
<i>Yehor Bashkevych, Oleh Zaritskyi. AUTOMATED SENTIMENT AND EMOTIONAL ANALYSIS OF GLOBAL NEWS ABOUT UKRAINE: DATA SCIENCE METHODS AND RESEARCH PERSPECTIVES</i>	19
<i>Stepan Bilan, Dmytro Bereznyi. OPTIMIZED KEYPOINT MATCHING IN UAV AGRICULTURAL IMAGERY WITH SIFT AND ORB</i>	22
<i>Oleh Bohutskyi, Hanna Krasovska. EVALUATING FREQUENCY ANALYSIS FOR LARGE LANGUAGE MODEL QUERIES</i>	24
<i>Kateryna Bondar, Iryna Domanetska. ANALYSIS OF DATASETS SUITABLE FOR USE BY AI MODELS IN DIABETES MELLITUS SCREENING AND EARLY DETECTION</i>	27
<i>Yurii Cherevach, Andrii Khlevnyi. AUTOMATED DETECTION OF PROPAGANDA TECHNIQUES IN MULTILINGUAL CONFLICT NARRATIVES</i>	30
<i>Vladislava Chernorai, Oleksandra Kovalenko, Vladimir Semenov. REGRESSION-TYPE PROBLEMS AND ALGORITHMS FOR VARIATIONAL INEQUALITIES</i>	33
<i>Anastasiia Chubukova, Oleh Zaritskyi. MODELING THE IMPACT OF EMOTIONALITY IN DIGITAL COMMUNICATIONS ON CUSTOMER BEHAVIOR</i>	37
<i>Ivan Demenko, Mykola Fesenko, Iryna Kutova. MULTI-AGENT SIMULATION ENVIRONMENT FOR INTELLIGENT OPTIMIZATION OF AUTOMATED TESTING PROCESSES</i>	39
<i>Pavlo Derevianko, Ihor Miroshnychenko. Generative AI IN THE STUDENT ENVIRONMENT: AN ANALYTICAL REVIEW OF EXPECTATIONS, RISKS, AND PROSPECTS</i>	42
<i>Maksym Donchenko, Andriy Klevny. A COMBINED METHOD FOR DETECTING AND ELIMINATING DUPLICATE PUBLICATIONS WHEN AGGREGATING PUBLICATION DATA FROM SCOPUS, WOS, AND ORCID DATABASES</i>	45

<i>Denis Drachuk, Ihor Miroshnychenko.</i> OBJECT DETECTION IN ADVERSE WEATHER CONDITIONS: CURRENT CHALLENGES, METHODS, AND FUTURE DIRECTION	47
<i>Illia Harbar, Kseniia Dukhnovska.</i> A HYBRID ONLINE-OFFLINE SYSTEM FOR ROBUST SPEECH RECOGNITION AND TRANSLATION OF KOREAN AUDIO	50
<i>Dariia Herasymchuk, Iryna Domanetska.</i> HYBRID APPROACHES TO COVID-19 EPIDEMIC FORECASTING BASED ON MECHANISTIC AND NEURAL NETWORK MODELS	54
<i>Vladyslav Hnatiienko.</i> CONCEPT OF A SPATIO-TEMPORAL DATA RECOVERY METHOD FOR SATELLITE IMAGERY DISTORTED BY CLOUD COVER	56
<i>Veronika Horbach, Victoria Volkogon.</i> APPLICATION OF A CUSTOM-TUNED RANKING ENGINE FOR MULTI-FACTOR USER PROFILE ANALYSIS	59
<i>Demyd Hredsov, Olena Reshetnyak.</i> ARTIFICIAL INTELLIGENCE AND DATA SCIENCE IN MANAGEMENT AND PUBLIC ADMINISTRATION: SOFTWARE ENGINEERING APPROACHES FOR DECISION-MAKING AND DIGITAL TRANSFORMATION	61
<i>Inna Hrynishak, Hanna Krasovska.</i> INTELLIGENT MULTI-AGENT SYSTEM ARCHITECTURE FOR DEEPFAKE DETECTION	64
<i>Vladyslav Hyzha, Dmitry Gololobov.</i> USING AI AGENTS TO ENHANCE USER EXPERIENCE IN PUBLIC TRANSPORT ROUTE PLANNING	68
<i>Andrii Ivanenko, Andriy Shatyrko.</i> METHODS OF TARGET IDENTIFICATION AND LOCALIZATION BY SOUND SIGNAL	70
<i>Yaroslav Khrolenko.</i> SEMANTIC SIMILARITY, CO-OCCURRENCE, AND SEMANTIC RELATEDNESS AS INDICATORS OF THEMATIC AFFINITY	73
<i>Victoriia Kodovbetska, Julia Minaeva.</i> RECOMMENDATION SYSTEM FOR FOOD ESTABLISHMENT SELECTION IN LVIV	76
<i>Volodymyr Kochuma.</i> IT ENSURING MODELS FOR INNOVATIVE DEVELOPMENT OF CONSTRUCTION (DEVELOPMENT) COMPANIES	78
<i>Serhii Korotenko, Roman Yarovy.</i> CHALLENGES OF ADAPTING MACHINE LEARNING MODELS TO STRUCTURAL DRIFT IN BUSINESS PROCESSES	82
<i>Volodymyr Kovalchuk, Lyubov Borkovska.</i> DIGITALIZATION OF INTERNAL BUSINESS PROCESSES USING INTELLIGENT CHATBOTS: DEVELOPMENT METHODOLOGY AND SOCIAL IMPACT	85

<i>Georgii Kozhevnikov, Serhii Mytrofanov, Hanna Zavolodko. AI: RESTORING AGRICULTURAL PRODUCTION AMID WAR AND CLIMATE CHALLENGES</i>	87
<i>Illia Krytskyi, Iryna Yurchuk. APPLICATION OF DEEP LEARNING FOR AUDIO OBJECT CLASSIFICATION AND METADATA GENERATION</i>	91
<i>Dmytro Masiuk. CHALLENGES OF DATA ANONYMIZATION IN EXPERIMENTAL RESEARCH</i>	93
<i>Oleksandr Mazurets, Oleksandr Ovcharuk, Roman Vit. PRACTICE INVESTIGATION OF NEURAL NETWORK DETECTING OF CONSTRUCTION WASTE BY PHOTOS</i>	95
<i>Taras Melnyk, Yevhen Tatarynov. APPLICATION OF LARGE LANGUAGE MODELS FOR ANALYZING THE COMPLIANCE OF PROGRAM CODE WITH SOLID PRINCIPLES</i>	97
<i>Viktoriia Moiseienko, Olena Fedusenko. FORECASTING FOOD PRICES USING ARIMAX MODEL</i>	100
<i>Oksana Moisiuk, Anna Kolomiiets. DEVELOPMENT OF AN AI-POWERED APPLICATION FOR SMART DINING DISCOVERY AND PERSONALIZED TABLE RESERVATION</i>	102
<i>Denys Muzyka, Stanislav Marchenko, Andrii Yarmilko. POLYMODEL VIDEO STREAM COMPRESSION IN COMPUTER VISION SYSTEMS</i>	104
<i>Oleksii Paharskyi, Georgy Gaina. LARGE LANGUAGE MODEL MULTI-AGENT SYSTEM FOR PROGRAMMING</i>	106
<i>Andrii Pavlov. A THEORY-BASED DECISION FRAMEWORK FOR ENCODING CATEGORICAL VARIABLES IN REGRESSION AND CLASSIFICATION PROBLEMS</i>	109
<i>Denys Popov, Anna Kolomiiets. DEVELOPMENT OF AN AI-DRIVEN APPLICATION FOR NUTRITION MONITORING AND HEALTHY HABIT FORMATION</i>	113
<i>Vladyslav Pylypenko, Vladyslava Skidan, Antonina Volivach. PREDICTING STUDENTS' SUCCESS AND GRADATION OF B USING MACHINE LEARNING</i>	115
<i>Oleksandr Redenskyi, Kseniia Dukhnovska. HEURISTIC OPTIMIZATIONS FOR THE TRANSPORTATION PROBLEM IN URBAN CONDITIONS</i>	117
<i>Maksym Rudenko, Mykyta Myronenko. INFORMATION-EXTREME MACHINE LEARNING OF AN AUTONOMOUS ON-BOARD SYSTEM FOR MILITARY EQUIPMENT RECOGNITION</i>	119
<i>Oleksii Skukis. A BRIEF ANALYSIS OF TOPOLOGIES IN PARTICLE SWARM OPTIMIZATION ALGORITHMS</i>	121

<i>Maksym Steblivskyi, Ihor Miroshnychenko. UNCOVERING SANCTIONS EVASION VIA THIRD COUNTRIES: A NETWORK ANALYSIS OF RUSSIAN IMPORT DATA</i>	123
<i>Maksym Suprunenko, Ihor Myroshnychenko. PHOTOREALISTIC SYNTHETIC DATA GENERATION FOR ROBUST UKRAINIAN FISCAL RECEIPT OCR</i>	125
<i>Yehor Symonov, Sergii Yershov. THE METHODOLOGY FOR CONSTRUCTING SYNTHETIC ANOMALIES FOR VALIDATING ML PROCESSES</i>	127
<i>Danylo Taranov, Oleh Moroz. ARTIFICIAL INTELLIGENCE AS A DIGITAL TWIN FOR OPTIMIZING PROJECT MANAGEMENT</i>	132
<i>Ivan Vakulov, Georgy Gaina. INTELLIGENT TOOLS FOR DEVELOPER SUPPORT: PROSPECTS FOR THE DEVELOPMENT OF IDE EXTENSIONS BASED ON LLM</i>	135
<i>Mykhailo Voitovych, Oleksii Bychkov. HYBRID AUTOMATA-GUIDED MULTI-AGENT REINFORCEMENT LEARNING ARCHITECTURE FOR MULTI-PHASE UAV SWARM MISSIONS</i>	137
<i>Eduard Zakabluk, Oleksandr Lapin. INFORMATION-EXTREME MACHINE LEARNING OF AN ONBOARD GEOINFORMATION SYSTEM FOR SEMANTIC SEGMENTATION OF AGRICULTURAL LAND</i>	139
<i>Olha Zalutska, Oleksandr Mazurets, Maryna Molchanova. EFFICIENCY ANALYSIS OF WRECKING WASTE CLASSIFICATION USING NEURAL NETWORK</i>	141
<i>Mariia Zheludkova, Oleh Zaritskyi. SOCIAL AND ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE</i>	143
<i>Vadym Ziuziun, Andrii Utvenko. APPLICATION OF REINFORCEMENT LEARNING TO OPTIMIZE INPUT PARAMETERS IN ORDER TO ACHIEVE PREDEFINED PERFORMANCE METRICS DURING BENCHMARKING OF A HARDWARE–SOFTWARE SYSTEM</i>	145
<i>Oksana Zolotukhina, Oleg Ilarionov. FEATURES OF BUILDING A RECOMMENDATION SYSTEM FOR ELECTIVE COURSES</i>	147
<i>I. Yarmolai. CONSTRUCTION OF A MATHEMATICAL MODEL OF SEISMOACOUSTIC MONITORING FOR REMOTE EXPLORATION</i>	151
<i>Iryna Yurchuk, Stanislav Golouh, Vladyslav Shulga, Ihor Pishta. ENHANCING THE RELIABILITY OF UAV DETECTION THROUGH MULTIMODAL SENSOR DATA FUSION</i>	155

TRACK 2: CYBERSPACE PROTECTION TECHNOLOGIES	158
<i>Stepan Bilan</i> . INTERNAL SECURITY OF ROBOTIC SWARM BEHAVIOR DURING MESSAGE EXCHANGE	159
<i>Ivan Bilokon, Yaroslava Fylenko, Sofiia Denichenko</i> . AUTOMATED OSINT DATA COLLECTION AND ANALYSIS FOR CYBER RISK ASSESSMENT	161
<i>Nykyta Biriukov, Yurii Shcheblanin</i> . FRACTAL-BASED CRYPTOGRAPHY WITH NORMALIZATION	163
<i>Serhii Buchyk, Vladyslav Lytvynenko</i> . KUBERNETES NETWORK POLICIES AS THE ZERO TRUST BASELINE: A CASE FOR MANDATORY DEFAULT DENY SEGMENTATION	165
<i>Serhii Buchyk, Vitalii Piatyhor</i> . IMPROVED SOCIAL MEDIA BOT DETECTION SYSTEM BASED ON PUBLICLY AVAILABLE DATA	167
<i>Oleksii Chalyi, Serhii Toliupa</i> . INFORMATION SECURITY AUDIT AUTOMATION USING ARTIFICIAL INTELLIGENCE: CHALLENGES AND OPPORTUNITIES	171
<i>Dmytro Dzhendzhero, Volodymyr Nakonechnyi</i> . DETECTION OF FALSE INFORMATION INDICATORS IN HYBRID WARFARE	173
<i>Tetiana Ivanova, Serhii Buchyk</i> . MACHINE LEARNING AS A TOOL FOR IMPROVING INTRUSION DETECTION SYSTEMS IN ENCRYPTED TRAFFIC	175
<i>Mikhail Kondratiuk, Andriy Kondratiuk, Mykola Brailovskyi</i> . DETECTION OF ANOMALOUS NETWORK ACTIVITY USING ENSEMBLE ALGORITHMS	177
<i>Yevhenii Korniienko, Yurii Scheblanin</i> . MODEL FOR DETECTING AND PREVENTING PHISHING ATTACKS BASED ON MACHINE LEARNING	180
<i>Liubov Kubiavka, Viktoriia Kodovbetska</i> . SOCIAL MEDIA INFORMATION ANALYTICS FOR DISINFORMATION DETECTION	182
<i>Andrii Kulko, Serhii Toliupa</i> . USE OF SUPPORT VECTORS IN INTRUSION DETECTION SYSTEMS	184
<i>Larysa Myrutenko, Ivan Mazur</i> . METHODS OF RESPONDING TO INCIDENTS IN INFORMATION SYSTEMS	188
<i>Volodymyr Nakonechnyi, Ilya Kachur</i> . METHODOLOGY FOR EVALUATING THE RESILIENCE OF ANTIVIRUS DETECTION TO POLYMORPHIC MALWARE TRANSFORMATIONS	191
<i>Volodymyr Nakonechnyi, Valeriia Kyzyma</i> . OSINT-ANALYSIS OF THREAT ACTOR INFRASTRUCTURE FOR FORECASTING AND COUNTERING PHISHING CAMPAIGNS	193

<i>Volodymyr Nakonechnyi, Vladyslav Lutsenko. ENHANCING DATA SECURITY IN DISTRIBUTED SYSTEMS THROUGH BLOCKCHAIN TECHNOLOGY</i>	195
<i>Volodymyr Nakonechnyi, David Mashkovich. METHODS FOR ENSURING CONTINUITY OF INTERNET OF THINGS SYSTEMS UNDER CYBER INCIDENTS</i>	197
<i>Volodymyr Nakonechnyi, Ilona Miroshnichenko. METHOD FOR EVALUATING THE EFFECTIVENESS OF DECEPTION TECHNOLOGIES FOR CYBERATTACK DETECTION</i>	199
<i>Denys Odnoochko, Anna Iliencko. CRYPTO-STEGANOGRAPHIC PROTECTION OF INFORMATION RESOURCES</i>	201
<i>Roman Ohiievych. HOW RTCT PROTOCOL IMPACTS THE COST OF SYBIL ATTACK</i>	203
<i>Vladyslav Ostrovshchuk, Iulia Khlevna. DETECTING ANOMALOUS STREAMING ACTIVITY IN SPOTIFY: AN INFORMATION ECOLOGY APPROACH TO DATA INTEGRITY</i>	205
<i>Ivan Parkhomenko, Lev Bakhmut. HIDING METADATA IN MULTIMEDIA USING MACHINE LEARNING METHODS</i>	207
<i>Ivan Parkhomenko, Ihor Kamniev. MITIGATING MISCONFIGURATIONS IN INFRASTRUCTURE-AS-CODE: TOWARDS RISK-AWARE DETECTION</i>	209
<i>Ivan Parkhomenko, Yelyzaveta Kazora. ENHANCING INFORMATION SECURITY THROUGH AUTOMATION</i>	211
<i>Ivan Parkhomenko, Yurii Sokyran. METHODS FOR ASSESSING SUPPLY CHAIN RISKS IN OPEN-SOURCE SOFTWARE</i>	213
<i>Yevhenii Ponomarenko, Oleksandr Laptiev. WHAT PNG 3 MEANS FOR STEGANOGRAPHY</i>	215
<i>Oleksandra Pryshchepa, Yurii Babenko. STEGANOGRAPHIC EMBEDDING OF DIGITAL SIGNATURES IN PDF DOCUMENTS</i>	217
<i>Mykhailo Savonik, Ivan Parkhomenko. EXPLAINABLE REAL-TIME ANOMALY DETECTION FOR EVASIVE ATTACKS IN AWS CLOUD ENVIRONMENTS</i>	221
<i>Kyryl Shamota, Yurii Shcheblanin. A METHOD OF CYBERATTACK DETECTION IN NETWORK TRAFFIC BASED ON COMBINED STATISTICAL ANALYSIS</i>	223
<i>Ihor Shylenko. APPROACHES TO MITIGATING CYBERSECURITY THREATS FROM CRYPTOCURRENCIES IN NATIONAL FINANCIAL SYSTEMS</i>	225

Dmytro Tsymbal, Anna Kolomiets. DEVELOPMENT OF A PROJECT FOR CREATING A SECURE WEB-PLATFORM FOR DATA MANAGEMENT IN OPERATIONAL-SEARCH ACTIVITIES 227

Illia Vavilov, Yurii Shcheblanin. A SECURE SYSTEM MODEL FOR SYNCHRONIZING SECRET KEYS IN MOBILE TOTP AUTHENTICATORS 229

TRACK 3: E-GOVERNMENT, E-COMMERCE, AND E-LEARNING TECHNOLOGIES 231

Maksim Gonchar, Andrii Khlevnyi. DEVELOPMENT OF AN INTERNET PORTAL FOR TRADING EDUCATION WITH BLOCKCHAIN TECHNOLOGY INTEGRATION AND TOKENIZED REWARD SYSTEM 232

Larysa Liashenko, Tetiana Andriichuk, Olena Stepanenko, Oksana Krasnenko. EMBEDDING RESEARCH BASED LEARNING in VIRTUAL INTERNATIONAL EDUCATION: ENHANCING STUDENTS' ACADEMIC ENGAGEMENT and INTERCULTURAL INTELLIGENCE 234

Dmytro Perepelytsia, Yurii Babenko. IMPLEMENTATION OF A SYSTEM FOR ELECTRONIC DIGITAL SIGNATURE VERIFICATION AND VISUALIZATION 236

Ilya Tauyrskii, Tetiana Zatonatska. MODELING THE DIGITAL INFRASTRUCTURE OF PUBLIC PROCUREMENT AS A TOOL FOR ENHANCING TRANSPARENCY AND EFFICIENCY OF PUBLIC FINANCE MANAGEMENT 238

TRACK 4: NETWORK, INTERNET, AND SMART TECHNOLOGIES 240

Hennadii Bulava, Kseniia Dukhnovska. SOCIAL NETWORK WITH GEOLOCATION SEARCH INTEGRATION: AN ALGORITHMIC APPROACH TO SERVICE RANKING 241

Danylo Chumachenko, Vira Liubchenko. A SEVEN-STEP METHOD FOR SELECTING ARCHITECTURAL PATTERNS FOR IOT 245

Yaroslav Hladkyi, Volodymir Druzhynin, Myroslava Gladka. IMMERSIVE VR/AR-ENVIRONMENTS FOR NEUROREHABILITATION AND MOTOR FUNCTION RECOVERY 249

Kyrylo Karmazin, Julia Pisarenko. OVERVIEW OF INTELLIGENT METHODS IN ADAPTIVE VIDEO STREAMING SYSTEMS 253

Yuliia Mavrodi, Tetiana Latysheva. PROJECT OF WEB SERVICE DEVELOPMENT FOR MONITORING POWER OUTAGES IN UKRAINE 255

Danylo Skydan, Yurii Shcheblanin. ANALYSIS OF PERFORMANCE AND STABILITY OF COMBINED METHODS IN WI-FI PROTOCOLS 258

Ihor Surovtsev, Valentyna Galimova, Yevheniya Savchenko-Syniakova, Volodymyr Liakhov, Andriy Khalevych, Yaroslav Antoniuk. SYSTEMS FOR RAPID DETERMINATION OF THE ECOLOGICAL STATE OF WATER RESOURCES 260

Orest Vovchak, Zenoviy Veres. COMPARATIVE ANALYSIS OF DATA STORAGE METHODS FOR IOT APPLICATIONS: ADVANTAGES AND LIMITATIONS 262

TRACK 5: SOFTWARE ENGINEERING, DATA STRUCTURES, AND COMPUTING 264

Maksym Demydov, Yurii Babenko. HIERARCHICAL ROLE MODEL IN CRM 265

Georgy Gaina, Hrygorii Hnatiienko, Iryna Domanetska, Oleh Ilarionov, Hanna Krasovska. PROFUNCTOR SEMANTICS IN CATEGORICAL LOGIC FOR THE VERIFICATION OF QUERIES AND DATA INTEGRITY IN RELATIONAL DATABASES 268

Maryna Hladych, Iryna Domanetska. ANALYSIS OF EXISTING APPROACHES TO TASK DURATION PREDICTION 271

Hrygorii Hnatiienko, Oleh Ilarionov, Hanna Krasovska. MULTI-ATTRIBUTE DETERMINATION OF MEASURES OF SIMILARITY BETWEEN BRANCHES OF KNOWLEDGE AND SPECIALTIES IN THE TASK OF DOCUMENTARY SELECTION OF CANDIDATES FOR POSITIONS 273

Volodymyr Khodatskyi, Anastasiia Nikolaienko. VISUALIZATION AND COMPARISON OF PATHFINDING ALGORITHMS BETWEEN TWO POINTS 277

Oleksandr Kolchyn, Vlad Volkov, Dariia Rudenko. TOWARDS REVEALING GAPS IN INPUT VALIDATION LOGIC VIA EVOLUTIONARY CONSTRAINT RELAXATION 280

Oleh Kotliar, Kseniia Dukhnovska. SOFTWARE FOR PERSONALIZED HRIS REPORT GENERATION: INTERACTIVE REPORT BUILDER WITH SERVER-SIDE PDF RENDERING 282

Yevhen Kutniak, Oleksandr Timinskyi. EFFECTIVE CONTROL OF TECHNICAL DEBT IN IT PRODUCTS 284

Nataliia Pashynska, Viktor Putrenko. THE SYSTEM APPROACH FOR DEVELOPMENT OF A DATABASE FOR THE PURPOSES OF REGIONAL CLIMATE CHANGE MODELING 288

Oleksandra Pryshchepa, Lyubov Kubiavka. INSURANCE BROKER ASSISTANT SYSTEM (IBAS): FROM IDEA TO IMPLEMENTATION 291

Mykhailo Rudyk, Yana Bielozerova. ARCHITECTURAL STRATEGIES AND TECHNICAL CHALLENGES OF DATABASE SHARDING IN HIGH-LOAD SYSTEMS 293

<i>Danylo Rybachok, Andrii Yarmilko. ARCHITECTURAL PARADIGMS FOR CONTESTED ENVIRONMENTS: A COMPARATIVE ANALYSIS OF MONOLITHIC AND MICROSERVICE DESIGNS FOR MODERN MILITARY LOGISTICS</i>	296
<i>Bohdan Semonov, Sergiy Pogorilyy. SEMANTIC ENRICHMENT OF SOURCE CODE IDENTIFIERS THROUGH AUTOMATIC ABBREVIATION EXPANSION FOR COMMIT MESSAGE GENERATION</i>	299
<i>Oleksandr Timinskyi, Denys Liashenko. ANALYSIS OF THE DEVELOPMENT DIRECTIONS OF COMPUTER SCIENCE MODELS IN THE FIELD OF IT PROJECT MANAGEMENT</i>	303
<i>Oleksandr Timinskyi, Artem Yurechko. COMPARATIVE ANALYSIS OF IT PROJECT MANAGEMENT METHODOLOGIES IN THE DEVELOPMENT OF WEB- AND CONTENT-ORIENTED SYSTEMS</i>	307
<i>Diana Vashchilina, Olena Vashchilina, Anastasiia Nikolaienko. STRUCTURING EDUCATIONAL IT PROJECTS WITH NPM WORKSPACES</i>	311
<i>Oleksii Vedmiediev. ASPECTS OF PROACTIVE ORGANIZATIONAL DEVELOPMENT OF CONSTRUCTION PROJECT TEAMS WHEN IMPLEMENTING VALUE HARMONIZATION WITH THE IT CONTEXT</i>	314
<i>Oleksandr Voitenko, Valerii Hudov. DAMPEN-TURBULENCE MODELS OF PROJECT PORTFOLIO MANAGEMENT TEAM OF A GENERAL CONTRACTING ORGANIZATION AND THEIR IT SUPPORT</i>	318
<i>Oleg Zhulkovskyi, Inna Zhulkovska, Anastasiia Shevchenko. OPTIMIZING NUMERICAL THERMAL SIMULATIONS THROUGH DATA-LEVEL PARALLELISM</i>	322
<i>Vadym Ziuziun, Tetiana Danilina. FEATURES OF DATABASE STRUCTURE DEVELOPMENT FOR THE "PETHEALTH" MOBILE APPLICATION</i>	324
<i>Vadym Ziuziun, Oleksii Korytnyi. RISK CONTROLING AND MONITORING SOFTWARES THAT CONSIDER HUMAN BIAS AS A MAJOR RISK</i>	326
<i>Vadym Ziuziun, Yuliia Koziuk. KEY CHALLENGES AND RISK MITIGATION STRATEGIES FOR UKRAINIAN IT STARTUPS DURING THE WAR</i>	328
<i>Vadym Ziuziun, Daria Shcherbak. MANAGEMENT OF QUALITY IN AN IT PROJECT DURING THE DEVELOPMENT OF THE ECOLOGICAL WEB APPLICATION "СМАЧХОГО"</i>	330

SECTION 1: ARTIFICIAL INTELLIGENCE TECHNOLOGIES AND DATA SCIENCE

Andrii Askerov

Bachelor, Master's Student

Lybov Borkovska

Candidate of Technical Sciences, Associate Professor

*STATE NON_COMMERTIAL ENTERPRISE "STATE UNIVERSITY "KYIV
AVIATION INSTITUTE", Ukraine*

AI IN SPORT: AN ARTIFICIAL PERSONAL TRAINER FOR DYNAMIC CYCLICAL PERIODIZATION BASED ON MACHINE LEARNING

Abstract. The paper presents the concept of an Artificial Personal Trainer (APT) system based on Machine Learning (ML) for dynamic and individualized cyclical periodization (training load planning). The system integrates multi-factor physiological data (HRV, sleep, recovery) from wearables with contextual information (weather, schedule) to automatically adjust training intensity and volume. This adaptive approach aims to minimize the risk of overtraining and maximize athletic performance, making professional-level periodization accessible to a wide audience.

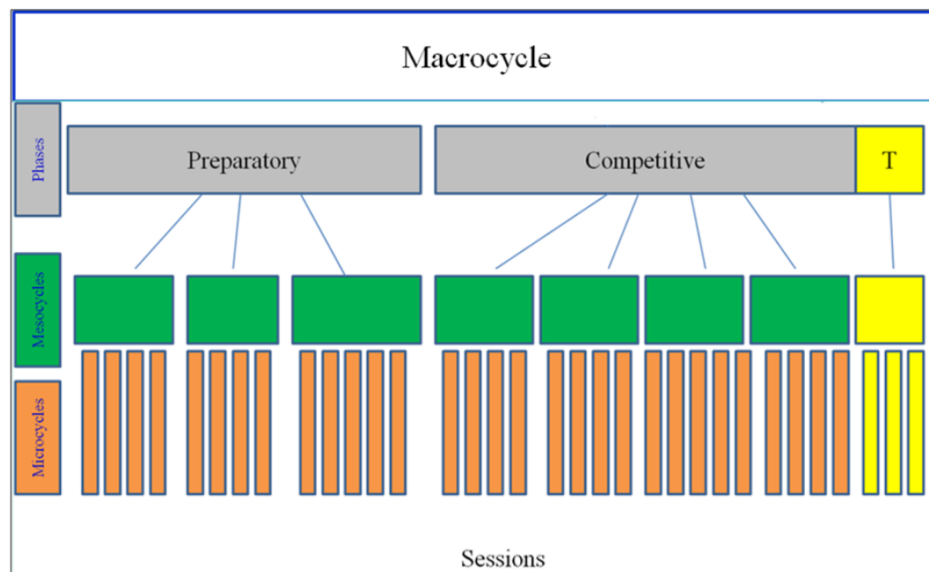
Keywords: Machine Learning, Wearable Technology, Training Plan, Digital Transformation, Periodization, Automatical Planning, HRV, Overtraining Prevention.

The rapid increase in the popularity of Wearable Technology devices (smartwatches, trackers) has led to an unprecedented accumulation of individual biometric and training data (heart rate, sleep quality, VO2 Max, exercise-specific metrics). This data is critical for athletes, as it is essential for the effective and safe planning of their training load – a foundation of professional training known as Periodization.

Periodization organizes the training process into phases and cycles to promote peak condition for the most important competitions. The macrocycle is divided into phases (preparatory, competitive and transition) each of which is allocated with a determined number of mesocycles that are designed to achieve specific physiological and performance goals (Figure 1).

Despite the wealth of data collected, most active individuals lack access to personalized expert analysis or a professional coach capable of effectively integrating this information into a dynamic training plan. Self-managing the cyclical nature of training load without deep knowledge of physiology and sports science often leads to overtraining, injury, or inefficient time use. Existing "smart" applications typically offer static plans, failing to adapt to instantaneous physiological changes (fatigue, illness) and external factors (weather, calendar commitments).

What I'm presenting is an Artificial Personal Trainer (APT) – an advanced Machine Learning (ML) system designed to deliver automated, dynamic, and individualized cyclical periodization.

Figure 1: Basic component of training plan. T - transition period

The APT integrates scientific methodologies with multivariate data gathered through two-way API integrations with major sports platforms (e.g., Garmin, Strava) and wearables (Garmin, Polar), analyzing both Physical Activity Data (power, heart rate, pace) and crucial Recovery Data (HRV, sleep quality). Its planning engine utilizes an adaptive periodization model based on key training stress metrics (like CTL, ATL, TSB). The final, structured workouts, complete with specific target zones (e.g., 105% of FTP), are stored internally and delivered directly to the athlete's device via device upload (Workout Push) and calendar synchronization, allowing the hardware to control the execution.

The core of the system is a specialized ML model (e.g., a Recurrent Neural Network or a Time Series model with an attention mechanism), trained to identify the athlete's unique physiological reaction patterns that lead to optimal progress or overtraining. Initial dynamic cyclical planning is performed by analyzing the athlete's goals, time frame, and baseline metrics, overlaying them onto the adaptive periodization model using training stress metrics (TSS, CTL, TSB).

In the event of an illness, the APT registers the illness (HRV drops 30% below baseline, temperature rises) and cancels workouts, replacing them with rest to allow recovery and prevent overtraining. the athlete's physical condition post-recovery is weaker than before the Mesocycle started, effectively nullifying over a month of effort (accumulated CTL is lost). Since the final competition date does not change, the APT cannot simply "postpone" the plan by a week. Instead, the APT rebuilds the subsequent two weeks for gradual recovery and adaptation (low volume, very low intensity) to safely restore the baseline level. Crucially, the system strategically sacrifices a less critical Mesocycle (e.g., shortening the general endurance phase) while preserving the peak phase (Tapering).

Thus, due to objective physiological constraints and the lack of time for full recovery, the APT predicts that achieving the initial goal result X is no longer possible. However, it ensures the best possible result (Result Y , where $Y < X$),

minimizing injury risks and preventing the athlete from forcing the preparation into an overtrained state. Instead of the planned intensive session, the system automatically assigns a recovery workout or suggests a rest day. Following the decision, the APT triggers a recalculation of the entire subsequent plan to integrate this deviation without losing the overall strategy for the competition peak.

The APT transforms passive data accumulation into an active, intelligent load management tool. It allows users to train effectively, minimizing injury risks and maximizing progress, without requiring constant consultation with a human coach. The system acts as a digital expert with 24/7 access to objective metrics of the body's condition. The emergence of the APT as a universally accessible tool will literally transform the lives of the active population, as it provides the ability to delegate the complex and time-consuming task of periodization to a specialized intellectual system. While today athletes entrust this task to professional coaches or spend time on self-planning, with this system, the plan will be created and adapted automatically according to goals, current status, and all other factors. This not only helps solve the task but can completely free the user from the need to constantly monitor and correct the training process, allowing them to focus exclusively on workout execution. This signifies a digital transformation of personal training, making it accessible, scientifically grounded, and dynamically adaptive for the general public.

In conclusion, the concept of the Artificial Personal Trainer, which uses advanced Machine Learning technologies for dynamic cyclical load planning based on athlete goals and multifactorial physiological data, formalizes the idea of an adaptive, self-adjusting system capable of mimicking the functions of a human expert coach and making personal training publicly available. Further research should focus on the development and validation of an MVP ML model for predicting the optimal load and recovery zone and testing the system prototype with a group of amateur athletes and qualified coaches to assess the system's effectiveness in improving athletic performance and preventing overtraining.

References:

1. Nanclerio, Fernando, Jeremy Moody, and Mark Chapman. 2013. "Applied Periodisation: A Methodological Approach". Cardiff Metropolitan University. <https://hdl.handle.net/10779/cardiffmet.19746763.v1>.
2. "A Coach's Guide to ATL, CTL & TSB". TrainingPeaks. <https://www.trainingpeaks.com/coach-blog/a-coachs-guide-to-atl-ctl-tsb/>.

Nikita Avdeev

Bachelor, student of the Faculty of Information Technologies,
Program systems and technologies department,

Iryna Yurchuk

Associate Professor, Department of Software Systems and Technologies, Faculty of
Information Systems

Taras Shevchenko National University of Kyiv

USING HYBRID NEURAL-NOISE METHODS TO DETECT ARTIFICIALLY GENERATED AND MANIPULATED IMAGES

Introduction. Rapid advances in generative AI have made it trivial to produce realistic images and subtle local manipulations (e.g., face swaps, mouth/eye edits). In parallel, the EU has adopted a risk-based regulatory approach that tightens verification requirements for higher-risk AI systems, which elevates the practical need for trustworthy detectors of synthetic and manipulated content [1–2].

This work presents a web-based detection system that combines statistical, noise-oriented analysis with neural network verification to deliver both an end-user verdict (real/fake/uncertain) and auditor-friendly evidence (heatmaps, masks, metrics) [1–2].

Background. Convolutional neural networks provide strong classification performance but often act as “black boxes,” motivating complementary explainable forensics signals [3].

We therefore pair neural models with noise-based modules: ELA (Error Level Analysis) for JPEG inconsistency, PRNU (sensor “fingerprint”) for camera-noise integrity, and a Noisesniffer-style ROI analyzer that detects statistically unlikely noise deficits in homogeneous blocks using an a-contrario Number of False Alarms (NFA) test (regions are accepted only if $NFA < \epsilon$) [3–4].

These signals guide neural verification, reducing false positives and helping capture subtle local edits [4].

System design (methods). Our system is a **hybrid pipeline** deployed as a web service (UI + REST API):

1. **Ingestion & preprocessing.** Safe intake (size limits, EXIF sanitization), standardization, and auto-resize to 512×512 enable fast, reproducible analysis [3].
2. **Noise-based detectors.**
 1. **ELA heatmaps** highlight compression inconsistencies [3].
 2. **PRNU correlation maps** validate sensor-noise integrity [3].
 3. **ROI noise analyzer** bins brightness, selects lowest-variance blocks, and performs region-growing with NFA significance to yield **explainable masks** [4].
3. **Neural verification (two-stage).**
 1. **Global model (250×250)** targets full-frame generations;
 2. **Local model (150×150)** focuses on small edits inside ROI (eyes, mouth, clothing fragments, objects) [3–4].

3. ROIs prioritize suspicious areas, stabilizing decisions in compressed or down-scaled imagery [4].
4. **DIRE module (our running example).**
 1. We illustrate the pipeline with a **DIRE-style reconstruction-error map** that exposes local reconstruction inconsistencies; it's rendered alongside ELA/PRNU/ROI overlays to show how neural residuals and noise cues align spatially [4–5].
5. **Fusion & scoring.** Rule-plus-weights fusion across modules with calibrated thresholds produces class probabilities and a **three-state verdict**: Real / Fake / Uncertain [3–4].
6. **Explainability & audit.** Interactive overlays (ELA/PRNU/ROI/DIRE), per-module scores, saved parameters (thresholds, ϵ , block sizes), and reproducible logs support audits and reporting [1–2].
7. **Interfaces.** Web dashboard for analysts (visual evidence + metrics) and REST/JSON for automated pipelines; export of reports (PDF/JSON) [1–2].

Visual example. Figure 1 — Example output. The UI shows: original image; **DIRE** residual heatmap; **ELA** inconsistency map; **PRNU** correlation visualization; and the **ROI/NFA** mask with the fused verdict (Real/Fake/Uncertain). This illustrates how statistical noise cues and neural residuals co-localize on manipulated regions [4–5].

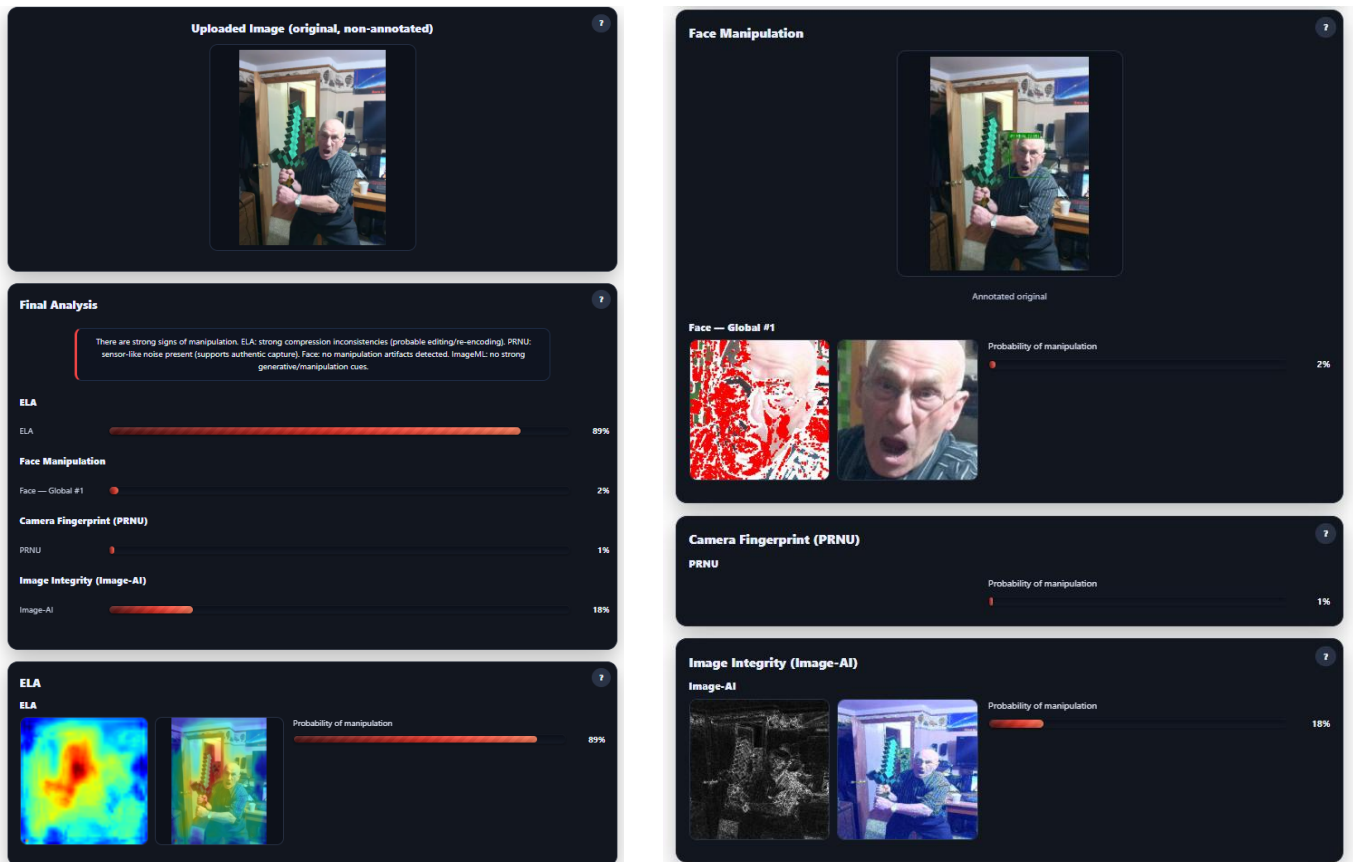


Figure 2. Example output program process

Conclusions. By integrating statistical noise forensics (ELA, PRNU, ROI/NFA) with neural verification (global + local, illustrated via a DIRE residual example), the system detects both fully generated images and localized manipulations while producing auditable, reproducible evidence (overlays, parameters, logs) appropriate for risk-aware environments [1–2,4–5].

References

1. European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*, L ... (12 July 2024). URL: <http://data.europa.eu/eli/reg/2024/1689/oj>
2. Coalition for Content Provenance and Authenticity (C2PA). (2025). *C2PA Technical Specifications*, v2.2. URL: <https://spec.c2pa.org/specifications/specifications/2.2/index.html>
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
4. Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*. URL: <https://arxiv.org/abs/1412.6980>
5. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *ICCV 2019*. URL: https://openaccess.thecvf.com/content_ICCV_2019/papers/Rossler_FaceForensics_Learning_to_Detect_Manipulated_Facial_Images_ICCV_2019_paper.pdf

Yehor Bashkevych

Master's student of the faculty of information technology

Oleh Zaritskyi

Doctor of engineering science, associate professor of management technology department

Taras Shevchenko National University of Kyiv

AUTOMATED SENTIMENT AND EMOTIONAL ANALYSIS OF GLOBAL NEWS ABOUT UKRAINE: DATA SCIENCE METHODS AND RESEARCH PERSPECTIVES

Abstract. This paper reviews modern Data Science approaches for automated sentiment and emotion analysis in global media texts about Ukraine, emphasizing how these methods and approaches jointly reflect international attitudes and emotional coloring in news discourse.

Keywords: Sentiment analysis, Emotion analysis, Multilingual NLP, Cross-lingual modeling, News framing.

Since 2022, multilingual news and social platforms have produced unprecedented volumes of Ukraine-related text, where sentiment extends beyond mere polarity to include emotional coloring (fear, anger, hope, sadness) shapes public attitudes and policy support. Today, mass media have become the main tool for distributing messages that affect the consciousness and subconsciousness of a person and, as a result, public consciousness, which should be considered as a social phenomenon, the main function of which is to influence the audience through the semantic assessment of information transmitted through mass media channels. Public consciousness should also be considered as a mechanism for updating information using various means of communication. Based on the above, the influence of mass media is not always negative, since in a globalized world it is difficult to imagine modernity without newspapers, social networks and other mass media. However, on the other hand, interested parties use mass media as a tool for influencing society, obtaining their own benefits [1].

Automated, cross-lingual models enable large-scale monitoring of emotional coloring dynamics across outlets and time, but evidence is scattered across methods, languages, and sources. In news specifically, headline/lede(*lead paragraph*) and full-text behave differently and warrant separate evaluation; standard emotion inventories (e.g., NRC or GoEmotions subsets) clarify what is being measured beyond polarity [2,3]. We consolidate what is known and surface gaps for high-quality, reproducible research.

Prior work is fragmented by (i) inconsistent taxonomies (sentiment vs. discrete emotions vs. valence–arousal), (ii) limited cross-lingual generalization (Ukr/EN/PL/DE/RU; domain shift news↔social), (iii) scarce emotion labels for news (esp. non-English), (iv) underuse of joint sentiment+emotion modeling for framing, and (v) weak robustness to bots/propaganda and source bias. These issues reduce comparability and risk misreading global attitudes [2-5].

We conducted a structured literature review across ACM/IEEE/Springer/ACL and open datasets. We mapped works along four axes: (A) model class (lexicon, classical ML, transformers); (B) supervision (supervised, distant/weak, zero-shot/LLMs); (C) cross-lingual strategy (MT translate-train/test, multilingual encoders, adapters/LoRA); (D) emotion inventory (Ekman/Plutchik/NRC vs. dimensional). We additionally tracked each study's data source (news vs. Twitter/Reddit/Telegram), dataset size, and whether/how it handled bots and propaganda [4-6].

Consistent with established practices for monitoring global media dynamics, we emphasize tracking the most popular news resources—such as Reuters, Xinhua, and major social platforms—to analyze emotional coloring and derive conclusions on evolving public sentiments and framing trends [2]. In practice, this involves large-scale data collection from high-impact outlets, preprocessing for multilingual NLP (e.g., tokenization, lemmatization), application of models like transformers for joint sentiment-emotion detection, and temporal analysis to uncover shifts in attitudes, such as dominant fear/anger in early 2022 phases transitioning to hope/resilience over time [3].

Across comparative baselines, multilingual transformers outperform lexicon-based and classical models for both polarity and discrete emotions; e.g., DistilRoBERTa achieves F1 0.43 (acc. 91%) vs. NRC's 0.23 (acc. 81%) on Ukraine-Russia tweets. Similarly, mBERT yields micro-F1 0.67 (acc. 70%) on Ukrainian social media sentiment, while hybrid setups remain attractive where labeled data are scarce. Translate-test is competitive on short news text (headlines/ledes) yet degrades on idiomatic opinion pieces; multilingual encoders reduce variance across large language sets. Integrating emotion distributions with sentiment logits sharpens framing signals and improves headline-level prediction around fear/anger-dominated events. Without bot/propaganda filtering, trendlines are unreliable; robustness checks are mandatory [4,5].

While these comparative baselines highlight the superiority of multilingual transformers in handling polarity and discrete emotions in Ukraine-related datasets, recent advancements underscore persistent gaps in real-world applicability. For instance, temporal drift in fast-evolving conflict narratives challenges model generalization across time, exacerbated by uneven non-English language coverage and machine translation's struggles with culture-specific idioms, sarcasm, and propaganda-laden stance cues [4,5]. These issues necessitate explicit evaluation criteria, diverse deduplicated datasets with human adjudication for ambiguities, and integrated robustness audits to ensure comparability. Moreover, ethical practice further requires attention to fairness, privacy, and accountability in real-time conflict monitoring [7].

To make the results comparable and decision-grade, we now focus on three steps: a small, event-stratified multilingual news benchmark; a joint sentiment-emotion baseline with calibrated, graded outputs; and a manipulation-aware pipeline that logs provenance and filters bots. In future developments, these elements will form the basis for an integrated NLP platform capable of real-time tracking of

emotional dynamics in Ukraine-related global news, leveraging scalable cloud infrastructure for multilingual data ingestion and visualization. This advancement aims to empower researchers and policymakers with reliable tools for detecting framing shifts and propaganda patterns, fostering more informed responses to evolving geopolitical narratives.

References:

1. S. Gnatyuk, N. Seilova, Yu. Polishchuk, O. Zaritskyi, The theoretical justification of quantitative evaluation method of mass media manipulative influence on public opinion. *Ukrainian Scientific Journal of Information Security* 24 (2018). doi: 10.18372/2225-5036.24.12953
2. Z. Jiang, Mining themes, emotions, and stance in the news coverage of the Russia–Ukraine war from Reuters and Xinhua. *Digital Scholarship in the Humanities* 39 (2024). doi:10.1093/llc/fqae015.
3. B. Breve, L. Caruccio, S. Cirillo, V. Deufemia, G. Polese, Analyzing the worldwide perception of the Russia–Ukraine conflict through Twitter. *Journal of Big Data* 11 (2024). doi:10.1186/s40537-024-00921-w.
4. A. Shevtsov, C. Tzagkarakis, D. Antonakaki, P. Pratikakis, S. Ioannidis, Twitter Dataset on the Russo-Ukrainian War. *arXiv* 2204.08530, 2022. doi:10.48550/arXiv.2204.08530.
5. M. Caprolu, A. Sadighian, R. Pietro, Characterizing the 2022- Russo-Ukrainian Conflict Through the Lenses of Aspect-Based Sentiment Analysis: Dataset, Methodology, and Key Findings, in: *Proc. IEEE ICCCN 2023*, 2023. doi:10.1109/ICCCN58024.2023.10230192.
6. M. Dean, B. Porter, Sentiment analysis of Russian-language social media posts discussing the 2022 Russian invasion of Ukraine. *Armed Forces & Society* 51 (2024). doi:10.1177/0095327X241235987.
7. O. Zaritskyi, Index of Artificial Intelligence Systems Ethics, in: *Information Technology and Implementation (IT&I-2022)*, Kyiv, Ukraine, Nov 30 – Dec 02, 2022, pp. 169–174. URL: <https://ceur-ws.org/Vol-3384/>.

Stepan Bilan

Ph.D., associate professor

Dmytro Bereznyi

Ph.D. student

Taras Shevchenko National University of Kyiv

OPTIMIZED KEYPOINT MATCHING IN UAV AGRICULTURAL IMAGERY WITH SIFT AND ORB

Abstract. Accurate keypoint detection is vital for stitching agricultural imagery used in crop monitoring and precision farming. This study compares ORB and SIFT algorithms, combined with RANSAC, on high-resolution UAV farmland images. Results highlight the impact of algorithm and parameter choice, suggesting that hybrid or learning-based methods may better balance accuracy and efficiency in future applications.

Keywords: keypoint detection, ORB, SIFT, RANSAC, image matching.

Introduction

Agricultural imagery is essential for crop monitoring, yield estimation, and precision farming. High-resolution UAV and satellite data offer detailed views of farmland, helping detect water stress, pests, and growth variability. A key challenge is stitching overlapping images into seamless mosaics, often hindered by repetitive textures and dynamic changes in agricultural scenes. Traditional feature-matching methods may fail under such conditions, leading to misalignments and reduced analysis accuracy. To address this, we propose an optimized image registration pipeline combining ORB and SIFT detectors with RANSAC for outlier rejection, aiming to improve alignment accuracy and support reliable agricultural analysis.

Data and methodology

We employed SIFT and ORB for keypoint detection and matching in high-resolution agricultural imagery. SIFT detects scale- and rotation-invariant features through four steps: identifying extrema using the Difference-of-Gaussians, refining unstable keypoints, assigning orientation from local gradients, and generating 128-dimensional descriptors [1]. These descriptors are robust, making SIFT effective in detecting subtle features in repetitive farmland scenes.

ORB provides a faster, lightweight alternative. It uses the FAST detector across image pyramids, applies rotation invariance via intensity centroid calculation, and ranks keypoints using the Harris corner measure. ORB descriptors are based on a rotation-aware BRIEF, encoding features through binary intensity comparisons [2]. While efficient, ORB may underperform in uniform regions.

To enhance robustness, we applied RANSAC for geometric verification. It estimates homographies and filters out inconsistent matches [3], improving alignment accuracy in the presence of noise, repetitive textures, and distortions—key for precise mosaicking and downstream agricultural analysis.

Results and analysis

Three ORB configurations were tested on UAV farmland imagery. Only the most optimized - using 5,500 keypoints, a finer scale factor (1.1), and more pyramid levels - achieved successful matches, yet still struggled in low-contrast areas. In contrast, SIFT consistently outperformed ORB, detecting more reliable matches in subtle or repetitive textures. After parameter tuning ($n_features = 500$, $ratio_threshold = 0.55$, $ransac_threshold = 2.0$), SIFT achieved a 100% inlier ratio, surpassing previous benchmarks. While robust overall, SIFT also faced difficulties in highly uniform regions. These results suggest hybrid or learning-based methods may further improve alignment accuracy in agricultural scenes.

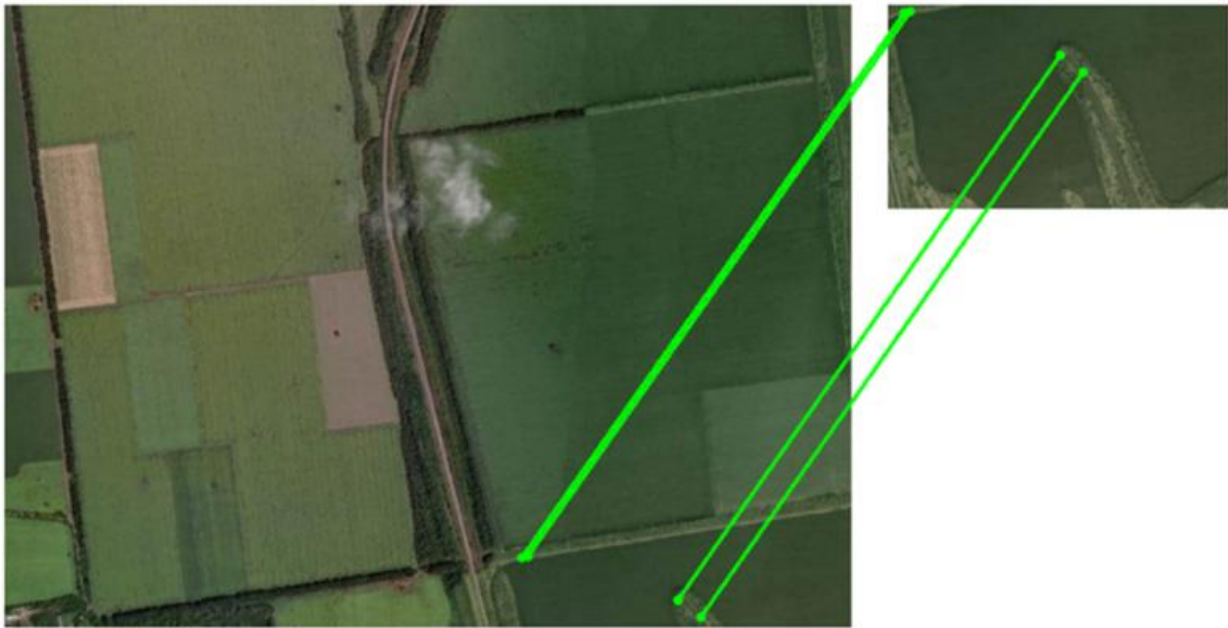


Figure 1. Results of SIFT+RANSAC model

Discussion and conclusion

ORB is fast but struggles with repetitive, low-contrast farmland imagery. SIFT is more computationally intensive but offers better robustness in detecting subtle features. Enhancements like adaptive thresholds and preprocessing can improve both methods. Alternatives such as AKAZE, BRISK, or deep learning-based SuperPoint show promise in texture-poor scenes. Our optimized ORB and SIFT configurations with RANSAC improved alignment, with SIFT proving more reliable overall for agricultural image analysis.

References

1. Lingua, A., Marenchino, D., & Nex, F. "Performance analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric applications ". (2009)
2. Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. "ORB: An efficient alternative to SIFT or SURF". (2011)
3. Baid, U. "Image Registration and Homography Estimation". (2015)

Oleh Bohutskyi

Bachelor in Computer Science, Master Student, Faculty of Information Technologies

Hanna Krasovska

PhD in Technical Sciences, Associate Professor, Department of Intelligent Technologies

Oleh Ilarionov

PhD in Technical Sciences, Associate Professor, Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv, olegbogutsky@knu.ua

EVALUATING FREQUENCY ANALYSIS FOR LARGE LANGUAGE MODEL QUERIES

Abstract. This study investigates the effectiveness of frequency analysis as an alternative to full-text inputs when querying large language models (LLMs). Within the context of a system for monitoring IT job market requirements, the research compares the quality of LLM-generated summaries when processing full job posting texts versus frequency-based representations of the same data. Experiments demonstrate that as the volume of analyzed vacancies increases, the quality of summaries generated from frequency data approaches that of full-text analysis, while offering significant advantages in processing efficiency and scalability. The findings have important implications for the design of large-scale text analysis systems that operate within LLM context window constraints.

Keywords: large language models, frequency analysis, text summarization, IT job market, context window optimization, natural language processing

1. Introduction

The rapid development of large language models (LLMs) has opened new possibilities for intelligent text analysis and summarization across various domains [1]. However, a significant limitation of these models is their restricted context window size, which constrains the amount of text that can be processed in a single query [2]. This limitation presents an essential challenge when analyzing large volumes of textual data, such as collections of job postings, research papers, or other extensive document sets.

Traditional approaches to handling large texts with LLMs often involve chunking or truncation, which can lead to loss of contextual information and degraded performance [3]. Recent studies have explored various techniques to optimize token usage, including compression methods, selective sampling, and information distillation [4]. However, less attention has been paid to the potential of frequency analysis as a lightweight alternative for representing large text collections when querying LLMs.

2. Methodology

The aim of this research is to develop and experimentally validate a novel approach that uses frequency analysis instead of full-text inputs when querying LLMs, and to assess its effectiveness in the specific context of monitoring IT job market requirements. The objectives include: analyzing current methods of LLM

utilization for document summarization, developing a methodology for comparing full-text versus frequency-based approaches, implementing a system for job posting collection and analysis, and conducting comparative experiments with varying data volumes.

This research employs methods from natural language processing, statistical analysis, and experimental computer science. The study utilizes architecture to collect job postings, perform frequency analysis on their content, and query LLMs using both full-text and frequency-based approaches. The quality of resulting summaries is evaluated using expert assessment.

The proposed methodology involves three main steps. First, job postings for IT positions are collected and organized into datasets of varying sizes (3, 5, and 10 job postings). Second, for each dataset, two types of LLM queries are performed: one with the full text of all job postings, and another with only the frequency analysis results (top-50 most frequently mentioned terms with their occurrence counts). Finally, the quality of the resulting summaries is evaluated based on criteria such as completeness, accuracy, structure, and generalizability.

The experiments demonstrate that while full-text analysis produces higher quality summaries when working with small datasets (3 job postings), the gap narrows significantly as the volume of data increases. With 10 job postings, frequency-based summaries achieve nearly identical quality scores (8.5 vs 8.7 on a 10-point scale) compared to full-text summaries. This finding aligns with previous research suggesting that frequency-based representations can effectively capture essential information from text collections [6]. Importantly, the frequency-based approach demonstrates superior processing efficiency, with query processing times up to 5.2 times faster than full-text queries for larger datasets.

These findings have significant implications for the design of systems that leverage LLMs for analyzing large text collections. By using frequency analysis, such systems can overcome context window limitations and achieve greater scalability without significant quality degradation. This approach is particularly valuable for applications such as job market monitoring, research literature analysis, and social media trend detection, where the volume of data often exceeds the context capacity of current LLMs [7].

In the specific context of IT job market analysis, the developed system demonstrates how frequency-based LLM queries can effectively identify key skill requirements and qualification expectations across numerous job postings. This capability provides valuable insights for educational institutions seeking to align their curricula with market demands, for professionals planning their career development, and for companies aiming to optimize their recruitment strategies.

3. Conclusion

The proposed research offers a practical solution to the context window limitations of large language models when analyzing extensive text collections. The experiments confirm the hypothesis that frequency analysis can serve as an effective alternative to full-text inputs for LLM queries, particularly when working with larger

datasets. The developed system for IT job market monitoring demonstrates the practical applicability of this approach, providing a scalable solution for extracting valuable insights from numerous job postings. Future research could explore the effectiveness of this methodology across different domains and investigate hybrid approaches that combine frequency analysis with selective inclusion of full-text segments for optimal results.

References:

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165. <https://arxiv.org/abs/2005.14165>
2. OpenAI. (2023). GPT-4 Technical Report. <https://cdn.openai.com/papers/gpt-4.pdf>
3. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. arXiv preprint arXiv:1901.02860. <https://arxiv.org/abs/1901.02860>
4. Wang, X., Salmani, M., Omid, P., Ren, X., Rezagholizadeh, M., & Eshaghi, A. (2024). Beyond the Limits: A Survey of Techniques to Extend the Context Length in Large Language Models. arXiv preprint arXiv:2402.02244. <https://arxiv.org/abs/2402.02244>
5. Novikova, J., Dušek, O., & Rieser, V. (2018). RankME: Reliable Human Ratings for Natural Language Generation. arXiv preprint arXiv:1803.05928. <https://arxiv.org/abs/1803.05928>
6. Radev, D. R., Jing, H., & Budzikowska, M. (2004). Centroid-based Summarization of Multiple Documents. In Proceedings of the Workshop on Text Summarization Branches Out. ACL. <https://aclanthology.org/W00-0403.pdf>
7. Wang, X., Chen, Z., Xu, T., Xie, Z., He, Y., & Chen, E. (2024). In-Context Former: Lightning-fast Compressing Context for Large Language Model. Findings of EMNLP 2024. <https://aclanthology.org/2024.findings-emnlp.138.pdf>

Kateryna Bondar, 4th year bachelor of computer science student

Iryna Domanetska, Ph.D., associate professor, associate professor at the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

ANALYSIS OF DATASETS SUITABLE FOR USE BY AI MODELS IN DIABETES MELLITUS SCREENING AND EARLY DETECTION

Abstract. This paper analyzes datasets suitable for training artificial intelligence (AI) models for screening and early detection of type 2 diabetes mellitus (T2D). The reliability of neural network-based systems depends on the representativeness and quality of data. Datasets are classified into five categories: tabular population surveys, electronic health records (EHR), signal data such as electrocardiograms (ECG), retinal fundus images, and magnetic resonance imaging (MRI). Key resources include NHANES for population data, MIMIC-IV for hospital EHR, and UK Biobank for multimodal deep learning. Smaller sets like Pima Indians Diabetes and Early Stage Diabetes Risk are useful for prototyping. The study concludes that robust AI models require integration of multimodal data, external validation, and standardized datasets to ensure clinical reliability and fairness.

Keywords: artificial intelligence, neural networks, diabetes mellitus type 2, early detection, screening, datasets.

Diabetes mellitus type 2 (T2D) remains one of the leading non-communicable diseases worldwide, imposing a significant burden on healthcare systems due to late diagnosis and the development of complications. Effective screening and early detection of T2D can substantially reduce the risk of micro- and macrovascular complications, simplify treatment, and lower overall costs. Recent advances in artificial intelligence (AI) and, in particular, artificial neural networks and ensemble learning models have demonstrated considerable potential to automate and enhance risk assessment for diabetes, especially at pre-diagnostic stages where traditional laboratory tests such as glycated hemoglobin (HbA1c) or fasting plasma glucose have not yet been performed. However, the reliability and transferability of such AI models depend critically on the quality and representativeness of the datasets used for training and validation. Consequently, identifying suitable datasets for the development of clinically reliable screening systems remains an essential research task.

All datasets can be classified into the following categories: tabular population surveys, electronic health records (EHR), signal data (ECG, electrocardiogram), retinal fundus images, and magnetic resonance imaging (MRI).

The National Health and Nutrition Examination Survey, maintained by the U.S. Centers for Disease Control and Prevention, provides one of the most representative population-level resources for AI-based diabetes modeling. It contains demographic, behavioral, anthropometric, and laboratory data, including HbA1c and fasting glucose levels. The dataset is publicly available through the CDC portal. NHANES is widely used for building population screening models and recalibrating algorithms to real-world disease prevalence. Nevertheless, it has certain limitations: variations in laboratory methods between survey cycles, potential mismatches

between HbA1c and glucose values, and predominantly cross-sectional design, which limits longitudinal prediction. An example of its application in machine learning modeling is presented by Qin [1].

The classical Pima Indians Diabetes Dataset, available via the UCI Machine Learning Repository [2] consists of 768 observations of female subjects from the Pima Indian community, with nine clinical predictors such as age, BMI, and blood pressure. It is frequently used as a benchmark in AI research because of its accessibility, yet its small size and narrow demographic profile make it unsuitable for building clinically generalizable screening systems.

Similarly, the Early Stage Diabetes Risk Prediction Dataset from Bangladesh [3] is convenient for methodological experimentation and ensemble model testing but lacks representativeness and diagnostic confirmation required for medical deployment.

EHR databases, such as MIMIC-IV, hosted on PhysioNet [4], contain laboratory results, diagnosis codes (ICD-9/10 E10–E14), and treatment histories for intensive-care patients. MIMIC-IV is valuable for the development of tabular AI models, imputation of missing data, and robustness testing. However, it is not representative of a screening population, as the subjects are already hospitalized, and diabetes is often pre-diagnosed. Thus, its use is limited to algorithmic development rather than population screening.

A rapidly growing research direction involves the use of non-invasive physiological signals, such as the electrocardiogram (ECG), for AI-based diabetes risk estimation. The UK Biobank ECG Dataset (Data-Field 20205) provides resting 12-lead ECG recordings for over 100 000 participants. Deep-learning approaches, including 1D-convolutional neural networks and residual networks, have been shown to predict prevalent and incident T2D from ECG waveforms alone. For instance, Pastika et al. (2025) demonstrated that AI-enhanced ECG analysis identifies cardiometabolic risk including diabetes mellitus [5]. These models open promising avenues for non-invasive population screening, though they require careful label alignment and external validation due to device and domain variability.

Another modality of interest is medical imaging. The UK Biobank Fundus Image Dataset (Data-Field 21016) contains more than 60 000 retinal fundus photographs accessible through the Biobank imaging portal. A convolutional neural network trained on this dataset achieved high accuracy for identifying individuals with T2D without laboratory data, as reported by Kim [6]. Likewise, whole-body Dixon MRI data have been applied to estimate visceral fat distribution and organ morphology using 3D-CNN architectures, as demonstrated in [7]. Although these datasets require substantial computational resources, they provide valuable insight into structural biomarkers associated with insulin resistance.

When selecting datasets for AI-based T2D screening, several factors must be critically assessed: the clarity and clinical validity of diabetes labels, the representativeness of the sample population, dataset size and class balance, the nature of predictors (invasive versus non-invasive), data quality and standardization, and the feasibility of external validation. Among the reviewed sources, the UK Biobank

emerges as the most comprehensive multimodal dataset suitable for deep-learning ensembles that combine ECG, retinal imaging, MRI, and clinical parameters. NHANES remains a benchmark for population calibration and epidemiological generalization, whereas MIMIC-IV offers a robust environment for model stress-testing and data-handling methods. Smaller datasets, such as Pima Indians and Early Stage Diabetes Risk, are better suited for educational or methodological prototyping rather than clinical deployment.

In conclusion, the development of AI systems for screening and early detection of type 2 diabetes depends not only on sophisticated neural architectures but, fundamentally, on the availability of large-scale, well-annotated, and representative datasets. Population-level resources like NHANES and UK Biobank, supplemented by curated EHR and imaging data, provide the necessary foundation for building robust, explainable, and transferable AI models capable of real-world implementation in preventive medicine.

References

1. Y. Qin, W. Liu et al. *Machine Learning Models to Predict Diabetes Risk Factors and Early Complications Using NHANES Data*. PLoS ONE 17(5) (2022) e0268194. DOI: 10.1371/journal.pone.0268194. Dataset: NHANES (<https://wwwn.cdc.gov/nchs/nhanes/>).
2. UCI Machine Learning Repository. *Pima Indians Diabetes Dataset*. Available at: <https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes>.
3. Kaggle. *Early Stage Diabetes Risk Prediction Dataset*. Available at: <https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset> (accessed 2025-11-02).
4. PhysioNet. *MIMIC-IV (Medical Information Mart for Intensive Care IV)*, Version 2.2 (2024). Available at: <https://physionet.org/content/mimiciv/2.2/>.
5. L. Pastika, S. H. Kim et al. *Artificial intelligence-enhanced electrocardiography identifies cardiometabolic risk including diabetes mellitus*. npj Digital Medicine 7 (2025) 256. DOI: 10.1038/s41746-024-01170-0. Dataset: UK Biobank ECG Data-Field 20205 (<https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=20205>).
6. S. H. Kim, J. H. Yun et al. *A deep learning model for screening type 2 diabetes from retinal photographs*. Computers in Biology and Medicine 144 (2022) 105366. DOI: 10.1016/j.compbiomed.2022.105366. Dataset: UK Biobank Fundus Data-Field 21016 (<https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=21016>).
7. C. Wachinger et al. *Deep learning prediction of type 2 diabetes from neck-to-knee MRI in the UK Biobank*. Heliyon 9(4) (2023) e15246. DOI: 10.1016/j.heliyon.2023.e15246. Dataset: UK Biobank Dixon MRI (<https://biobank.ndph.ox.ac.uk/ukb/imaging/>).

Yurii Cherevach

Master's degree Student, Department of Technologies Management

Andrii Khlevnyi

PhD, Associate Professor of the Department of Technologies Management

Taras Shevchenko National University of Kyiv, Ukraine

AUTOMATED DETECTION OF PROPAGANDA TECHNIQUES IN MULTILINGUAL CONFLICT NARRATIVES

Abstract. This work explores a transformer-ensemble pipeline augmented with argumentation-level features and multilayer explainability for automatic, traceable detection of propaganda techniques in Ukrainian, English, and Arabic conflict narratives – producing span-level outputs with technique labels, confidence scores, autogenerated fact-check queries and human-review flags, while addressing annotation protocols, adversarial robustness, and multilingual evaluation challenges.

Keywords: Propaganda detection, multilingual narratives, transformer models, argumentation features, explainable AI, NLP, information warfare, machine learning, fact-checking.

Modern information warfare involves the unprecedented spread of automated propaganda in multilingual conflict zones, undermining information integrity and democratic societies through emotional appeals, institutional distrust, and polarization [1, 2]. The speed and scale of disinformation require automated detection solutions inaccessible to manual analysis.

This research analyzes advanced detection methods and develops a structured methodology for identifying propaganda techniques in Ukrainian, English, and Arabic conflict narratives.

Natural Language Processing (NLP) technologies form the fundamental basis for automatic text content analysis. Basic NLP techniques apply tokenization, lemmatization, sentiment analysis, and named entity recognition to identify manipulative structures [1]. Despite scalability advantages, traditional NLP approaches demonstrate significant limitations related to contextual ambiguity, difficulties in detecting sarcasm and irony, and risks of training data bias [1].

Comparative analysis of three generations of machine learning models demonstrates clear methodological evolution [3]. Traditional approaches (SVM, Random Forest) showed lowest performance due to inability to capture context. Recurrent neural networks (LSTM, GRU) demonstrated better results but struggled with long textual dependencies. Transformer architectures (BERT, RoBERTa, ELECTRA) showed highest effectiveness through self-attention mechanisms. The ukr-electra-base model achieved 0.96 accuracy for labeling, 87% for appeal to fear,

and 82% for causal oversimplification [3]. XLM-RoBERTa proves effective for multilingual tasks [2].

An innovative direction is integration of argumentation features, based on the assumption that propaganda mimics logical argumentation (claims and premises) but fills it with manipulative content [2]. This approach identifies structural components and significantly improves model accuracy. Large language models (Llama 3.1) not only ensure high accuracy but also generate comprehensible explanations of their decisions, enhancing system transparency [4].

Analysis of multilingual corpora revealed common propaganda patterns. Emotional manipulation through loaded language (words with strong connotations) is most prevalent, with propaganda texts showing negative emotion levels of 0.064 versus 0.034 in neutral texts and subjectivity of 0.365 versus 0.283 [1]. Name-calling assigns characteristics without justification, while appeal to fear exploits anxiety and prejudices [5]. Logical fallacies include causal oversimplification (single cause for complex problems, e.g., "If France had not declared war on Germany, World War II would never have happened" [5]), false dilemma (binary choice, e.g., "No negotiations until aggression ends" [2]), straw man (distorting opponent's argument), and whataboutism (hypocrisy accusations). Other techniques include repetition, slogans, flag-waving, and appeal to authority [5].

The developed system concept integrates four feature types. Lexical features include word and bigram analysis, sentiment scores, subjectivity, emotional valence, and specialized lexicons. Syntactic features focus on parts of speech and causal conjunctions. Pragmatic features identify emotional appeals and decontextualization. Critically important are argumentation features identifying major claims, statements, and premises [1, 2, 3].

The computational pipeline includes preprocessing (cleaning, stopword removal, lemmatization), feature extraction, modeling through transformer ensemble with separate binary classifiers for each technique to handle imbalanced data [3]. Crucially important is integration of the explainability layer through LIME methods for visualizing influential words or generative LLMs for textual justifications [3, 4].

The system includes an automatic flagging mechanism for human review cases involving sarcasm, low model confidence, novel linguistic patterns, or overlapping spans with conflicting labels [1, 2]. Automatic generation of fact-checking queries for potentially manipulative claims is provided.

Transformer architectures supplemented with argumentation features represent the most effective solution for propaganda detection in multilingual conflict narratives, achieving up to 96% accuracy for individual techniques. The primary challenges include annotated data scarcity for less common languages, semantic

ambiguity in detecting sarcasm [1], imprecise span boundaries requiring refined protocols [2], and adaptability to evolving tactics. Development perspectives involve expanding language coverage, integrating multimodal analysis, enhancing explainability through XAI, and establishing international consortia for data exchange. Critical for practical success are structured annotation processes, continuous validation with human expertise, and adaptation to new manipulative tactics.

References:

1. V. M. Danylyk, V. A. Vysotska, and M. A. Nazarkevych, “Disinformation, fakes and propaganda identification methods in mass media based on machine learning,” 2024. <https://doi.org/10.28925/2663-4023.2024.25.449467>
2. S. Nabhani, C. Borg, K. Micallef, and K. Al-Khatib, “Integrating Argumentation Features for Enhanced Propaganda Detection in Arabic Narratives on the Israeli War on Gaza,” in *Proceedings of the First International Workshop on Nakba Narratives as Language Resources*, Abu Dhabi, UAE, 2025, pp. 127–149, Association for Computational Linguistics. Available: <https://aclanthology.org/2025.nakbanlp-1.14/>. Accessed: Oct. 29, 2025.
3. I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, E. Manziuk, and O. Barmak, “Method for neural network detecting propaganda techniques by markers with visual analytic,” *CEUR Workshop Proceedings*. Available: <https://ceur-ws.org/Vol-3790/paper14.pdf>. Accessed: Oct. 29, 2025.
4. M. Hasanain, M. A. Hasan, M. B. Kmainasi, E. Sartori, A. E. Shahroor, and F. Alam, “PropXplain: Can LLMs Enable Explainable Propaganda Detection?”. <https://doi.org/10.48550/arXiv.2502.16550>
5. G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, and P. Nakov, “SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles,” in *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain, 2020, pp. 1377-1414. Available: <https://aclanthology.org/2020.semeval-1.186.pdf>. Accessed: Oct. 29, 2025.

Vladislava Chernorai

PhD student

Oleksandra Kovalenko

PhD student

Vladimir Semenov

D.Sc., Professor, Professor at Department of Computational Mathematics

Taras Shevchenko National University of Kyiv

REGRESSION-TYPE PROBLEMS AND ALGORITHMS FOR VARIATIONAL INEQUALITIES

Abstract. The report discusses several applications of new algorithms for variational inequalities to solving regression-type problems. The idea is to first reduce the regression problem to a saddle point problem and then apply extragradient-type algorithms.

Keywords: regression-type problems, saddle-point problems, variational inequality, extragradient algorithms, operator extrapolation algorithm.

Some data analysis and machine learning problems can be formulated as

$$F(Kx) + G(x) \rightarrow \min_{x \in R^n}, \quad (1)$$

where $F: R^m \rightarrow R \cup \{+\infty\}$, $G: R^n \rightarrow R \cup \{+\infty\}$ are proper closed and convex functions, $K: R^n \rightarrow R^m$ is a linear operator [1]. Problem (1) and its dual

$$-G^*(-K^*y) - F^*(y) \rightarrow \max_{y \in R^m}$$

can be formulated as the following saddle-point problem (primal-dual formulation)

$$\min_{x \in R^n} \max_{y \in R^m} (\langle Kx, y \rangle + G(x) - F^*(y)). \quad (2)$$

Here

$$F^*(y) = \sup_{x \in R^n} (\langle y, x \rangle - F(x))$$

is the conjugate function for $F: R^m \rightarrow R \cup \{+\infty\}$. Details can be found in [1].

Suppose that problem (2) has a solution $(\bar{x}, \bar{y}) \in R^n \times R^m$. Then it satisfies the operator inclusion (which is equivalent to the variational inequality)

$$\begin{cases} 0 \in K^*\bar{y} + \partial G(\bar{x}), \\ 0 \in -K\bar{x} + \partial F^*(\bar{y}), \end{cases} \quad (3)$$

where ∂G and ∂F^* are subdifferentials of the functions G and F^* [1]. The inclusion (3) can be written in the following form:

$$\text{find } (\bar{x}, \bar{y}) \in R^n \times R^m \text{ such that } \begin{cases} G(x) - G(\bar{x}) + \langle K^*\bar{y}, x - \bar{x} \rangle \geq 0 & \forall x \in R^n, \\ F^*(y) - F^*(\bar{y}) - \langle K\bar{x}, y - \bar{y} \rangle \geq 0 & \forall y \in R^m. \end{cases} \quad (4)$$

Furthermore, problem (3) (or (4)) can be solved by algorithms for variational inequalities [2].

For example, if the functions G and F^* are smooth, then for (3) the operator extrapolation method [2] has the form

$$\begin{cases} x^{k+1} = x^k - \lambda_k (K^* y^k + \nabla G(x^k)) - \lambda_{k-1} (K^* y^k + \nabla G(x^k) - K^* y^{k-1} - \nabla G(x^{k-1})), \\ y^{k+1} = y^k + \lambda_k (Kx^k - \nabla F^*(y^k)) + \lambda_{k-1} (Kx^k - \nabla F^*(y^k) - Kx^{k-1} + \nabla F^*(y^{k-1})), \end{cases}$$

where $\lambda_k > 0$. In the non-smooth case, the scheme can be used

$$\begin{cases} x^{k+1} = \text{prox}_{\lambda_k G} (x^k - \lambda_k K^* y^k - \lambda_{k-1} (K^* y^k - K^* y^{k-1})), \\ y^{k+1} = \text{prox}_{\lambda_k F^*} (y^k + \lambda_k Kx^k + \lambda_{k-1} (Kx^k - Kx^{k-1})). \end{cases}$$

In the form (1), the ROF and TV-L1 models in the noisy image restoration problems are formulated [1].

Let us consider several examples.

Let us start with the problem of ℓ_2 -regularized ℓ_2 -regression:

$$\frac{1}{2} \|Kx - b\|_2^2 + \frac{\lambda}{2} \|x\|_2^2 \rightarrow \min_{x \in R^n},$$

where $b \in R^m$, $K: R^n \rightarrow R^m$ is a linear operator, $\lambda > 0$. The saddle formulation has the form

$$\min_{x \in R^n} \max_{y \in R^m} \left(\langle Kx, y \rangle + \frac{\lambda}{2} \|x\|_2^2 - \frac{1}{2} \|y\|_2^2 - \langle b, y \rangle \right), \quad (5)$$

which is equivalent to the equation

$$\begin{cases} 0 = K^* \bar{y} + \lambda \bar{x}, \\ 0 = -K\bar{x} + \bar{y} + b. \end{cases}$$

The operator extrapolation method for problem (5) has the form

$$\begin{cases} x^{k+1} = x^k - \lambda_k (K^* y^k + \lambda x^k) - \lambda_{k-1} (K^* y^k + \lambda x^k - K^* y^{k-1} - \lambda x^{k-1}), \\ y^{k+1} = y^k + \lambda_k (Kx^k - y^k - b) + \lambda_{k-1} (Kx^k - y^k - Kx^{k-1} + y^{k-1}), \end{cases}$$

where $\lambda_k > 0$. Various selection rules λ_k are described in [2].

Let us consider the problem of ℓ_1 -regularized ℓ_2 -regression (Lasso, least absolute shrinkage and selection operator):

$$\frac{1}{2} \|Kx - b\|_2^2 + \lambda \|x\|_1 \rightarrow \min_{x \in R^n},$$

where $b \in R^m$, $K: R^n \rightarrow R^m$ is a linear operator, $\lambda > 0$. The saddle formulation has the form

$$\min_{x \in R^n} \max_{y \in R^m} \left(\lambda \|x\|_1 + \langle Kx, y \rangle - \frac{1}{2} \|y\|_2^2 - \langle b, y \rangle \right).$$

The operator extrapolation method for this problem has the form

$$\begin{cases} x^{k+1} = \text{prox}_{\lambda_k \lambda \|\cdot\|_1} (x^k - \lambda_k K^* y^k - \lambda_{k-1} (K^* y^k - K^* y^{k-1})), \\ y^{k+1} = y^k + \lambda_k (Kx^k - y^k - b) + \lambda_{k-1} (Kx^k - y^k - Kx^{k-1} + y^{k-1}), \end{cases}$$

where $\lambda_k > 0$. Note that $\text{prox}_{\mu\|\cdot\|_1}$ it is calculated explicitly [1]:

$$\left(\text{prox}_{\mu\|\cdot\|_1} x\right)_i = \begin{cases} x_i - \mu, & x_i \geq \mu, \\ 0, & |x_i| < \mu, \\ x_i + \mu, & x_i \leq -\mu. \end{cases}$$

Consider the ℓ_2 -regression problem (least squares method) with the non-negativity condition:

$$\frac{1}{2}\|Kx - b\|_2^2 \rightarrow \min, \quad x \geq 0,$$

where $b \in R^m$, $K: R^n \rightarrow R^m$ is a linear operator. The saddle formulation has the form

$$\min_{x \in R^n} \max_{y \in R^m} \left(\langle Kx, y \rangle + \delta_{R_+^n}(x) - \frac{1}{2}\|y\|_2^2 - \langle b, y \rangle \right),$$

where $R_+^n = \{x \in R^n : x \geq 0\}$, δ_M is an indicator function of the set M . We obtain the following inclusion

$$\begin{cases} 0 \in K^* \bar{y} + N_{R_+^n} \bar{x}, \\ 0 = -K\bar{x} + \bar{y} + b, \end{cases}$$

where $N_C z$ is a normal cone of the set C at the point z . The operator extrapolation method for this problem has the form

$$\begin{cases} x^{k+1} = P_{R_+^n} \left(x^k - \lambda_k K^* y^k - \lambda_{k-1} (K^* y^k - K^* y^{k-1}) \right), \\ y^{k+1} = y^k + \lambda_k (Kx^k - y^k - b) + \lambda_{k-1} (Kx^k - y^k - Kx^{k-1} + y^{k-1}), \end{cases}$$

where $\lambda_k > 0$.

Consider the ℓ_1 -regression problem (the method of least moduli) with the non-negativity condition:

$$\|Kx - b\|_1 \rightarrow \min, \quad x \geq 0,$$

where $b \in R^m$, $K: R^n \rightarrow R^m$ is a linear operator. The saddle formulation has the form

$$\min_{x \in R^n} \max_{y \in R^m} \left(\langle Kx, y \rangle + \delta_{R_+^n}(x) - \delta_{B_\infty(R^m)}(y) - \langle b, y \rangle \right),$$

where $B_\infty(R^m) = \{y \in R^m : \|y\|_\infty = \max_i |y_i| \leq 1\}$ is a closed unit ball in the ℓ_∞ -norm. We obtain the following inclusion

$$\begin{cases} 0 \in K^* \bar{y} + N_{R_+^n} \bar{x}, \\ 0 \in -K\bar{x} + b + N_{B_\infty(R^m)} \bar{y}. \end{cases}$$

The operator extrapolation method for this problem has the form

$$\begin{cases} x^{k+1} = P_{R_+^n} \left(x^k - \lambda_k K^* y^k - \lambda_{k-1} (K^* y^k - K^* y^{k-1}) \right), \\ y^{k+1} = P_{B_\infty(R^m)} \left(y^k + \lambda_k (Kx^k - b) + \lambda_{k-1} (Kx^k - Kx^{k-1}) \right), \end{cases}$$

where $\lambda_k > 0$.

Let us now consider the problem of ℓ_r -regularized ℓ_p -regression ($p, r \in (1, +\infty)$):

$$\frac{1}{p} \|Kx - b\|_p^p + \frac{\lambda}{r} \|x\|_r^r \rightarrow \min_{x \in R^n},$$

where $b \in R^m$, $K: R^n \rightarrow R^m$ is a linear operator, $\lambda > 0$. The corresponding saddle formulation has the form

$$\min_{x \in R^n} \max_{y \in R^m} \left(\langle Kx, y \rangle + \frac{\lambda}{r} \|x\|_r^r - \frac{1}{q} \|y\|_q^q - \langle b, y \rangle \right), \quad (6)$$

where $q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, which is equivalent to the equation

$$\begin{cases} 0 = K^* \bar{y} + \lambda |\bar{x}|^{p-1} \cdot \operatorname{sgn} \bar{x}, \\ 0 = -K\bar{x} + |\bar{y}|^{q-1} \cdot \operatorname{sgn} \bar{y} + b, \end{cases}$$

where the operations on the right are understood component-wise. The operator extrapolation method for saddle point problem (6) has the form

$$\begin{cases} x^{k+1} = x^k - \lambda_k \left(K^* y^k + \lambda |x^k|^{p-1} \cdot \operatorname{sgn} x^k \right) - \\ \quad - \lambda_{k-1} \left(K^* y^k + \lambda |x^k|^{p-1} \cdot \operatorname{sgn} x^k - K^* y^{k-1} - \lambda |x^{k-1}|^{p-1} \cdot \operatorname{sgn} x^{k-1} \right), \\ y^{k+1} = y^k + \lambda_k \left(Kx^k - |y^k|^{q-1} \cdot \operatorname{sgn} y^k - b \right) + \\ \quad + \lambda_{k-1} \left(Kx^k - |y^k|^{q-1} \cdot \operatorname{sgn} y^k - Kx^{k-1} + |y^{k-1}|^{q-1} \cdot \operatorname{sgn} y^{k-1} \right), \end{cases}$$

where $\lambda_k > 0$.

The authors are partially supported by the NASU (0124U002162).

References:

1. E. Ryu, W. Yin, Large-Scale Convex Optimization, Cambridge University Press, Cambridge, 2023.
2. V.V. Semenov, S.V. Denisov, Convergence of the Method of Extrapolation from the Past for Variational Inequalities in Uniformly Convex Banach Spaces. Cybernetics and Systems Analysis, volume 58, (2022) p. 564-575.

Anastasiia Chubukova

Master's degree student of the Faculty of Information Technology

Oleh Zaritskyi

Doctor of Engineering Science, Professor (Associate) of information technology department, Taras Shevchenko National University of Kyiv

Taras Shevchenko National University of Kyiv

MODELING THE IMPACT OF EMOTIONALITY IN DIGITAL COMMUNICATIONS ON CUSTOMER BEHAVIOR

Abstract. This paper explores the impact of emotionality in brand digital communications on customer behavior. Based on an analysis of international studies, it is confirmed that positively toned content increases engagement, trust, and the likelihood of repeat purchases. The proposed approach integrates emotional tone analysis of digital messages with customer transactional data to quantitatively measure this impact.

Keywords: communication emotionality, customer behavior, sentiment analysis, loyalty, digital marketing, NLP, customer analytics

With the rise in popularity of social media and online marketing, emotional content has become an important part of brands' digital communications. It plays a significant role in building trust, loyalty, and emotional attachment among consumers. Companies increasingly rely on online channels, while consumers have become more sensitive to emotional tone of communication. Consumers now evaluate products and services not only from a functional perspective - the emotional dimension is equally important [1]. Therefore, interest in studying how the emotionality of online communications affects customer behavior continues to grow.

This is confirmed by international research. For example, according to NeuroScienceMarketing, campaigns with emotional content show almost twice as high results (31% vs. 16%) as rational ones [2]. Nielsen research has shown that advertising that elicits an above-average emotional response contributes to a 23% increase in sales [3]. In addition, 71% of customers recommend a brand with which they have an emotional connection [4].

Given the growing role of emotional factors in consumer behavior, research on analyzing brand communications through the lens of emotionality has gained popularity. In particular, Hongxin Li (2025), using NLP methods of emotional tone analysis and user clustering based on behavioral characteristics, shows that a positive emotional tone of content increases the level of interaction by an average of 47% [5].

The article "The Power of Emotional Advertising Appeals" (2023), which uses experimental design and structural equation modeling (SEM), proves that emotional stimuli in advertising increase the likelihood of purchase by 28–35%, and positive emotions strengthen brand loyalty [6].

The study "The Role of Sentiment Analysis in Brand Management and Marketing" (2024), based on the analysis of more than 50,000 posts using BERT and Random Forest models, showed that a 10% increase in the share of positive content

increases user engagement by 8–12%. This confirms that systematic sentiment analysis can more accurately predict customer behavior and enhance loyalty [7].

Thus, all three studies confirm the importance of measuring the emotionality of digital communications and prove its significant impact on customer activity. However, most of the available work is based on social media data and surveys, while opening up the prospect of integrating such approaches with transactional data for a deeper understanding of actual purchases.

Building on these findings, this paper proposes an approach that combines emotional tone analysis of digital messages with behavioral customer data to quantitatively assess how emotional coloration influences repeat purchases and engagement. The methodology involves sentiment analysis to measure the emotional profile of content, clustering customers by sensitivity to emotional communication, and regression modeling to quantify the strength of the effect. The expected outcome is an analytical model that enables companies to predict customer reactions and optimize communication strategies based on emotional tone.

Conclusions. Emotionality in online communications has a significant impact on brand perception, trust levels, and customer behavior. Research shows that positively colored content increases engagement and the likelihood of repeat purchases, while excessive emotionality can have the opposite effect. Studying these patterns opens up new opportunities for developing more effective strategies for interacting with customers in the digital environment.

References:

1. Xingjie He, Lixiao Zhu, Lin Sun, Linqian Yang. The influence of brand marketing on consumers' emotion in mobile social media environment. 2022. doi: 10.3389/fpsyg.2022.962224
2. We're Ruled by Our Emotions, and So Are the Ads We Watch. URL: <https://www.nielsen.com/insights/2016/were-ruled-by-our-emotions-and-so-are-the-ads-we-watch/>
3. Roger Dooley. Emotional Ads Work Best. URL: <https://www.neurosciencemarketing.com/blog/articles/emotional-ads-work-best.htm>
4. Viraj T. Motista's Study Results Conclude That Emotional Connection Is the Key to Brand Success. URL: <https://martechseries.com/sales-marketing/customer-experience-management/motistas-study-results-conclude-emotional-connection-key-brand-success/>
5. Hongxin Li. Social Media Data Mining and Online Consumer Behavior Analysis. *Procedia Computer Science*. Volume 261, 2025, Pages 406-413. doi: 10.1016/j.procs.2025.04.220
6. David Vrtana, Anna Krizanovna. The Power of Emotional Advertising Appeals: Examining Their Influence on Consumer Purchasing Behavior and Brand–Customer Relationship. *Sustainability*, 2023, 15(18), 13337. doi: 10.3390/su151813337
7. Mirza Selimovic, Ali Abd Almisreb, Salue Amanzholova. The Role of Sentiment Analysis in Brand Management and Marketing: A Comparative Study. *Procedia Computer Science*. Volume 251, 2024, Pages 579-584. doi: 10.1016/j.procs.2024.11.152

Ivan Demenko

PhD student

Mykola Fesenko

Candidate of Technical Sciences, Associate Professor, Senior Researcher at the Laboratory №206 for Applied Informatics Problems

Iryna Kutova

Head of Postgraduate Department

V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine

MULTI-AGENT SIMULATION ENVIRONMENT FOR INTELLIGENT OPTIMIZATION OF AUTOMATED TESTING PROCESSES

Abstract. This paper presents a multi-agent simulation environment designed for intelligent optimization of automated testing processes. The system models CI/CD workflows and enables the study of adaptive test planning strategies. The architecture includes testing, optimization, and monitoring agents that interact through a central controller using reinforcement learning methods. The proposed approach improves testing efficiency, coverage, and overall process stability.

Keywords: multi-agent systems, simulation modeling, automated testing, reinforcement learning, CI/CD optimization.

Automated testing is a key component of modern software development approaches, particularly within continuous integration and delivery (CI/CD) pipelines. As system complexity increases, so does the need for rapid and accurate execution of numerous tests under dynamic deployment conditions. Traditional methods based on fixed scenarios lack sufficient adaptability, whereas the use of intelligent algorithms – such as multi-agent systems and machine learning techniques – enables the creation of self-learning systems capable of optimizing testing processes in real time.

The goal of this research is to develop a multi-agent simulation environment for studying and optimizing automated testing processes. It models the interaction of CI/CD components and provides a safe environment for experimenting with various test planning and execution strategies, allowing for the evaluation of algorithm efficiency and agent behavior under changing conditions.

The simulation environment reproduces the complete software product lifecycle within a CI/CD pipeline – from code commit to deployment and feedback. Its architecture is based on an event-driven model, where each modification in the code repository triggers an automated chain of testing activities. The environment integrates several specialized modules responsible for event generation, workload scheduling, execution time measurement, error detection, and performance metric aggregation. Additionally, it supports configurable simulation parameters such as resource constraints, concurrency levels, and test prioritization policies. This design enables researchers to emulate realistic testing workflows, analyze the impact of varying load conditions, and evaluate the adaptability and robustness of optimization algorithms in a controlled yet representative CI/CD context.

Figure 1 presents the architecture of the multi-agent simulation environment. It illustrates the interaction of different types of agents through a shared simulation environment and a central controller that provides coordination of actions and data exchange.

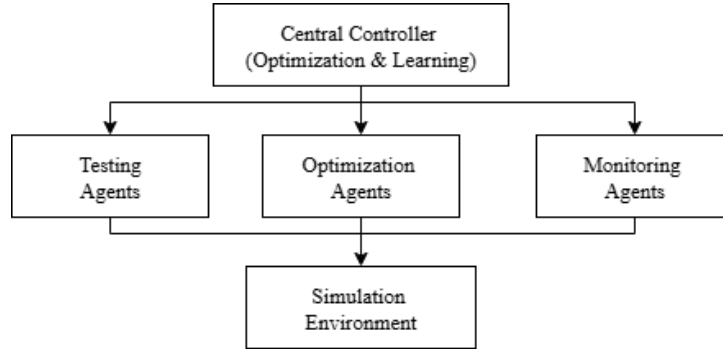


Figure 1 – Architecture of the multi-agent simulation environment for automated testing process optimization

Additionally, the environment includes tools for monitoring agent performance and collecting statistics in formats compatible with analytics systems such as Prometheus and Grafana. This enables interactive analysis of experimental results, including visualization of key metrics such as test coverage, execution time, resource utilization, and synchronization level among agents.

The system implements four types of agents, each with clearly defined roles:

1. Testing agents – execute test cases, collect results, evaluate performance, and report detected errors while considering execution history and task context.
2. Optimization agents – adjust the order of test execution and resource allocation using the Multi-Agent Reinforcement Learning (MARL) approach to minimize testing time and improve coverage.
3. Monitoring agents – track the overall system state and report deviations or performance degradation.
4. Coordinator – aggregates data and performs meta-optimization by combining Q-learning algorithms with evolutionary methods such as the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to search for global optimal solutions.

The system implements a hybrid approach that combines reinforcement learning methods with evolutionary algorithms. This integration allows the achievement of a balance between local and global optima when distributing tests among agents. The agents' behavior is defined through a reward function that integrates key testing metrics:

$$R_i = \alpha \frac{C_i}{C_{max}} + \beta \frac{1}{T_i} - \gamma D_i, \quad (1)$$

where R_i is the reward of agent i , C_i represents test coverage, T_i is test execution time, D_i is the duplication coefficient, and α, β, γ are weighting coefficients that determine the priority between optimization criteria.

Agent coordination is managed through a message broker (RabbitMQ or MQTT) that supports asynchronous communication between system components.

The Contract Net Protocol is used to align actions, providing a flexible task distribution mechanism based on the «call-for-proposal» interaction model.

Algorithmically, agent interaction is implemented using Multi-Agent Deep Deterministic Policy Gradient (MADDPG) and Proximal Policy Optimization (PPO) methods. These techniques enable agents to form cooperative strategies in a partially observable environment and avoid redundant test execution. To enhance scalability, Federated Learning elements are incorporated, allowing agents to exchange policies without centralized data storage, thereby improving efficiency and system robustness.

To evaluate the system's efficiency, simulation experiments were conducted comparing three approaches: static test scheduling, heuristic adaptation, and the proposed multi-agent approach.

The results, summarized in Table 1, clearly demonstrate the advantage of the agent-based model across all performance indicators.

Table 1 – Comparison of testing strategy efficiency

Strategy Type	Average Execution Time (s)	Code Coverage (%)	Redundant Tests (%)
Static scheduling	120	82	15
Heuristic adaptation	95	85	10
Multi-agent approach	78	91	6

The main advantage of the multi-agent approach lies in the system's ability to self-organize, learn, and adapt without manual intervention. Agents can operate both autonomously and collaboratively, achieving global objectives through local optimizations. This ensures fault tolerance, flexible scalability, and adaptability to various types of tests and infrastructure environments.

The integration of the developed simulation environment into existing CI/CD platforms (GitLab CI, Jenkins, GitHub Actions) involves the use of REST APIs and containerization technologies such as Docker and Kubernetes. A hybrid layer is planned between the simulation platform and real CI/CD pipelines, enabling agents to be trained on historical test data and then deployed as an optimization module in production environments. This approach paves the way for adaptive, intelligently managed CI/CD processes, where testing becomes an active mechanism for improving the overall efficiency of software development.

References:

1. J. Doddapaneni, AI test design and script generator: Enhancing software testing through AI-driven automation, *International Journal of Multidisciplinary Research and Growth Evaluation*, 2025. doi:10.54660/ijmrge.2025.6.1.1942-1943.
2. R. Nooyens, T. Bardakci, M. Beyazit, and S. Demeyer, Test amplification for REST APIs via single and multi-agent LLM systems, in: *Proceedings of the International Conference on Testing Software and Systems*, 2025. doi:10.48550/arXiv.2504.08113.
3. E. Kesavan, AI-driven and autonomous testing, *International Scientific Journal of Engineering and Management*, 2024. doi:10.55041/isjem01651.
4. D. I. Symonov, Identification and control of chaotic processes in complex technical systems, *Scientific Bulletin of Uzhhorod University. Series Mathematics and Informatics*, 46(1), pp. 273–284, 2025.

Pavlo Derevianko

Bachelor in Software engineering, Master Student, Faculty of Information Technologies

Ihor Miroshnychenko

PhD in Economics, Associate Professor at the Department of Management Technologies, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

Generative AI IN THE STUDENT ENVIRONMENT: AN ANALYTICAL REVIEW OF EXPECTATIONS, RISKS, AND PROSPECTS

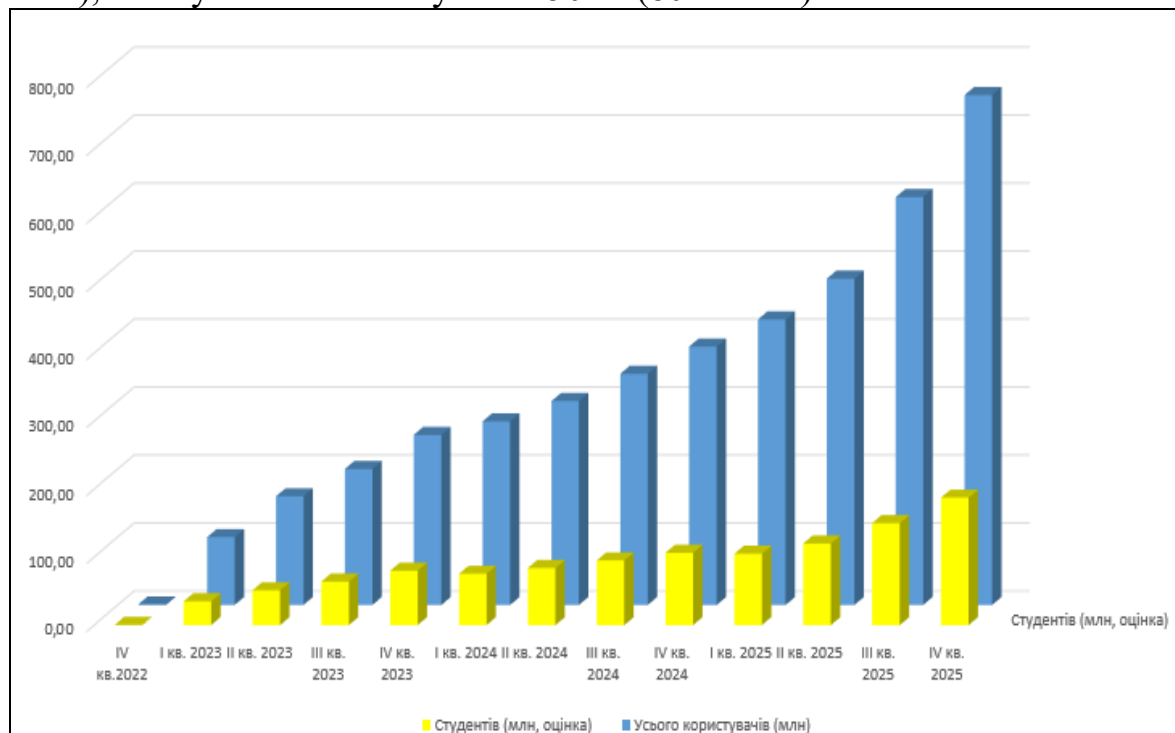
Abstract. The study examines the rapid growth of generative AI (GAI) users, many of whom are students, focusing on its integration into education and its potential to enhance learning effectiveness and digital literacy.

Keywords: GAI, education, students, analysis, risks, expectations, prospects, Ukraine, USA, China, EU.

GAI is becoming increasingly integrated into higher education as an innovative tool and an indispensable assistant in the learning process. Studies show that students view GAI positively, using it as an effective resource for independent learning.

Review of the Dynamics of GAI User Growth (including students).

Between late 2022 and the end of 2025, the number of GAI users increased at an exceptional rate. In Q4 2022, the global audience reached approximately 1M users, of whom around 0.35M were students. By Q1 2023, this number had risen to nearly 100M (35M stud), representing more than a hundred-fold increase within just three months-a record-breaking growth rate for digital technologies in the past decade. Throughout 2023, the trend continued: Q2 - 160 M (51.2M stud.), Q3 - 200M (64M stud.), and by the end of the year - 250 M (80M stud.).



In 2024, the number of GAI users grew by 20-30M per quarter (10-12M stud), reaching ~380M total (~106M stud). In 2025 the growth accelerated: Q1 - 420M (105M stud), Q2 - 480M (120M stud), Q3 - 600M (150M stud). Forecasts suggest nearly 750M users by the end of 2025, including ~187M stud.

Over three years, the total GAI audience has grown almost 700-fold, and the number of student users has increased more than 500-fold.

It is therefore appropriate to analyze the specifics of GAI integration into the learning process, considering the risks, expectations, and prospects of its use in different countries, such as the United States, China, the EU, and Ukraine.

1 Analysis of the risks of using GAI in the educational process:

1.1. **Plagiarism and declining academic integrity:** USA - 58% (Inside Higher Ed, 2024); China - 72% (Shanghai Education Review, 2024); EU - 49% (European Digital Education Hub, 2024); Ukraine - 37% (IITZN NAPS of Ukraine, 2024).

1.2. **Dependency and reduced cognitive engagement.** USA: 41% (Campus Technology, 2024); China: 47%, among first-year students - over 60%; EU: 34% (EU Youth AI Survey, 2024); Ukraine: about 28% (IITTA, 2024).

1.3. Security risks and potential data leakage are acknowledged by: EU - 61%, USA - 42%, China - 25%, Ukraine - 46% (Deloitte AI Trust Survey, 2024).

1.4. Network coverage. In the USA, China, and the EU, more than 70% of students have stable access to GAI, while in Ukraine this indicator in 2024 is only 48%.

2 Analysis of expectations from the integration of GAI into the learning process

2.1. Increasing student learning efficiency - USA: 82% (EDUCAUSE Review, 2024); China: 88%; EU: 74%; Ukraine: 59% (IITTA, 2024).

2.2. Improvement of personalized learning trajectories for students - USA: 56%; China: 71%; EU: 63%; Ukraine: 52%.

3 Analysis of the prospects of the impact of GAI on the quality of education

3.1. Integration of GAI into university curricula - USA: 38% (EDUCAUSE Survey, 2025); China: over 60%; EU: 44%; Ukraine: only 22%, however, in 2025/2026 the Ministry of Education and Science plans to publish methodological guidelines on the implementation of GAI for educational institutions.

3.2. Educational effectiveness when using GAI (comparison for 2023-2024) among students in: USA - +12%; China - +18%; EU - +9-11%; Ukraine: pilot projects (KPI, LNU) demonstrated an increase in academic performance of +7-9%.

3.3. Skills market and student employment. LinkedIn Education Analytics predicts that the ability to work with GAI will become one of the top 10 key requirements that increase employment prospects and make a candidate more competitive: USA - 72%, China - 81%, EU - 67%, Ukraine - 54%.

Conclusion

The conducted analysis shows that GAI is rapidly integrating into the global educational space. In 2022-2025, the number of GAI users, particularly among students, increased hundreds of times, demonstrating an unprecedented speed of

adoption of new technologies in higher education and a growing level of digital readiness within the academic community.

The potential benefits of GAI in higher education outweigh the risks, improving learning efficiency and enabling personalized education. The integration of GAI into higher education will lead to the formation of a new model of interaction - “student-GAI-educator,” where the educator becomes a moderator and mentor, while artificial intelligence functions as a personal educational assistant that supports individual learning trajectories.

Summarizing the above, it can be noted that GAI will become a key instrument for the transformation of higher education, increasing academic performance, supporting personalized learning paths, and enhancing educational outcomes.

In October 2025, Ukraine for the first time took part in a high-level meeting of the European AI Board, which represents an important step toward aligning with the standards of the AI Act. Ukraine presented its progress in implementing the White Paper on AI, a roadmap for the AI Act together with EU4Digital, the development of a National AI Strategy until 2030, the launch of an AI assistant in the Diia platform, an AI tool for EU integration, and the creation of a large language model.

Considering this, I hope that the further integration of GAI into the learning process will take place with the support of the state (control, assistance, and guidance), ensuring responsible and safe use of generative artificial intelligence.

References

1. Campus Technology. (2024). *Impact of Generative AI on Student Learning Engagement*. <https://campustechnology.com>
2. Deloitte. (2024). *AI Trust Survey 2024: Privacy and Transparency in the Age of Generative Models*. <https://www.deloitte.com>
3. EDUCAUSE Review. (2024). *Generative AI in Higher Education: Opportunities and Challenges*. <https://er.educause.edu>
4. EDUCAUSE Survey. (2025). *University AI Policies and Implementation*. <https://er.educause.edu>

Maksym Donchenko

First-year Master's student, Faculty of Information Technologies

ORCID: 0009-0004-3084-4834

Andriy Klevny

PhD in Technical Sciences, Assistant, Faculty of Information Technologies

ORCID: 0000-0001-8942-6670

Taras Shevchenko National University of Kyiv

A COMBINED METHOD FOR DETECTING AND ELIMINATING DUPLICATE PUBLICATIONS WHEN AGGREGATING PUBLICATION DATA FROM SCOPUS, WOS, AND ORCID DATABASES

Abstract. The study examines the problem of duplicate publications when integrating data from Scopus, Web of Science (WoS), and ORCID databases. A combined deduplication method is proposed, combining verification of unique identifiers, textual similarity of titles, and prioritization of sources. The method should reduce the number of duplicate records and increase the reliability of aggregated data. The results can be used in analytical systems to form unified author profiles and publication databases.

Keywords: deduplication, data aggregation, Scopus, Web of Science, ORCID, DOI, text similarity, source prioritization, fuzzy comparison, Jaro–Winkler distance.

The automatic aggregation of publication data from various databases (Scopus, WoS, ORCID) provides a comprehensive picture of scientific activity, but the problem of duplicate records accompanies it. A single publication may appear in several sources, which distorts citation metrics and reduces the accuracy of analytics. The proportion of potential duplicates when integrating Scopus and WoS is significant, and ORCID often duplicates records when importing from multiple sources simultaneously [3].

Bramer W.M. et al. [4] described a manual approach to finding duplicates in EndNote. Forbes C. [2] created Deduplicator, an automated tool that utilizes the Jaro–Winkler algorithm, with a primary focus on medical databases such as PubMed. Madsen H.H. et al. [5] emphasize the importance of unique identifiers, but note that even unique identifiers can contain errors or duplicates, so they cannot be considered the only ideal option for deduplication. Kumpulainen M. et al. [6] note in their study the difficulties in combining Scopus and WoS due to differences in metadata formats. The problem of integrating ORCID with external systems also remains relevant, as described in the study by Ivanov S. et al. [7].

To solve the problem of record deduplication, a combined method for identifying and eliminating this issue was developed. This method combines the analysis of unique identifiers, normalized names, and prioritization of sources in records received from ORCID when merging data from Scopus, WoS, and ORCID databases.

The main stages of the developed method:

1. Data normalization. The system converts publication titles to lowercase, removes special characters, and creates standardized DOI fields. Normalization ensures accurate comparison even when metadata is obtained from different sources and has minor differences in spelling due to the presence of special characters or differences in case.

2. Comparison by DOI. If the DOI matches, an additional check of the titles is performed, as the DOI may refer to an entire journal issue or volume.

3. Comparison of all publications, including those remaining after comparison by DOI, by normalized title, using the Jaro-Winkler metric with a similarity threshold of 0.95. The similarity threshold is critical because it prevents the accidental detection of publications with very similar titles, such as two different chapters of a book.

4. Prioritization of sources for ORCID is accepted in the following order: Scopus → WoS → Crossref → others (institutional repositories, DataCite, PubMed, author, etc.).

5. Aggregation of results. The system forms a consolidated record that contains unique attributes (e.g., EID from Scopus, UID from WoS). Records with uncertainty are marked as “requires manual verification.”

The proposed combined method provides multi-level data verification based on unique identifiers, names, and sources. It significantly reduces duplication when combining data from Scopus, WoS, and ORCID, thereby facilitating the work of analysts and improving data quality.

References:

1. Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2024). Analysis of the publication and document types in openalex, web of science, scopus, pubmed and semantic scholar. <https://doi.org/10.48550/arXiv.2406.15154>
2. Forbes, C., Greenwood, H., Carter, M., & Clark, J. (2024). Automation of duplicate record detection for systematic reviews: Deduplicator. *Systematic Reviews*, 13(1). <https://doi.org/10.1186/s13643-024-02619-9>
3. ORCID Support. Group multiple versions of the same work together. <https://support.orcid.org/hc/en-us/articles/360006894774-Group-multiple-versions-of-the-same-work-together>
4. Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA*, 104(3), 240–243. <https://doi.org/10.3163/1536-5050.104.3.014>
5. Madsen, H. H., Madsen, D., & Gauffriau, M. (2016). Evaluation of unique identifiers used as keys to match identical publications in Pure and SciVal – a case study from health science. *F1000Research*, 5, 1539. <https://doi.org/10.12688/f1000research.8913.2>
6. Kumpulainen, M., & Seppänen, M. (2022). Combining Web of Science and Scopus datasets in citation-based literature study. *Scientometrics*. <https://doi.org/10.1007/s11192-022-04475-7>
7. Ivanov, S., Ivohin, E., & Makhno, M. (2024). Automation of accounting of publications using the ORCID application programming interface. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics and Mathematics*, (1), 141–146. <https://doi.org/10.17721/1812-5409.2024/1.26>

Denis Drachuk, 1st year Master's student in Computer Science

Ihor Miroshnychenko, PhD, associate professor of the department of technologies management

Taras Shevchenko National University of Kyiv

OBJECT DETECTION IN ADVERSE WEATHER CONDITIONS: CURRENT CHALLENGES, METHODS, AND FUTURE DIRECTION

Abstract. This study reviews modern approaches to object detection under adverse weather, focusing on challenges from rain, fog, snow, and low illumination. It traces the evolution from classical methods to deep learning architectures like YOLO, Faster R-CNN, and Transformers, along with robustness techniques such as augmentation and domain adaptation. Key issues include domain shift and limited data, with future directions toward lightweight hybrid models and generative augmentation for robust detection in complex environments.

Keywords: object detection, object recognition, adverse weather, difficult weather.

Object detection is a core task in computer vision that combines classification and localization of objects in images or videos [1]-[3]. However, adverse weather conditions such as rain, fog, snow, and low illumination significantly degrade detection performance due to reduced image quality, contrast, and informative features [4]-[8]. Deep learning models like YOLOv4, Faster R-CNN, and Mask R-CNN suffer accuracy drops caused by domain shift between training and test data [9]-[11]. Ensuring robust object detection in such conditions is essential for autonomous driving, surveillance, and intelligent transportation systems [1], [2], [5].

Object detection has evolved from classical algorithms using manually designed features like HOG, SIFT, and Haar-like features to deep learning models that greatly improve accuracy under challenging conditions [2]. Hybrid approaches combining classical and deep learning methods, such as HOG-Bag-of-Visual-Words with Back Propagation Neural Networks, have demonstrated superior performance, achieving 79% accuracy versus 65% for conventional CNNs on degraded images [12]. Modern object detection in adverse weather relies mainly on CNN-based models: two-stage detectors (e.g., Faster R-CNN, Mask R-CNN) and one-stage detectors (e.g., YOLO, SSD); as well as emerging Transformer-based architectures [1], [2], [5], [13], [14]. Two-stage models provide higher accuracy but suffer from slower inference, while one-stage detectors like YOLO provide an optimal balance between speed and precision, achieving up to 97.93% AP on thermal images [5], [8]. Despite their effectiveness, CNN-based models degrade under foggy or rainy conditions due to domain shift, which motivates the use of preprocessing CNNs for deraining and denoising [10], [15]. Recent Transformer and diffusion-based architectures, such as RT-DETR and WeatherDiffusion, further improve detection robustness but demand greater computational resources [16].

To mitigate accuracy degradation under adverse weather, researchers employ data processing, multisensor fusion, and architectural improvements such as training on all-weather datasets (ACDC, DAWN), synthetic augmentation via GANs or diffusion models, and domain adaptation techniques [2], [6], [17], [18]. Enhanced architectures like MFA-YOLO with attention and wavelet-based modules, as well as thermal-image fine-tuned YOLOv3 models, significantly boost detection reliability in fog, rain, and low illumination [16]. Key challenges include poor generalization and domain shift, causing models trained in clear weather to lose accuracy in adverse conditions [17]. The scarcity of large annotated datasets covering diverse weather and lighting scenarios further limits performance [1]. Additionally, preprocessing methods like deraining or dehazing may distort essential features, while low interpretability of deep models hampers their safe use [2], [17].

Future work should address data scarcity, enhance model generalization, and reduce computational complexity. Promising directions include synthetic data generation via generative models and style-transfer augmentations to improve robustness. From an architectural perspective, future research could explore combining CNNs and Transformers with attention mechanisms, creating end-to-end models that jointly perform restoration and detection, and designing lightweight versions suitable for edge devices.

References:

1. Ahmed, Muhammad et al. "Survey and Performance Analysis of Deep Learning Based Object Detection in Challenging Environments." *Sensors (Basel, Switzerland)* 21 (2021): n. pag.
2. Tahir, Noor Ul Ain et al. "Object Detection in Autonomous Vehicles under Adverse Weather: A Review of Traditional and Deep Learning Approaches." *Algorithms* 17 (2024): 103.
3. Oksuz, Kemal et al. "Imbalance Problems in Object Detection: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2019): 3388-3415.
4. Liu, Wenyu et al. "Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions." *AAAI Conference on Artificial Intelligence* (2021).
5. Krišto, Mate et al. "Thermal Object Detection in Difficult Weather Conditions Using YOLO." *IEEE Access* 8 (2020): 125459-125476.
6. Hnewa, Mazin and Hayder Radha. "Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques." *IEEE Signal Processing Magazine* 38 (2020): 53-67.
7. Walambe, Rahee et al. "Lightweight Object Detection Ensemble Framework for Autonomous Vehicles in Challenging Weather Conditions." *Computational Intelligence and Neuroscience* 2021 (2021): n. pag.
8. Wang, Kunyi and Yaohua Zhao. "Improving object detection in challenging weather for autonomous driving via adversarial image translation." *PLOS One* 20 (2025): n. pag.
9. Xu, Lida. "Data Shift of Object Detection in Autonomous Driving." *ArXiv* abs/2508.11868 (2025): n. pag.

10. Li, Jinlong et al. "Domain Adaptation Based Object Detection for Autonomous Driving in Foggy and Rainy Weather." *IEEE Transactions on Intelligent Vehicles* 10 (2023): 900-911.
11. Rothmeier, Thomas et al. "I Had a Bad Day: Challenges of Object Detection in Bad Visibility Conditions." *2023 IEEE Intelligent Vehicles Symposium (IV)* (2023): 1-6.
12. Osipov, Aleksey et al. "Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions." *Sustainability* (2022): n. pag.
13. Bilous, Nataliya et al. "Comparison of CNN-Based Architectures for Detection of Different Object Classes." *AI* (2024): n. pag.
14. Pfeuffer, Andreas and Klaus C. J. Dietmayer. "Optimal Sensor Data Fusion Architecture for Object Detection in Adverse Weather Conditions." *2018 21st International Conference on Information Fusion (FUSION)* (2018): 1-8.
15. Wang, Kaige et al. "Image Deraining and Denoising Convolutional Neural Network For Autonomous Driving." *2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)* (2021): 241-245.
16. Li, Zhongyang and Haiju Fan. "MFA-YOLO: Multi-Scale Fusion and Attention-Based Object Detection for Autonomous Driving in Extreme Weather." *Electronics* (2025): n. pag.
17. Gupta, Himanshu et al. "Robust Object Detection in Challenging Weather Conditions." *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2024): 7508-7517.
18. Gurbindo, Unai et al. "Object detection in adverse weather conditions for autonomous vehicles using Instruct Pix2Pix." *ArXiv abs/2505.08228* (2025): n. pag.

Illia Harbar

2nd year Master's student, Department of Software Systems and Technologies

Kseniia Dukhnovska

Ph.D., Associate Professor, Department of Software Systems and Technologies

Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

A HYBRID ONLINE-OFFLINE SYSTEM FOR ROBUST SPEECH RECOGNITION AND TRANSLATION OF KOREAN AUDIO

In an increasingly interconnected world, the demand for effective cross-lingual communication tools has surged. Automated audio translators, capable of processing speech and rendering it in another language, are becoming indispensable in sectors such as education, international business, and tourism.¹ However, the development of such systems presents significant technical hurdles, particularly when dealing with languages that possess unique linguistic characteristics. The effectiveness of these systems is often constrained by real-world conditions like background noise, regional accents, and dialects, which can severely degrade performance.¹

This research focuses on the specific challenges posed by the Korean language. Korean presents a formidable test for standard audio processing pipelines due to its complex phonetic system, where the pronunciation of Hangeul characters can change based on context, and its unique grammatical structures, such as a sophisticated hierarchy of politeness (honorifics) that must be correctly interpreted to preserve meaning [1]. Existing solutions often force users into a trade-off: leverage powerful, cloud-based online services that offer high accuracy but require constant internet connectivity and involve transmitting potentially sensitive data to third-party servers, or use less accurate offline tools that guarantee privacy and availability.

This paper introduces a system designed to resolve this dilemma. The primary scientific contribution is the design, implementation, and evaluation of a novel software architecture that comprehensively integrates offline and online models for both Automatic Speech Recognition (ASR) and Machine Translation (MT) [2]. As stated in the foundational research for this work, this represents the first comprehensive integration of an offline speech recognition library (Vosk), a large-scale neural network for ASR (Whisper), and a heterogeneous ensemble of machine translation systems (Google Translate and MarianMT) [3-5]. This hybrid approach provides users with the flexibility to select the optimal processing mode based on their specific needs, whether prioritizing the highest accuracy in a public setting or ensuring data confidentiality in a private business meeting. The system demonstrates that by strategically combining the strengths of different models, it is possible to create a tool that is not only highly accurate but also adaptable to diverse user constraints and environments.

System Architecture and Methodology. The system is built upon a modular software architecture, a core principle of robust software engineering that ensures component independence, maintainability, and scalability. This architecture

comprises four distinct, interconnected modules: Audio Pre-processing, Hybrid ASR, Ensemble MT, and the User Interface.

To ensure optimal and consistent performance of the downstream recognition models, all input audio undergoes a standardized pre-processing pipeline implemented using the librosa library. This critical first step is designed to normalize the audio signal and mitigate the impact of environmental noise.¹ The pipeline consists of three stages:

1. **Standardization:** Input audio files, accepted in WAV or MP3 formats, are first resampled to a uniform sampling rate of 16 kHz. Signal amplitude is then normalized to a range of $[-1, 1]$. This stage ensures that the ASR models receive data in a consistent format, regardless of the source recording's characteristics.

2. **Noise Reduction:** A spectral subtraction method, `librosa.decompose.decompose`, is applied to reduce ambient background noise. This algorithm analyzes the signal's spectrogram, identifies persistent noise components, and subtracts them, resulting in a noise reduction of up to 10 dB. This step is crucial for enhancing recognition accuracy in real-world scenarios such as public lectures or outdoor recordings.¹

3. **Segmentation:** The cleaned audio signal is segmented into 20 ms frames with a 10 ms overlap. This framing is a standard technique in speech processing that breaks the continuous signal into short, quasi-stationary segments suitable for feature extraction and analysis by the ASR models.¹

The core of the speech-to-text functionality lies in the hybrid ASR module, which provides the user with a choice between two distinct models. The fundamental task of ASR is to find the most probable sequence of words T given an input audio signal A , which can be formally expressed as:

$$T^* = \underset{T}{\operatorname{argmax}} P(T|A)$$

The system leverages two models to solve this problem, each with unique strengths:

- **Vosk:** This component serves as the primary offline recognition engine. Based on the Kaldi speech recognition toolkit, it employs a Hidden Markov Model-Deep Neural Network (HMM-DNN) architecture. The Vosk model used in this system is specifically trained and optimized for the Korean language, making it highly efficient and accurate for this task in environments without internet access. Its lower computational footprint also makes it suitable for devices with limited resources.¹

- **Whisper:** Developed by OpenAI, Whisper is a state-of-the-art ASR model based on a Transformer architecture. Its training on a massive and diverse dataset grants it exceptional robustness to background noise, accents, and dialectal variations. While more computationally intensive, Whisper provides superior accuracy, especially in challenging audio conditions, and can be run either locally (offline) or via an API (online).¹

This dual-model design empowers the user to make a context-aware choice: selecting Vosk for fast, private, offline transcription or Whisper for maximum

accuracy in noisy environments.

The translation module is designed not just to convert text but to do so with a high degree of linguistic fidelity, particularly for the nuances of Korean. This is achieved by moving beyond a single-model approach and implementing a weighted ensemble. This method can be conceptualized as combining a powerful "generalist" model with a "specialist" model fine-tuned for specific linguistic challenges. The generalist provides broad coverage, while the specialist acts as a corrective mechanism, improving the handling of domain-specific or complex grammatical structures like Korean honorifics.¹

The ensemble integrates two models:

- Google Translate API: A leading online translation service known for its high performance across a vast number of language pairs. It serves as the powerful generalist baseline in the ensemble.
- MarianMT: An efficient, open-source Neural Machine Translation framework based on the Transformer architecture, developed by the Helsinki-NLP group. This model, which can be run offline, is highly effective for specific language pairs and serves as the specialist component, offering strong performance on Korean grammatical structures [6-7].

The outputs of these two models are combined using a weighted scoring function. The system seeks to find the optimal translated text T' that maximizes a combined probability score. The score for a candidate translation T' given a source text T is calculated as:

$$Score(T') = 0.6 \cdot P_{\text{Google}}(T'|T) + 0.4 \cdot P_{\text{MarianMT}}(T'|T)$$

The weights, $\omega_1=0.6$ for Google Translate and $\omega_2=0.4$ for MarianMT, were determined empirically on a validation dataset to maximize the final BLEU score. This configuration leverages the broad accuracy of Google's service while incorporating significant corrective input from MarianMT to enhance the overall quality and contextual appropriateness of the final translation [8-9].

Conclusion and Future Work. The research successfully demonstrated that by strategically integrating specialized offline models (Vosk, MarianMT) with powerful, noise-robust online services (Whisper, Google Translate), it is possible to create a flexible and high-performance translation tool. The system achieved a state-of-the-art Word Error Rate of 5% and a Bilingual Evaluation Understudy score of 0.85, validating its effectiveness. The modular architecture provides a user-centric solution that balances the competing demands of accuracy, privacy, and connectivity, making it a practical tool for applications in education, business, and tourism.

Building on this successful foundation, future work will focus on addressing the system's current limitations and expanding its capabilities. The following specific directions are proposed :

- To mitigate the high error rates on regional dialects, the immediate priority is to fine-tune the ASR models. This will involve augmenting the training data with a large and diverse corpus of Korean dialects, aiming to significantly reduce the WER gap between standard and non-standard speech.

- The architectural framework is language-agnostic. Future iterations will expand support to include other phonetically complex languages, applying the same hybrid and ensemble principles.

To broaden the system's impact, two key development paths will be pursued. First, a public API will be created to allow for seamless integration into third-party applications and educational platforms like Moodle. Second, a dedicated mobile application will be developed, primarily targeting the tourism use case and offering a convenient, on-the-go translation solution.

References

1. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. In *Advances in Neural Information Processing Systems*, 30.
2. Radford, A., Kim, J. W., Xu, T., et al. (2023). Robust Speech Recognition via Large-Scale Weak Supervision. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*.
3. Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*.
4. Povey, D., Ghoshal, A., Boulianne, G., et al. (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.
5. Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., et al. (2018). Marian: Fast Neural Machine Translation in C++. In *Proceedings of ACL 2018, System Demonstrations*.
6. Vosk Speech Recognition Toolkit. (2025). Retrieved from <https://alphacephei.com/vosk/>
7. OpenAI. (2023). Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. Retrieved from <https://openai.com/research/whisper>
8. Google Cloud Translation API Documentation. (2025). Retrieved from <https://cloud.google.com/translate/docs>

Dariia Herasymchuk

4th year bachelor of computer science student

Iryna Domanetska

Ph.D., associate professor, associate professor at the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

HYBRID APPROACHES TO COVID-19 EPIDEMIC FORECASTING BASED ON MECHANISTIC AND NEURAL NETWORK MODELS

Abstract. The study examines modern approaches to forecasting the course of the COVID-19 epidemic. It compares classical statistical methods (ARIMA), compartmental epidemiological models (SEIR), modern machine learning techniques (LSTM, Bi-LSTM), and hybrid architectures (ARIMA-LSTM, SEIR+LSTM). Special attention is paid to the advantages and limitations of each group of models in terms of interpretability, consideration of nonlinear factors (vaccination, mutations, population mobility), and forecast accuracy. The comparative analysis showed that the hybrid SEIR+LSTM model combines the high interpretability of the mechanistic approach with the adaptability of neural networks to nonlinear data dynamics, leading to improved prediction accuracy. This approach is promising for decision support systems in the healthcare sector.

Keywords: COVID-19, forecasting, ARIMA, SEIR, LSTM, hybrid model, time-varying parameters, machine learning.

Forecasting the course of pandemics, particularly COVID-19, is critically important for strategic planning of medical resources, risk assessment, and the development of adaptive anti-epidemic measures. Given the high variability of data and rapid virus evolution, classical models require the complement of flexible machine learning algorithms. Effective pandemic forecasting demands models capable of capturing both linear and nonlinear patterns of disease spread.

Existing scientific studies on COVID-19 forecasting demonstrate the evolution from classical statistical models to hybrid intelligent systems capable of accounting for the complex dynamics of a pandemic. The most effective have proven to be models that combine mechanistic approaches with neural network algorithms.

Classical statistical methods, such as ARIMA, provide high interpretability and computational speed but fail to account for nonlinear fluctuations caused by new virus strains or quarantine measures [1].

Compartmental models such as SEIR allow the description of the main mechanisms of infection transmission through systems of differential equations [2]. Despite the transparent parameter structure (β , γ , σ), these models are sensitive to initial conditions and fail to account for behavioral factors [3]. Modern neural approaches, such as LSTM and Bi-LSTM, show higher accuracy due to their ability to model long-term dependencies in time series but lag behind classical models in interpretability [4].

Further improvement of LSTM architectures—through hyperparameter optimization and dropout regularization—enhances model stability and generalization ability.

A promising direction is hybrid modeling, which combines mechanistic and ML-based approaches.

In particular, ARIMA-LSTM integrates the linear forecasts of ARIMA with nonlinear residual correction using LSTM [5], while the SEIR+LSTM model enables dynamic updates of parameters $\beta(t)$, σ , $\gamma(t)$, predicted by the neural network [6]. This combination allows the model to account for changes in mobility, vaccination levels, and virus mutations.

Experimental results indicate that SEIR+LSTM achieves the best balance of accuracy, interpretability, and flexibility, outperforming individual models by 15–25% in forecast accuracy [6]. Therefore, hybrid approaches based on SEIR and LSTM can serve as the foundation for developing intelligent decision support systems in epidemiological monitoring.

Hybrid models, despite their effectiveness, are accompanied by a number of challenges that must be considered during their development and implementation: the integration of structurally heterogeneous approaches, the risk of overfitting and loss of interpretability, and the complexity of neural network hyperparameter selection. These issues can be mitigated through modular architecture design, the use of regularization techniques, and the implementation of explainable mechanisms.

Hybrid models demonstrate potential for building decision support systems where mechanistic components provide explainability, and neural components ensure adaptability to real-world data. This approach enables the integration of epidemiological models with UX-oriented interfaces for medical analysts, ensuring transparency and flexibility in scenario modeling.

References

1. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. URL: <https://otexts.com/fpp3>
2. Eddin, M. S., et al. (2025). *Systematic Comparison of Different Compartmental Models for Predicting COVID-19 Progression..* DOI: 10.3390/epidemiologia6030033
3. Zhao, W. et al. (2022). *Prediction of COVID-19 Data Using Hybrid Modeling. Frontiers in Public Health.* DOI: 10.3389/fpubh.2022.923978
4. Shao, Y., Wan, T. K., Chan, K. H. K., et al. (2025). *Prediction of COVID-19 cases by multifactor-driven long short-term memory (LSTM) model.* DOI: 10.1038/s41598-025-86698-1
5. Mahmud, A., Syed Husni Noor Syed Hatim Noor, K. I. Musa, F. M. Hamzah, Z. M. Yudin, N. Kamaruddin, A. M. Madawana, & M. A. Awang Nawi (2025). *Hybrid ARIMA-LSTM for COVID-19 forecasting: a comparative AI modeling study.* DOI: 10.7717/peerj-cs.3195
6. Feng, L., Chen, Z., Lay Jr., H. A., Furati, K., & Khaliq, A. (2022). *Data-driven time-varying SEIR-LSTM/GRU algorithms to track the spread of COVID-19.* 19(9):8935–8962. DOI: 10.3934/mbe.2022415

Vladyslav Hnatiienko

2nd-year PhD Student, Department of Information Systems and Technologies

Taras Shevchenko National University of Kyiv, Faculty of Information Technology

CONCEPT OF A SPATIO-TEMPORAL DATA RECOVERY METHOD FOR SATELLITE IMAGERY DISTORTED BY CLOUD COVER

Abstract. The paper addresses the loss of up to 90% of crop-related satellite data due to cloud cover, which limits yield prediction accuracy. A spatio-temporal recovery method is proposed based on the principles that nearby plots and close-in-time observations have similar vegetation indices. The method combines spatial V_p and temporal V_t recovery with an optimally selected weight α , improving data completeness and prediction stability. The results enable fuller use of satellite imagery and open prospects for automated hyperparameter tuning and scalable recovery across different crops and conditions.

Keywords. Satellite data recovery, satellite imagery recovery method, spatio-temporal data recovery method.

Introduction. Previous studies have shown that satellite imagery enables highly accurate prediction of agricultural crop yields. In addition, obtaining such data is significantly simpler than acquiring UAV imagery or field sensor measurements. When using local observation methods, the main limitation is the need for an individual approach to each field, which results in additional financial costs. In contrast, a single satellite allows for daily monitoring of multiple fields, and modern technologies make it possible to acquire detailed data at a much lower cost than with local observations.

However, a major limitation of satellite monitoring is data loss caused by cloud cover, which obscures observed areas [1]. As much as 90% of the data may be lost, since clear days are rare during the crop maturation period. Accurate yield prediction requires a high degree of correspondence between sensor-recorded indicators and the actual biological parameters of plants in each field segment. Currently, distorted satellite images are mostly excluded from analysis, and models are built on limited datasets. This approach allows for detecting some partial dependencies and achieving prediction accuracy of up to 96%, but in many cases, the models lack reliability, as the data may not align temporally with those used during training; moreover, observations at unrecorded time points may differ significantly even when the indicators in clear-sky conditions match [2]. The discrepancy problem could potentially be mitigated by augmenting the training set with data from hundreds or even thousands of fields, but this would require substantial resources. At the same time, it is unlikely that uncertainty in unobserved time points can be fully resolved merely through dataset expansion or model refinement.

Concept of the Data Recovery Method. Satellite images discarded due to even minor distortions often contain a substantial amount of additional information that can be used to improve the correspondence between the actual state of vegetation in the field and its representation in the data. Therefore, to increase the informativeness of the input data, it is necessary to utilize all images captured during the crop maturation period by restoring all incorrectly recorded measurements. In [3], a method is presented for

recognizing image fragments distorted by clouds. The next step after recognition is to reconstruct the missing data based on the valid elements in the dataset.

We introduce two heuristics regarding the reduction of differences in vegetation index values (indicators derived from satellite data that describe vegetation status) with decreasing observation interval.

Heuristic 1. The difference in vegetation index values is smaller for spatially closer field segments than for more distant ones.

Heuristic 2. The difference in vegetation index values is smaller for observations that are temporally closer than for those further apart in time.

Each plot in the training set is represented by a parameter vector $d_{x_n, y_n}(t)$ (where n is the plot index) for each time moment t , which describes the vegetation state on that plot. At time t , this vector is an element of the satellite image captured for the entire area of observation. We introduce the spatial recovery operation V_p to estimate the corrected parameter values for plot n at time t : $\hat{d}_{x_n, y_n}(t) = V_p(x_n, y_n, t)$, where the valid elements of the satellite image at time t are used for the computation. The operation V_p can be applied to elements that have spatially adjacent plots with valid data. After recovery, the element is considered valid.

At the same time, the parameter vector of each plot forms a time series: for each day of the growing season $t \in 0, 1, \dots, 100$ (assuming the season lasts 100 days), the value $d_{x_n, y_n}(t)$ is recorded for plot n . We introduce the temporal recovery operation V_t to estimate the corrected parameter values for plot n at time t : $\hat{d}_{x_n, y_n}(t) = V_t(x_n, y_n, t)$, where valid elements of the time series for the same plot are used for the computation. Similar to V_p , the operation V_t can be applied to elements adjacent to valid observations, and after recovery, the element is also considered valid.

Both operations may be performed using different methods (e.g., local interpolation, nearest neighbors, weighted averaging, etc.). In addition, recovery for each element can be performed using V_p , V_t , or their weighted combination:

$\hat{d}_{x_n, y_n}(t) = \alpha V_p(x_n, y_n, t) + (1 - \alpha) V_t(x_n, y_n, t)$, $\alpha \in [0, 1]$ — the recovered value of the vegetation index for plot n at time t .

A schematic example of this recovery process based on neighboring elements is shown in Figure 1.

Depending on the volume and spatial distribution of valid data, different recovery configurations may be applied. The main objective is to improve prediction accuracy; therefore, the recovery parameters are selected by solving the optimization problem:

$$\min_{\alpha \in [0, 1], \mathcal{M}_p, \mathcal{M}_t} \|F(d_{x_n, y_n}^j - Y_{x_n, y_n})\|,$$

where d_{x_n, y_n}^j is the time series whose elements are valid or restored values of the indicators describing the state of plot n at each time moment $t \in 0, 1, \dots, 100$;

$\mathcal{M}_p$ and $\mathcal{M}_t$ are the selected methods of spatial and temporal recovery, respectively.

For each distorted element in the dataset, we introduce characteristics $\chi_{x_n, y_n}(t)$, which define its relation to other valid elements and serve as the information basis for selecting the optimal value of α and the methods of spatial and temporal recovery $\mathcal{M}_p$ and $\mathcal{M}_t$. Thus, the task is to determine such optimal configurations that enable the formulation of a general rule for selecting recovery hyperparameters in any case, with the goal of maximizing the accuracy of yield prediction.

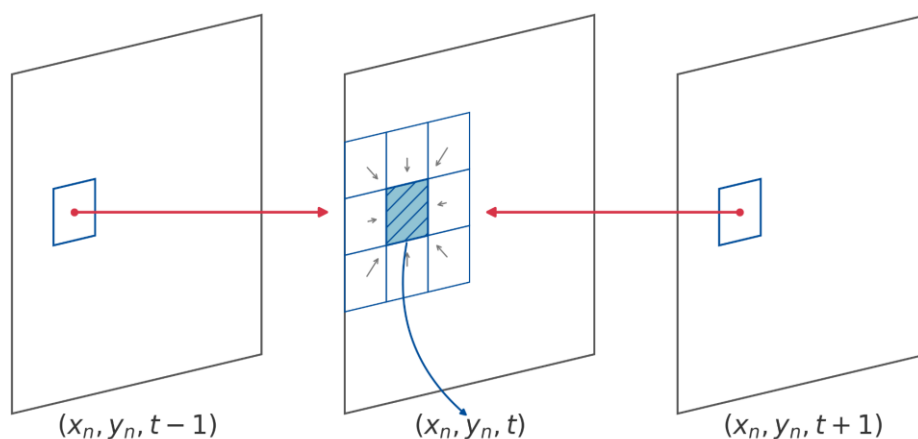


Figure 1. Schematic illustration of the spatio-temporal recovery process: the value at location (x_n, y_n, t) is reconstructed using spatial neighbors from the same image and temporal neighbors from adjacent time steps.

Conclusions and Future Work. A concept has been proposed for a method of restoring satellite image elements based on valid observations. Further development stages include constructing a dataset for experimental verification and tuning the hyperparameters. Additionally, methods for performing the operations V_p and V_t will be evaluated, and a set of the most effective ones will be defined for selection during the data recovery process.

References:

1. Vladyslav Hnatienko, Vitaliy Snytyuk. Site-specific forecasting of agricultural crop yield as a technology and service // Proceedings of the 8th International Scientific and Practical Conference "Applied Information Systems and Technologies in the Digital Society" (AISTDS 2024), Kyiv, Ukraine, October 1, 2024. CEUR Workshop Proceedings, Vol. 3942.
2. V.H. Hnatienko, H.M. Hnatienko, O.L. Zozulya, V. Ye. Snytyuk, V.V. Schwartau. Site-Specific sunflower yield forecasting based on spatial analysis and machine learning // Reports of the National Academy of Sciences of Ukraine. 2025. – P. 17–26. <https://doi.org/10.15407/dopovidi2025.04.017>.
3. Гнатієнко В.Г., Зозуля О.Л., Снитюк В.С. Концептуальні елементи технології ідентифікації супутникових показників в умовах захмареності знімків // Матеріали 2-ї Міжнародної науково-практичної конференції «Інформаційні системи та технології: результати і перспективи» (IST 2025), Київ, Україна, 5 березня 2025 р. – Київ: ФІТ КНУТШ, 2025. – С. 312–315.

Veronika Horbach

Student

Victoria Volkogon

Ph.D. in Technical Sciences

State University «Kyiv Aviation Institute»

APPLICATION OF A CUSTOM-TUNED RANKING ENGINE FOR MULTI-FACTOR USER PROFILE ANALYSIS

Abstract. The conducted work addresses semantic matching for complex user profiles in personalization and recommendation systems. It proposes a customized deterministic ranking mechanism (e.g., based on Elasticsearch technology) as an alternative to keyword search. The main idea is implementing a weighted scoring algorithm for multi-dimensional profile analysis considering tags, contextual parameters, and user preferences. This approach ensures increased transparency, interpretability, and relevance, crucial for intelligent decision support systems.

Keywords: ranking algorithm, weighted scoring, profile analysis, Elasticsearch, semantic matching, rule-based system, expert system.

Introduction

Modern information systems are increasingly faced with the task of matching complex, multi-dimensional profiles. Traditional methods, based on rigid keyword search, are often ineffective. They fail to place the most relevant documents at the top of the list [1] and do not account for the relative importance of different criteria (e.g., a "mandatory" requirement versus a "desirable" one).

This creates a need for methods that enable semantic matching while remaining transparent and controllable, similar to the principles of classical expert systems [2].

Problem Statement

The purpose of this work is to research and describe the application of a custom-tuned search engine to solve the problem of multi-factor profile matching. The main task is to describe a method for applying a deterministic weighted scoring algorithm. This algorithm must analyze a structured user profile (skill sets, tags, preferences) and match it against a set of requirements, providing a transparent and relevant compliance score.

Methodology: Application of a Custom Ranking Engine

The proposed method focuses not on developing a new engine, but on the specific application of existing high-performance search systems, such as Elasticsearch [3]. The method is based on two key concepts:

1. **User-Managed Profile Model:** The method assumes that a user's profile consists not only of unstructured text but also of a set of structured tags. A key feature is that the user can independently manage these tags and categorize them by their level of importance (e.g., "mandatory criteria" and "desirable criteria").
2. **Application of a Weighted Ranking Algorithm:** Instead of a simple count of matches, it is proposed to use the advanced capabilities of the search query

(Query DSL) [4]. The essence of "custom-tuning" lies in the application of a weighted scoring algorithm. This is achieved by assigning different numerical "weights" to different criteria:

- A match with a tag in the "mandatory" category provides a high multiplier to the final score.
- A match with a "desirable" tag provides a smaller weight increment.
- Keywords found in descriptive fields may carry less weight than explicit tags.

Thus, the ranking logic is clearly defined within the query itself. This approach transforms the search engine into a transparent, rule-based mechanism that provides the computational mechanisms found in most expert systems [5].

Discussion of the Approach

This application of a search engine has significant advantages:

- **Transparency (Explainability):** Unlike ML models, the logic is deterministic. The system enables a consistent explanation why one profile received a higher score.
- **User Control:** The ranking directly reflects the preferences that the user has consciously set in their profile.
- **No "Cold Start" Problem:** The method operates without a preliminary training phase or requiring prior data collection.
- **Flexibility:** The matching logic (weighting coefficients) can be changed directly at the API level, adapting it to new requirements.

Conclusions

The presented work describes a method for applying a custom-tuned ranking engine to the task of multi-factor profile analysis. The proposed approach, which uses user-managed tags and a deterministic weighted scoring algorithm, is a transparent, flexible, and effective alternative to traditional search methods. The described methodology is relevant for any information system that requires accurate and interpretable matching of complex entities.

References:

1. T. M. Parks, Search Engines, in: H. Bidgoli (Ed.), Encyclopedia of Information Systems, Academic Press, New York, NY, 2003, pp. 23–29.
2. Y. Shang, Expert Systems, in: W. Chen (Ed.), The Electrical Engineering Handbook, Academic Press, Burlington, MA, 2005, pp. 367-377.
3. Elastic, Elasticsearch, 2025. URL: <https://www.elastic.co/elasticsearch>.
4. Elastic, Query DSL, 2025. URL: <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html>.
5. F. J. Kurfess, Artificial Intelligence, in: R. A. Meyers (Ed.), Encyclopedia of Physical Science and Technology, 3rd. ed., Academic Press, New York, NY, 2003, pp. 609-629.

Demyd Hredsov

Master's Degree Field of Study "Information technologies" National Technical University "Kharkiv Polytechnic institute".

Olena Reshetnyak

Dr. Sc. (Economics), Assoc. Prof., Head of Industrial Policy & Innovation Development Sector, Dept. of Industrial Policy & Energy Security, Research Centre for Industrial Development Problems, NAS of Ukraine.

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE IN MANAGEMENT AND PUBLIC ADMINISTRATION: SOFTWARE ENGINEERING APPROACHES FOR DECISION-MAKING AND DIGITAL TRANSFORMATION

Abstract. This paper explores the integration of Artificial Intelligence and Data Science technologies in management and public administration. The role of software engineering is emphasized as the foundation for developing reliable and transparent digital solutions. The proposed PRISM framework covers the full data cycle – from collection to monitoring outcomes. Key applications include strategic planning, process optimization, risk management, transparency, and trust enhancement. Expected results involve improved efficiency, decision quality, and institutional resilience in the digital economy.

Keywords: Artificial Intelligence, Data Science, Management, Public Administration, Software Engineering, Decision Support, Digital Transformation, MLOps, Risk Management, Strategic Planning.

Relevance

Organizations and public institutions alike are undergoing rapid digital transformation. The complexity of socio-economic processes, coupled with the explosive growth of data, requires decision-making tools that exceed the capabilities of traditional management and administrative methods. Artificial Intelligence (AI) and Data Science technologies, supported by modern software engineering practices, enable evidence-based decision-making, predictive analysis, and automation of both business and governmental processes. This creates new opportunities for efficiency, transparency, and sustainable development.

Purpose and Contribution

The purpose of this research is to substantiate the role of AI and Data Science in enhancing decision-making processes in both management and public administration, while emphasizing software engineering as the foundation for reliable, scalable, and accountable digital solutions. The study contributes by:

1. Introducing the PRISM Framework (Policy- and Result-Oriented Insight Scoring & MLOps), applicable to organizational management and public administration;
2. Identifying shared areas of application – strategic planning, resource allocation, risk management, customer/citizen engagement, and monitoring policy or business outcomes;

3. Proposing accountability mechanisms for AI models that ensure transparency and trust in managerial and administrative contexts.

Methodology (PRISM Framework)

Pipelines: Integration of heterogeneous data sources (financial systems, open data, citizen feedback, IoT sensors, HR and CRM platforms).

Risk & Rights: Assessment of business and policy risks, protection of citizens and customers rights, compliance with regulations.

Insight Scoring: Evaluation of analytical results by accuracy, relevance, and value creation.

Safeguards: MLOps practices ensuring reproducibility, monitoring of model drift, and logging of decisions.

Monitoring: Continuous evaluation of managerial and administrative interventions via KPIs, dashboards, and controlled experiments.

Applications

Strategic and Policy Planning: AI supports forecasting in both business development and public policy.

Operational Efficiency: Optimization of workflows, supply chains, and administrative services through predictive analytics.

Engagement and Trust: NLP tools analyze customer needs and citizen feedback, enhancing service personalization and policy responsiveness.

Risk Management: Early detection of financial, operational, and social risks to improve resilience and governance.

Expected Outcomes

Enhanced efficiency through automation and real-time analytics;

Higher quality of decisions thanks to predictive insights and scenario modeling;

Greater transparency and accountability of managerial and administrative processes;

Strengthened competitiveness and trust in organizations and institutions through innovative, data-driven strategies.

Limitations and Future Research

The main challenges involve fragmented data ecosystems, skills gaps among managers and civil servants, and ethical concerns regarding automated decision-making. Future research should focus on causal inference models for complex decision-making, interdisciplinary training programs, and scalable frameworks that bridge business and public sector needs.

Conclusions

The integration of Artificial Intelligence and Data Science into both management and public administration represents a transformative pathway toward

evidence-based, adaptive, and citizen/customer-oriented decision-making. Software engineering provides the technological foundation for developing robust, secure, and transparent solutions. As a result, organizations and public institutions alike can enhance efficiency, accountability, and resilience in the digital economy.

References:

1. Brynjolfsson, E., & McAfee, A. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W. W. Norton & Company.
2. Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, 96(1), 108–116.
3. OECD (2021). *The OECD Digital Government Policy Framework: Six Dimensions of a Digital Government*. OECD Publishing.
4. European Commission (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Brussels.
5. Tapscott, D., & Tapscott, A. (2016). *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Penguin.
6. *Digital Transformation in Ukraine: Challenges and Prospects (2022)*. Analytical Report of the National Institute for Strategic Studies.
7. Cabinet of Ministers of Ukraine (2021). *Concept of Artificial Intelligence Development in Ukraine*. Kyiv.

Inna Hrynishak

Bachelor in Computer Science, Master Student, Faculty of Information Technologies

Hanna Krasovska

PhD in Technical Sciences, Associate Professor, Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv, inna_grynishak@knu.ua

INTELLIGENT MULTI-AGENT SYSTEM ARCHITECTURE FOR DEEPFAKE DETECTION

Abstract. This research presents an improved deepfake detection system, expanding upon our previous hybrid CNN-LSTM approach[1], which achieved 85% accuracy in video-based detection. By implementing a modular Multi-Agent System (MAS), where each agent is responsible for a specific task — such as data collection, preprocessing, visual and temporal analysis, and decision-making — we significantly enhance both detection accuracy and system efficiency. The redesigned architecture reaches 94% accuracy for video-based detection and improves processing speed by over 36% compared to the monolithic CNN-LSTM model. The system's scalability and real-time capabilities make it a powerful solution for critical applications in media verification, cybersecurity, and digital forensics.

Keywords: deepfake detection, multi-agent system (MAS), CNN-LSTM hybrid, modular architecture, temporal analysis, real-time detection

1. Introduction

The proliferation of deepfakes — synthetic media generated through deep learning — continues to erode trust in digital content. These hyper-realistic manipulations are increasingly employed in disinformation campaigns, financial fraud schemes, and reputational attacks. According to recent analytics [2], the volume of deepfake content online has surged more than tenfold from 2018 to 2024, with growth expected to accelerate in 2025. This trend underscores the widespread availability of generative AI tools and the ease with which even non-experts can create convincing fake media.

Recent reports [3] from 2024 and early 2025 confirm that this trend persists, highlighting the increasing use of deepfakes in financial fraud, political disinformation, and social engineering.

Our previous work introduced a hybrid CNN-LSTM model capable of identifying spatial and temporal inconsistencies in video data. However, limitations in speed and scalability prompted the development of a more advanced architecture.

This paper introduces a Multi-Agent System (MAS)[3] for deepfake detection that builds upon the CNN-LSTM pipeline and decomposes functionality into autonomous, cooperating agents. This redesign significantly improves the system's flexibility, robustness, and performance.

2. Methodology

This study builds upon the previously developed hybrid CNN-LSTM architecture for deepfake detection, which integrates spatial analysis through Convolutional Neural Networks (CNNs) and temporal modeling via Long Short-

Term Memory (LSTM) networks [1]. CNNs effectively extract visual cues such as facial inconsistencies, compression artifacts, and abnormal textures from individual frames, while LSTMs analyze frame sequences to detect unnatural temporal transitions, including asynchronous lip movements and blinking patterns.

To enhance the model's performance, scalability, and accuracy, a Multi-Agent System (MAS) architecture was introduced. Instead of relying on a monolithic pipeline, the updated system distributes responsibilities among a set of autonomous, task-specific agents. This approach enables the segmentation of long videos into smaller fragments for independent and parallel processing, which not only accelerates the analysis but also improves the accuracy of fake detection and increases the system's robustness against local errors by allowing more detailed examination of individual video parts.

The multi-agent deepfake detection system is built according to a hierarchical principle and includes the following key agents: Internet Retrieval Agent, Preprocessing Agent, Parallelization Agent, Visual Analysis Agent, Temporal Interpretation Agent, Aggregation Agent, Decision-Making Agent, Result Interpretation Agent. The logical architecture of the system is presented in Figure 1.

The detection process begins with agents responsible for user interaction and external data retrieval, which handle inputs such as uploaded files or links to online media. These are followed by preprocessing agent that perform face detection, frame normalization, and segmentation of the video into smaller fragments for parallel analysis. This step is crucial, as it allows the system to reduce the computational burden on individual processing units and improves the overall responsiveness of the detection pipeline. By breaking the video into smaller, manageable segments, the system can process multiple fragments simultaneously, leading to faster inference times and more detailed analysis — especially in long or high-resolution videos.

The spatial characteristics of each fragment are then extracted by agents based on CNN architectures, while temporal dependencies are modeled by agents using LSTM-based interpretation. Intermediate results are collected and unified by an aggregation agent, which passes the synthesized data to a decision-making agent responsible for determining the authenticity of the content. Finally, the system provides interpretable outputs to the user, aided by visualization agents that highlight suspicious regions in the analyzed frames.

The overall workflow of agent interactions and data exchange is illustrated in Figure 1. This sequence diagram captures the asynchronous yet coordinated behavior of agents in the MAS-based architecture, demonstrating how detection is carried out through parallelized and modular stages. It shows both system entry points — through direct upload and through web-based retrieval — and highlights the collaboration among specialized agents throughout the detection pipeline.

By distributing responsibilities across multiple intelligent agents, the system achieves high concurrency, reduced latency, and greater interpretability. These improvements result from the ability to execute tasks simultaneously, reduce processing delays, and provide clear insight into intermediate steps of analysis. Furthermore, the modular agent-based structure enables independent testing and

updating of components, making the system adaptable to emerging deepfake generation techniques and suitable for real-time deployment across various platforms.

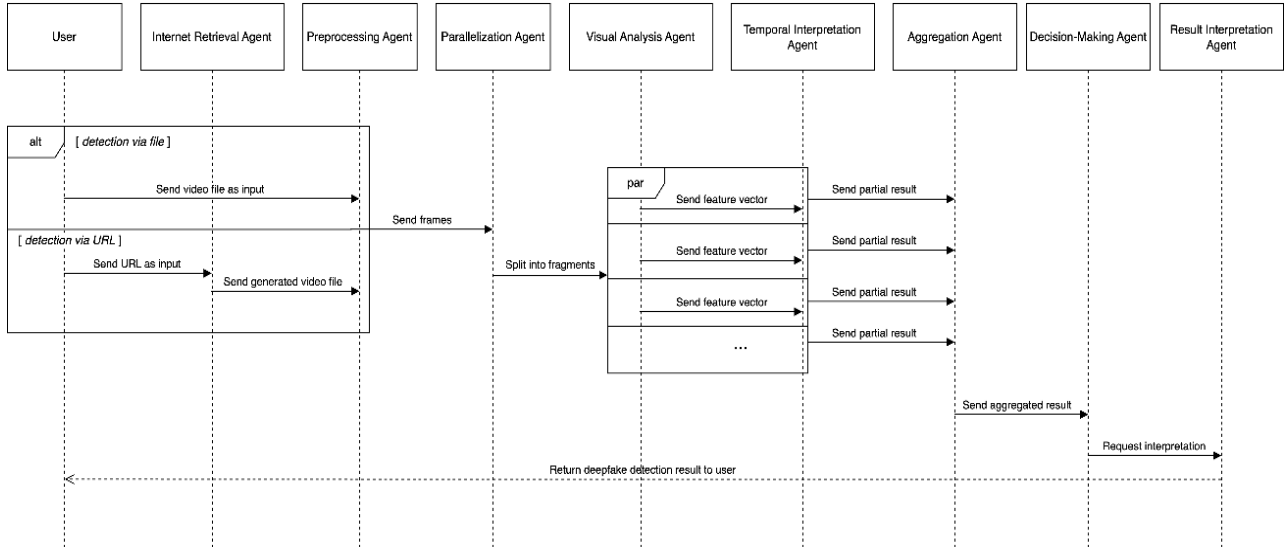


Figure 1 – Sequence diagram of agent interactions in the proposed MAS-based deepfake detection system

In addition to architectural improvements, this approach also enhances the depth and quality of analysis. In the previous system, due to computational constraints, only a limited number of frames were selected for evaluation, which reduced the representativeness of the video data. The updated system, however, leverages the parallelization agent, which dynamically assesses the total number of frames in a video and determines how many CNN+LSTM processing pairs should be instantiated, thereby controlling the degree of parallelization. This allows for more comprehensive frame coverage and a significantly deeper analysis of long and high-resolution videos, ensuring that subtle manipulations are not overlooked.

3. Results Comparison

We conducted comparative experiments using the same datasets as in our initial study (FaceForensics++, DFDC, Celeb-DF). To evaluate how detection performance scales with video length, we compared the three models using both short and long video inputs. The CNN-LSTM+MAS architecture maintains low inference time even on longer videos due to dynamic parallelization. Moreover, it achieves the highest accuracy on long-form content (94.7%), thanks to its ability to analyze all fragments comprehensively rather than sampling a subset of frames, as done in earlier approaches.

4. Conclusion

This study demonstrates that transitioning from a monolithic hybrid CNN-LSTM model to a modular multi-agent architecture yields substantial benefits in both accuracy and efficiency of deepfake detection. The proposed system not only

improves classification performance across various video lengths but also reduces inference time through intelligent parallelization. Its modular nature supports scalability, adaptability to new datasets and manipulation techniques, and seamless integration into real-time media verification pipelines.

Future work will focus on enhancing model robustness against adversarial attacks and integrating additional agents for context-aware analysis and cross-platform deployment, particularly in social media and streaming environments.

References:

1. Hrynishak, I., & Krasovska, H. Hybrid CNN-LSTM Model for Deepfake Detection. Taras Shevchenko National University of Kyiv, 2024. Information Technology and Implementation (Satellite): Conference Proceedings, November 21, 2024, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publishing House «Caravela», 2024. Pp. 34-36
2. Deeptrace Labs. "The State of Deepfakes: Landscape, Threats and Impact." Deeptrace Report, 2024. Available at: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf
3. Sumsb & Sensity.ai. "Deepfake Fraud Report 2025: Financial and Political Manipulation Trends." Global Identity & Deepfake Analytics, January 2025.
4. Xie, L., & Zhao, T. "A comprehensive review of deepfake technology." IEEE Transactions on Multimedia, vol. 21, no. 8, pp. 1974–1986, 2021.

Vladyslav Hyzha

Student of the Department of Software Engineering

Dmitry Gololobov

PhD, Associate Professor of the Department of Software Engineering

State University «Kyiv Aviation Institute», Ukraine

USING AI AGENTS TO ENHANCE USER EXPERIENCE IN PUBLIC TRANSPORT ROUTE PLANNING

Abstract. The purpose of this publication is to examine the role of AI agents in improving user experience within public transport route-planning systems. The study focuses on how conversational interaction, supported by large language models, simplifies trip planning and enhances accessibility for diverse users.

Keywords: artificial intelligence, agent, user, OpenTripPlanner.

In recent years, driven by the rapid advancement of large-language models (LLMs) and associated conversational artificial intelligence (AI) frameworks, the possibilities for deploying AI agents in mobility Applications have grown significantly. When combined with routing and multimodal transit engines, such agents enable new forms of human-machine interaction that go beyond the traditional drop-down and map-based user interfaces. The subject of consideration in this study is the design, implementation, and user-experience impact of a conversational AI agent integrated into a public-transport route-planning application that uses the open-source OpenTripPlanner (OTP) platform as its routing backend.

Traditionally, public-transport route-planner applications require users to navigate a sequence of UI steps: select origin, select destination, choose departure or arrival time, set preferences (e.g., fewest transfers, least walking), then examine results (often on a map). However, many users today are more familiar with conversational interfaces – chat apps or messenger-style interactions. By presenting a chat-based interface driven by an AI agent, it will reduce cognitive overhead and enable a more inclusive experience, particularly for occasional transit riders or less tech-savvy users. The user simply issues a natural-language request (for example, “From home to the university, how do I get there by bus or tram leaving in 20 minutes?”), and the agent interprets the request, generates a routing query, and returns a structured response – enriched with a visual map component if needed.

From an architectural perspective, the system consists of three main parts. The first is the conversational AI agent, which interprets the user’s message and understands what kind of route they are looking for. The second is the routing engine, powered by OTP, which calculates the best possible routes based on available transport data. The third is the user interface, where the results are shown in a simple chat format, often complemented by a small map for better visualization. OTP was chosen because it is open-source, supports different modes of travel, and is already widely used in public transport systems [1][2]. When the user types a request, the agent identifies the key details and sends a request to the routing system. After receiving possible route options, it converts them into a clear, conversational reply.

The process is designed to feel natural and effortless: the user only needs to send one message instead of navigating through several screens or menus. The agent also may ask follow-up questions, remember travel habits, and update routes when delays or disruptions occur. Because OTP supports real-time data feeds and accessibility information, it allows the system to provide inclusive and current travel advice [3].

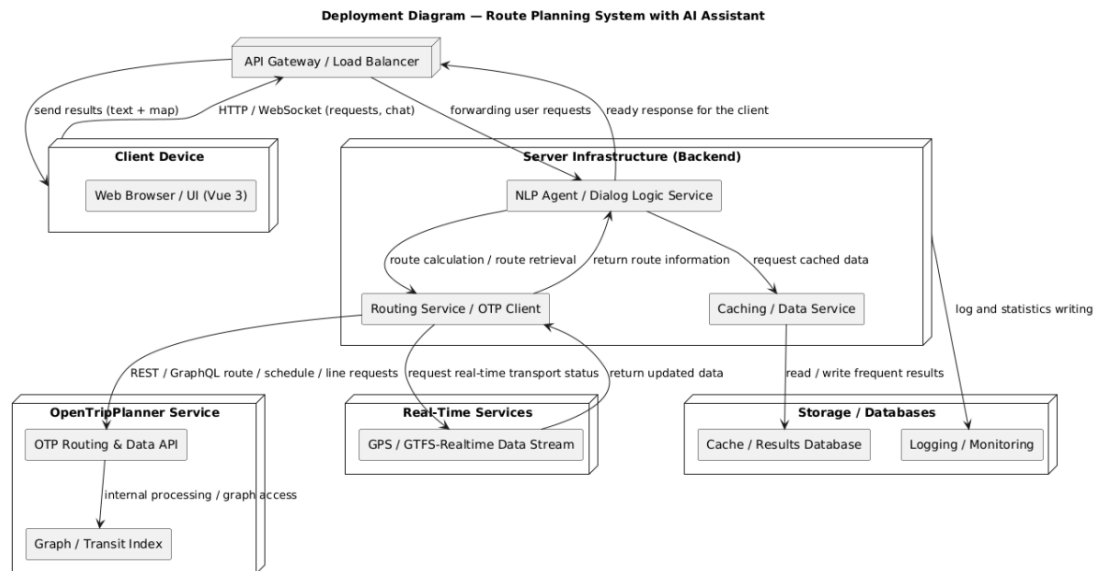


Figure 1 – Deployment diagram of route planning system with AI agent

Conclusion. Overall, the synergy of conversational AI agents and open-source routing engines such as OTP offers a promising direction for public-transport route-planning applications. By replacing complex interface interactions with natural-language dialogue and providing visual route representations, such systems can make public transport planning more accessible, efficient, and user-friendly.

References:

1. Jevinger, Åse & Zhao, C. & Persson, Jan & Davidsson, Paul. (2023). Artificial intelligence for improving public transport: a mapping study. *Public Transport*. 16. 1-60. URL: https://www.researchgate.net/publication/375771439_Artificial_intelligence_for_improving_public_transport_a_mapping_study
2. “OpenTripPlanner – Multimodal Trip Planning,” OpenTripPlanner, accessed 2025. URL: <https://www.opentripplanner.org>
3. Morgan, Malcolm & Young, Marcus & Lovelace, Robin & Hama, Layik. (2019). OpenTripPlanner for R. *Journal of Open Source Software*. 4. 1926. 10.21105/joss.01926. URL: https://www.researchgate.net/publication/337697116_OpenTripPlanner_for_R

Andrii Ivanenko, PhD student

Andriy Shatyrko, Doctor of Phys.-Math. Sciences, Professor

Taras Shevchenko National University of Kyiv

METHODS OF TARGET IDENTIFICATION AND LOCALIZATION BY SOUND SIGNAL

Abstract. This paper examines a two-stage method for the identification and localization audio source, particularly a UAV, based on its acoustic signal. The first stage involves sound classification using self-supervised learning models, where the BEATS model achieved 93% accuracy. The second stage entails determining the spatial position of the source using Direction of Arrival (DOA) estimation methods with microphone arrays.

Keyword: Audio classification, Audio source localization, Direction of arrival (DOA) Deep learning Self-Supervised Learning BEATS, Microphone array

The use of UAVs has significantly increased since the late 2010s, posing an everyday threat to critical infrastructure. Traditional detection methods are ineffective against them due to their small size and composite materials.

Every classic method has limitations in detecting various objects: radar is weak against small drones; optical/thermal cameras rely on visibility; and RF monitoring only works with radio-controlled devices.

Therefore, the most reliable approach for the comprehensive detection of threats (UAVs, gunfire, etc.) is multi-sensor data fusion, which combines the advantages of different sensors. In this context, acoustic detection is an important component for solving this problem.

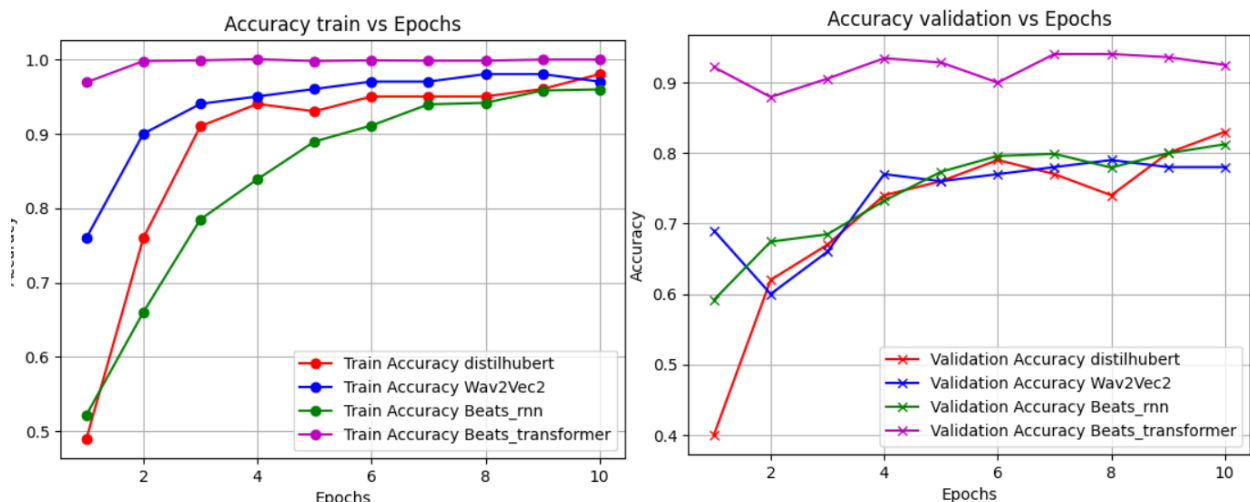


Figure 1. Models' accuracy

To ensure reliable acoustic classification of sound signals in real-world conditions, an approach based on Self-Supervised Learning (SelfSL) was chosen. The advantage of SelfSL lies in its training on large, unlabeled datasets, which is critically important where manual annotation is limited.

In the study, three leading SelfSL models were compared: BEATS [1], DistilHuBERT, and wav2vec. Experiments were conducted using five sound classes: "airplane," "car/truck," "UAV" (Unmanned Aerial Vehicle), "gunfire," and "tank."

The comparative analysis showed that the BEATS model, supplemented with a Transformer layer, achieved the best result, reaching an overall classification accuracy of 93%.

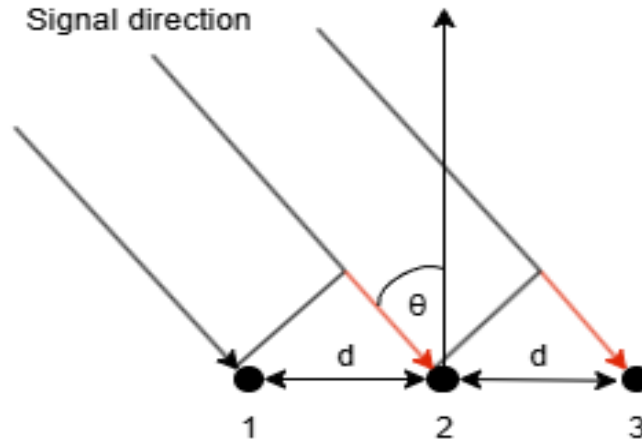


Figure 2. Diagram of a linear antenna array

The next critical task after the acoustic model classifies a sound event is to determine its precise location. This is achieved using Direction of Arrival (DOA) estimation methods.

DOA allows for determining the direction from which a sound or radio signal arrives. This relies on an array of microphones or antennas arranged in a specific geometry (linear, circular, or grid). When a planar wave reaches the array at a certain angle, it first hits one element and then others with a slight time and phase delay (Fig-2). These delays are directly related to the angle of arrival: the larger the angle, the greater the difference in phase and time between adjacent sensors. This allows us to estimate the source's origin.

In practice, various algorithms are used for this: simple energy methods look for the maximum total power in a specific direction, correlation methods determine the time delay between signals, and more complex spectral and subspace approaches can locate multiple sources simultaneously and increase accuracy in challenging conditions. As a result, DOA converts a set of signals from the sensor array into spatial information about the source's location.

An extensive review study was conducted on this topical issue, resulting in an overview article. Our article examined both classical approaches and modern algorithms based on optimization and statistical models. For each method, both the principle of operation and the underlying mathematical apparatus were described.

If we summarize the results of the analysis, we obtain a generalized table (Table 1) below, which presents the key characteristics of the discussed classic DOA methods [2].

Method	Resolution	Works with Coherent Sources?	Noise Robustness	Computational Complexity
Bartlett	Low	Yes	Low	Low (1-D Search)
GCC-PHAT	Medium	Yes	High	Low (1-D Search)
MVDR	High	No	High	High (1-D Search)
MUSIC	Very High	No	High	High (EVD + 1-D Search)
ESPRIT	High	No	Medium	Medium (EVD)
LASSO	High	Yes	Medium	Very High (Iterative Method)

Table 1. DOA Methods

Conclusions. The work presents an acoustic approach to object detection and localization. The system first performs sound classification, recognizing a wide spectrum of targets, including UAVs, transport, gunfire, tanks, and aircraft. After source identification, DOA (Direction of Arrival) algorithms, specifically methods like Bartlett [3], MVDR, and MUSIC [4], are used to determine its spatial location.

Further research focuses on improving robustness and accuracy. Plans include the integration of deep learning with classical DOA algorithms and the development of hybrid systems that combine acoustic methods with optical sensors, use of multiple microphone arrays for triangulation, which will allow for determining the object's precise 3D coordinates, vital for applications in robotics.

References:

1. Chen S. BEATs: Audio pre-training with acoustic tokenizers / S. Chen, [et al.] // arXiv.org. – 2022. – 18 Dec. – DOI: 10.48550/arXiv.2212.09058.
2. Krim H. Two decades of array signal processing research: the parametric approach / H. Krim, M. Viberg // IEEE Signal Processing Magazine. – 1996. – Vol. 13, No. 4. – pp. 67–94. DOI: 10.1109/79.526899
3. Yaro A. S. Multiangulation position estimation performance analysis using a Bartlett's Beamforming Method / A. S. Yaro, [et al.] // Nigerian Journal of Technology. – 2017. – Vol. 36, No. 4. – pp. 1155–1161. DOI: <https://doi.org/10.4314/njt.v36i4.23>
4. Fang Y. Low-Complexity MUSIC Algorithm: From Finite-Precision Perspective / Y. Fang, [et al.] // arXiv.org. – 2025. – 12 p. – DOI: 10.48550/arXiv.2503.12316.

Yaroslav Khrolenko

PhD Student, Decision Support Systems Laboratory

Institute for Information Recording of NAS of Ukraine, Kyiv

SEMANTIC SIMILARITY, CO-OCCURRENCE, AND SEMANTIC RELATEDNESS AS INDICATORS OF THEMATIC AFFINITY

Abstract. This study examines theoretical and methodological approaches to assessing the semantic affinity of scientific publications based on three interrelated indicators: semantic similarity, co-occurrence, and semantic relatedness. These indicators are shown to serve as the foundation for thematic analysis, knowledge clustering, and the automated generation of expert profiles in modern scientific data aggregators such as OpenAlex, Semantic Scholar, Dimensions AI, and Scite.ai. The paper characterizes typical models, metrics, and tools that facilitate the computation of thematic affinity based on linguistic and bibliometric features.

Keywords: semantic similarity, co-occurrence, semantic relatedness, thematic affinity, OpenAlex, MAG, NLP, scientometrics.

Modern information-analytical systems for handling scientific publications increasingly rely on natural language processing (NLP) and scientometric methods. One of the key tasks of such systems is identifying the thematic affinity of scientific works, i.e., the degree of similarity in their content, methodology, or research objects. Thematic affinity is a critical indicator for various tasks, including the development of recommendation systems, grouping experts by research domains, constructing knowledge maps, automated peer review, and intelligent search for relevant studies. The primary data sources for these systems are open scientific publication aggregators, such as OpenAlex, Semantic Scholar, Microsoft Academic Graph (MAG), Scopus, and Dimensions. These platforms integrate publication metadata, author information, citations, keywords, and concepts, enabling the construction of multi-level models of semantic affinity.

Semantic similarity is used to measure the closeness between two units of knowledge—words, terms, concepts, or documents. The literature identifies three main approaches to its evaluation: taxonomic, statistical, vector-semantic.

Taxonomic (ontology-based) approach relies on determining the distance between concepts within a structured hierarchy (e.g., WordNet, UMLS, ACM Computing Classification System). Metrics such as Resnik, Lin, and Jiang–Conrath assess similarity based on shared ancestors in the concept tree [1].

Statistical or distributional approach assumes that words or phrases with similar contexts in text corpora are semantically similar. It employs frequency analysis of co-occurrences within context windows or documents. Typical statistical models and methods include term-context frequency matrices, probabilistic and entropy-based measures (Jaccard similarity, KL-divergence) [2], and corpus-based models (arXiv, OpenAlex Works, MAG Abstracts) [3].

Vector-semantic (embedding-based) approach represents knowledge units as vectors in a multidimensional space, with the distance between vectors (often cosine similarity) indicating the degree of semantic similarity. These models are trained on large text corpora to capture contextual and distributional regularities in language,

allowing them to encode both syntactic and semantic information. Unlike purely statistical methods, embedding-based models learn dense, continuous representations that generalize across contexts and domains. Common implementations include Word2Vec, GloVe, and FastText for word-level embeddings, as well as BERT, SciBERT, and SPECTER for contextualized document-level representations.

In contrast to semantic similarity, which reflects the degree of resemblance in meaning between concepts, semantic relatedness encompasses a broader spectrum of relationships — associative, causal, thematic, functional, or part–whole. It may exist even between concepts that are not synonymous but belong to a shared context or research domain. In practical terms, semantic relatedness serves as an interdisciplinary indicator, enabling the identification of conceptual “bridges” between different scientific fields, thereby fostering knowledge integration and the emergence of new research directions.

The quantitative assessment of relatedness is based on analyzing relationships among knowledge elements within a multidimensional semantic space. Methods commonly used include correlation analysis of vector representations of concepts, measures of entropy and mutual information, weighted coefficients in co-occurrence graphs, and semantic analysis of citation contexts, which accounts for the nature and direction of scholarly references [4]. These approaches make it possible not only to measure the degree of similarity between terms or documents but also to uncover structural connections among research topics, construct knowledge maps, and identify interdisciplinary intersections.

In scientific analytics systems (e.g., OpenAlex, Semantic Scholar, Dimensions), semantic relatedness is utilized for clustering thematic spaces, constructing “knowledge landscapes,” and identifying emerging research directions.

Co-occurrence refers to the frequency of joint appearances of two terms within a single document, abstract, or bibliographic reference. Its computation enables the construction of co-occurrence networks, which visualize the thematic structure of a knowledge domain.

Typical metrics include - frequency of joint term appearances, degree and betweenness centrality of concepts, clustering coefficient, modularity of thematic groups [4].

The OpenAlex aggregator, a successor to Microsoft Academic Graph (MAG), implements a multilevel hierarchy of Fields of Study. For each publication, an automated classifier trained on the MAG corpus identifies a set of concepts at varying levels (e.g., from “Artificial Intelligence” to “Convolutional Neural Networks”). This facilitates the computation of semantic proximity between works based on shared concepts and their hierarchical distance.

Semantic Scholar employs the SPECTER and SPECTER2 models, which generate vector representations of scientific articles by integrating the text of titles, abstracts, and citation contexts. Similarity between works is measured through cosine similarity of vectors, enabling the identification of “similar articles” even without matching keywords [5].

Systems like Dimensions AI and Scite.ai combine textual and citation features, analyzing the context of citations. Instead of merely counting citations, these systems consider their nature—supportive, neutral, or critical—thereby refining the assessment of semantic affinity.

Thematic affinity between scientific works can be determined through a combination of several similarity types:

- Content-based (semantic similarity based on titles, abstracts, and keywords);
- Conceptual (overlap of Fields of Study);
- Bibliographic (shared references, co-citation, bibliographic coupling);
- Structural (distance in co-occurrence graphs).

Hybrid models of thematic affinity integrate vector embeddings with graph-based metrics, achieving better alignment with expert judgments.

Such approaches are applied in modern research to construct maps of scientific disciplines, identify scientific communities, and generate reviewer profiles based on publication histories.

Thus, semantic similarity, co-occurrence, and semantic relatedness form a system of indicators for the thematic affinity of scientific works. These indicators not only enable the grouping of studies by shared themes but also uncover latent connections between disciplines.

The integration of these methods into scientific aggregators (e.g., OpenAlex, Semantic Scholar, Dimensions) creates the foundation for developing intelligent systems to support scientific decision-making, from automated reviewer searches to constructing expert competency models.

Future research should focus on comparative analyses of the accuracy of hybrid models combining vector-semantic embeddings with graph-based and citation-based indicators. This will contribute to the development of more transparent and reproducible approaches to analyzing scientific knowledge.

References:

1. S. Harispe, S. Ranwez, S. Janaqi, J. Montmain. *Semantic Similarity from Natural Language and Ontology Analysis*. Springer, 2015. DOI: 10.1007/978-3-031-02156-5
2. M. Shao, L. Qin. *Text Similarity Computing Based on LDA Topic Model and Word Co-occurrence*. In: *Proceedings of the International Conference on Software Engineering, Knowledge Engineering, and Information Engineering (SEKEIE-14)*, 2014, pp. 47–52. DOI: 10.2991/sekie-14.2014.47.
3. J. Priem, H. Piwowar, R. Orr. *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. *arXiv preprint*, arXiv:2205.01833, 2022. DOI: 10.48550/arXiv.2205.01833.
4. O. Marchenko. *Methods for Evaluating the Semantic Similarity and Relatedness of Natural Language Words*. *Artificial Intelligence*, No. 4, 2012, pp. 213–219. URL: <https://nasplib.isoftware.kiev.ua/server/api/core/bitstreams/c447b50d-68c9-482d-9cea-8bf67a9e9cc0/content>.
5. A. Cohan, S. Feldman, I. Beltagy, D. Downey, D. S. Weld. *SPECTER: Document-level Representation Learning Using Citation-informed Transformers*. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 2270–2282. Association for Computational Linguistics. URL: <https://arxiv.org/pdf/2004.07180>.

Victoriia Kodovbetska

Bachelor's student of the faculty of information technology

Julia Minaeva

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Taras Shevchenko National University of Kyiv

RECOMMENDATION SYSTEM FOR FOOD ESTABLISHMENT SELECTION IN LVIV

Abstract. Nowadays, restaurants, cafes, and food establishments in general are important places for socializing. Selecting a venue among various options with different cuisines can be challenging for citizens, especially newcomers. The proposed system addresses such challenges and allows users to find food establishments suited to their personal preferences. The developed application uses a cascade hybrid technique, combining collaborative and knowledge-based filtering to provide personalized suggestions that enhance the user experience.

Keywords: recommendation system, food establishments, collaborative filtering, knowledge-based filtering, cascade hybrid method.

Introduction. The research focuses on a relevant task of developing a recommender system for food establishments. Selecting a food establishment has become a challenging task, especially due to various options with different cuisines and varying budgets, for local citizens and newcomers as well. According to the situation in recent years, many people have had to relocate, and an effective way of getting familiar with a new city, socializing, and adapting to a new environment is by exploring the local cuisine.

The development of a recommender system for selecting food establishments in large urbanized cities is an important task. The designed system will allow not only to find venues according to users' preferences, but also to improve the promotion of local restaurant businesses. As a case study for the development of the recommender system, the city of Lviv was chosen.

Methodology. After analyzing various approaches for designing recommender systems, a hybrid cascade method was chosen. This approach has advantages due to combining collaborative and knowledge-based filtering.

The food establishment selection task requires identifying the recommender function $f: U \times I \rightarrow R$, where U represents a set of users, I represents a set of restaurants, and R denotes a set of ratings. The goal is to maximize the accuracy and relevance of recommendations through a hybrid approach that includes collaborative filtering, content-based analysis, and knowledge-based rules.

Implementation. A cascade hybrid approach was chosen for implementing the application, combining collaborative filtering and knowledge-based filtering [1,2].

The advantages of collaborative filtering include the ability to identify unexpected options for the user that go beyond their usual preferences, using rating information to form recommendations based on the preferences of other users with similar interests.

Thus, the knowledge-based approach allows for providing recommendations for new users without the detailed historical data required for applying collaborative filtering. The method uses comprehensive information about food establishments, such as type of cuisine, location, and price range, which allows obtaining a high level of accuracy for recommendations and ensures the ability to create recommendations for specific user requests.

Results and Expected Outcomes. The developed recommender system successfully provides personalized recommendations by combining both filtering methods [2-3]. For enhancing the accuracy and relevance of the recommendation, the users' feedback changes are used for dynamic updates.

It effectively addresses the cold start problem by using the knowledge-based approach for new users while enhancing recommendations with collaborative filtering as user data is gathered and accumulated. The designed application includes options for users to leave reviews and ratings for establishments that have been added.

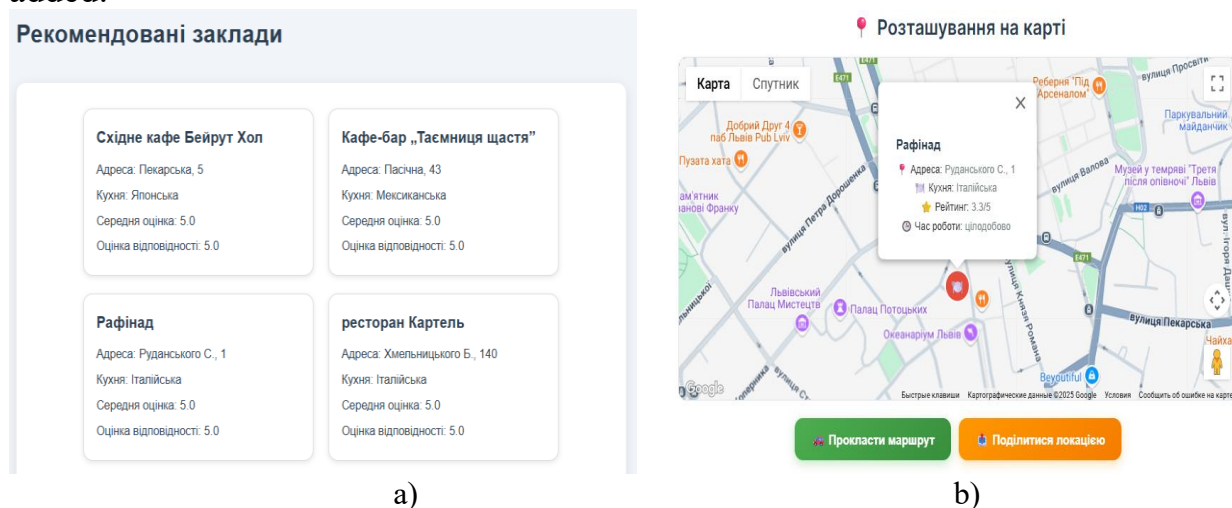


Figure 1.a) Example of the recommended food establishments.

b) Example of an interactive map display for the recommended restaurant.

Conclusion. The recommender system for food establishments was designed, utilizing the cascade hybrid approach. The developed system provides recommendations according to users' preferences, and not only solves the task of personalized selection of venue, but also promotes and supports local businesses. The possible ways of enhancing the system are the implementation of contextual filtering (weather, time of day) and integration with mapping services.

References:

1. Ricci F., Rokach L., Shapira B. *Recommender Systems Handbook* [Electronic resource]. Available at: https://www.researchgate.net/publication/227268858_Recommender_Systems_Handbook
2. Sarwar B., Karypis G., Konstan J., Riedl J. *Application of Dimensionality Reduction in Recommender System.* [Electronic resource]. Available at: https://www.researchgate.net/publication/2824548_Application_of_Dimensionality_Reduction_in_Recommender_System_-_A_Case_Study
3. Burke R. *Hybrid Recommender Systems: Survey and Experiments* [Electronic resource]. Available at: https://www.researchgate.net/publication/263377228_Hybrid_Recommender_Systems_Survey_and_Experiments/

Volodymyr Kochuma

PhD student of the Project Management Department

Kyiv National University of Construction and Architecture

IT ENSURING MODELS FOR INNOVATIVE DEVELOPMENT OF CONSTRUCTION (DEVELOPMENT) COMPANIES

Abstract. The innovative development of construction (development) companies as a subject area for the application of IT development tools is considered. The main problems in the field of IT support of the development business are identified. The directions of innovative development of construction (development) companies are proposed. The main IT technologies that ensure the innovative development of construction (development) companies are identified. The models of IT support of the innovative development of construction (development) companies are considered. A conceptual model of IT support of the innovative development of a developer is proposed, the model includes the data level, the digital platform level, the management level and the client interaction level. The directions of further research are determined; the conclusions are formulated.

Keywords: IT ensuring, project management in development, digitalization, innovative development, IT architecture.

Introduction. Digitalization of the economy and the active development of information technologies are significantly transforming business processes in the field of development. In modern conditions, construction companies operate in an environment of high competition, market uncertainty, increased requirements for environmental friendliness and energy efficiency of facilities. In these conditions, the key factor for sustainable development and competitiveness is the ability of construction (development) companies to generate, integrate and effectively use innovations, among which digital technologies play a dominant role.

IT ensuring is no longer an auxiliary function - it is becoming a strategic asset, belonging to the “organization’s process assets” according to the formulation of the world's leading project management standard PMBOK (Guide to the Project Management Body of Knowledge) [1] of the American PMI Institute.

It is worth noting that construction companies face the problem of the lack of holistic models for integrating digital solutions (due to the increasing dynamics of digital development and the resulting increase in chaos), which leads to data fragmentation, inefficiency of project management, and increased costs and risks. That is why there is a need to form system models that would describe the relationship between IT solutions and the processes of innovative development of construction (development) companies.

The relevance and problems of digital transformation of the development industry and companies operating in the industry can be determined by analyzing the features of the functioning of development companies in the digital economy. The methodological part of research in this area should be based on both classical work on project management [2] and the use of Agile methodology [3], as well as combinations of Agile with classical models and methods [4], since IT development of construction (development) companies contains both a traditionally managerial and a significant IT component.

Main part. Development (in the construction sector) is a complex multidisciplinary activity that covers the following areas: market analysis and search for land plots; design and coordination of design and estimate documentation; financial planning; construction management; facility operation; marketing and sales [5].

Each of these areas generates large volumes of heterogeneous data, which requires integrated IT systems.

Let's identify the main problems in the field of IT support for the development business.

1. Fragmentation of digital infrastructure – different divisions of the company (as well as different teams in a development project) use incompatible IT products.
2. Low level of digital culture of personnel.
3. Lack of a unified digital development strategy.
4. Insufficient development of analytical systems and data-based forecasting systems.
5. Slow implementation of innovations due to regulatory restrictions and lack of investment.

These problems create barriers to improving the efficiency of business processes in development.

The analysis allows us to formulate theoretical principles for the formation of models of IT support for innovative development in development. Let us explain the concept of innovative development of construction (development) companies. Innovative development in development involves:

- creation of new products (innovative residential and commercial complexes);
- implementation of technological innovations (BIM, digital twins, “smart” buildings);
- increasing the efficiency of the business model (automation, cost optimization);
- applying new management methods.

A decisive contribution to ensuring innovative development is made by modern information technologies. IT provides:

- data collection and integration;
- process automation; decision support;
- optimization of design and construction solutions;
- high accuracy and transparency of project management;
- improving the quality of customer service.

Let's identify the main IT technologies that ensure the innovative development of development companies.

1. BIM (Building Information Modeling)

BIM provides a digital representation of the object at all stages: design, construction, operation. Its implementation allows:

- reduce design errors;
- optimize construction schedules;
- plan costs;
- create digital twins with the possibility of further use in facility management.

2. *CIM (City Information Modeling)*

CIM allows you to model entire urban areas taking into account traffic flows, environmental parameters, building density, and engineering infrastructure, which is extremely important for the successful implementation of complex development projects.

3. *IoT infrastructure*

The Internet of Things provides data collection from sensors in buildings, which forms the key basis for “smart” management of real estate objects, the transition from “smart home” technologies to “smart city”.

4. *Big Data analytics (Big Data) and artificial intelligence (AI).*

The use of AI allows you to forecast demand, optimize a marketing campaign, model financial risks, automate management decisions, etc.

5. *Cloud technologies*

The use of cloud data storage and cloud services provides scalability, remote access, rapid deployment of digital solutions, the ability to monitor and manage online.

Let's consider models of IT support for innovative development of construction (development) companies.

1. Process model.

Focused on optimizing business processes using IT solutions. Provides for the creation of a process map, identification of bottlenecks and duplications in processes and their elimination, automation of individual stages of processes.

2. Integrated IT architecture model.

It involves building a single IT architecture that includes:

- Enterprise Resource Planning (ERP);
- Project Management System (PMS);
- Customer Relationship Management (CRM);
- BIM platform;
- Data Warehouse (DWH);
- Analytical services.

3. Innovation model

Forms an environment for generating innovations through R&D departments, innovation hubs, interaction with startups, open innovation tools.

4. Digital ecosystem model

The most modern approach, which involves the integration of all market participants through a single digital environment in which the developer, designers, contractors, investors and buyers communicate.

A conceptual model of IT support for the developer's innovative development is proposed, which includes four levels:

1) Data level: BIM/CIM data, geospatial data, project management data, IoT data, corporate data;

2) Digital platform level: integration platform (ESB), cloud infrastructure, analytics modules;

3) Management level: project management (PM) processes, financial control processes, marketing and sales management processes;

4) Level of customer interaction: mobile applications, customer services, digital marketing.

It is proposed to determine the effectiveness of the implementation of IT models by indicators:

- reduction of construction time;
- reduction of non-target costs;
- increase in project profitability;
- increase in the accuracy of forecast models;
- reduction of risks;
- increase in customer satisfaction.

At the same time, it is necessary to identify barriers and risks of implementing IT solutions in development: lack of qualified specialists, high costs of digital infrastructure, resistance to innovations from middle management, the need to ensure a high level of cybersecurity and data protection, inconsistency of the regulatory framework.

The following prospects for the development of IT in construction development should be formulated: the development of digital twins of megacities, full automation of design using AI, robotization of construction, the use of blockchain in contract management, the creation of “smart districts” with autonomous infrastructure. To create a scientific component of the specified direction, the following prospects for further research in the chosen direction should be identified: identification of processes of development companies that require digitalization, creation of models and methods of comprehensive digitalization of development companies, development and implementation of metrics for measuring the effectiveness of digitalization.

Conclusions. In the current conditions of digitalization of the economy, IT support is becoming a key tool for innovative development of construction (development) companies. Building integrated IT architecture models allows to increase the efficiency of processes, reduce project implementation times, optimize costs and ensure sustainable development of the company. The analysis summarizes the most promising technological solutions and forms the basis for further research into the ways of digital transformation in the field of construction development.

References:

1. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition (2021). USA, *Project Management Institute*, 250 p.
2. Kerzner, H. (2015) Project Management 2.0: Leveraging Tools, Distributed Collaboration, and Metrics for Project Success. *Wiley*, 226 p.
3. Stellman, A. Greene, J. (2013) Learning Agile: Understanding Scrum, XP, Lean, and Kanban. *O'Reilly Media*, 420 p.
4. Spundak, M. (2014) Mixed agile/traditional project management methodology – reality or illusion? 27th IPMA World Congress. *Procedia – Social and Behavioral Sciences*, Vol.119, pp. 939-948.
5. Organizational and technological development and integrated implementation of projects in construction: [collective] monograph (2025) [author collective]; Kyiv National University of Construction and Architecture. Kyiv, *KNUBA*, 201 p.

Serhii Korotenko

PhD Candidate, ORCID ID: 0009-0003-9236-4775

Roman Yarovy

PhD, Associate Professor, Scientific Supervisor

Private Higher Education Establishment "European University", Kyiv, Ukraine

CHALLENGES OF ADAPTING MACHINE LEARNING MODELS TO STRUCTURAL DRIFT IN BUSINESS PROCESSES

Abstract. This paper discusses the problem of structural drift in business processes and its impact on the application of machine learning methods. It shows that classical BPMN models and process mining do not provide sufficient alignment between formal models and real executions, while current ML architectures mostly focus on sequential data. This creates a gap between the business need for flexible monitoring and the ability to adapt to changes. The paper highlights the key challenges that need to be addressed to build adaptive ML solutions for business processes.

Keywords: BPM, ML, structural drift, process mining, adaptation

Business processes in modern organizations are in constant change. Regulatory requirements, new business strategies, and internal reorganizations make process models quickly outdated. This phenomenon is known as structural drift. As van der Aalst (2014) noted, “there is no knowledge without processes” since they reflect the actual activities of organizations. However, there is always a gap between the recorded model and the real execution, which negatively affects prediction accuracy and the practical value of systems [1][3].

Structural drift can be defined as a set of changes $D = (\Delta T, \Delta R, \Delta P)$, where ΔT represents task changes, ΔR - refers to resource or role changes, and ΔP - refers to policy changes. This helps to formally describe different transformations in a dynamic environment.

A business process can be represented as a directed graph $G = (V, E)$, where V - is the set of tasks and events, and E - is the set of control-flow relations. Structural drift occurs when G evolves into G' with modified V or E , which reflects the addition of new steps, reordering or removal tasks. At this point, ML models face significant difficulties, because the training data becomes structurally inconsistent. The situation becomes even more complex when both old and new versions of the process coexist and run in parallel. In this case, the ML model must interpret a mixed event space, where the same business case may follow different structural rules.

Traditional notations, such as BPMN or Petri nets, describe sequences of actions but focus on an “ideal scenario,” leaving many constraints implicit [7]. As a result, they do not capture actual behavior, including deviations that often cause major problems. Process mining tries to bridge this gap by analyzing event logs, but it also has limitations: results are often sequential, making it difficult to account for parallelism, and there is a lack of integration with other analytical tools [2]. In this context, Graph Neural Networks appear promising because they can process data directly as graphs and better capture parallelism and complex structural dependencies in business processes.

Modern ML methods, including LSTM and Transformer, are applied to predict next activities and execution times. However, they usually treat logs as linear sequences, ignoring structural constraints, organizational roles, and policies [4][8]. This limits their generalizability and robustness in dynamic environments. Formally, a model M trained on logs $L(t)$ becomes inconsistent with the process version $P(t + 1)$, which can be expressed as $M(L(t)) \neq P(t + 1)$.

Structural drift appears in many forms: adding or removing tasks, changing their order, altering roles and resources, and updating policies. All these factors complicate the task of building adaptive ML solutions that can work online. Key challenges include handling multi-variant structures, integrating organizational context, tracking process versioning, and keeping the results interpretable. Figure 1 illustrates structural drift, where different versions of the same process (v1, v2, v3) coexist and are executed in parallel.

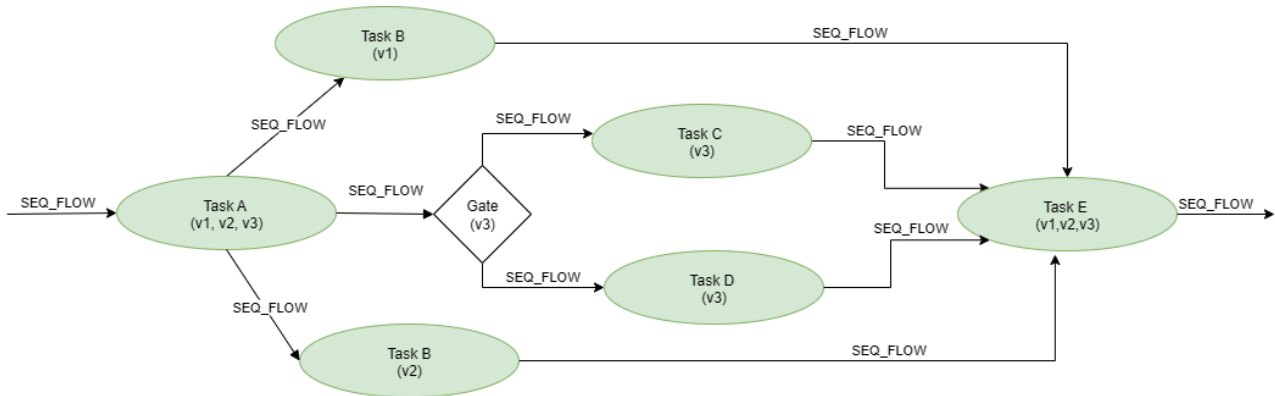


Figure 1. Example of structural drift with parallel versions of a business process

Conclusion

Structural drift in business processes creates serious barriers for both traditional and modern ML models. A promising approach that can combine the flexibility of ML with business context is the use of Knowledge Graphs [5][6], which can be defined as $KG = (V, E, C)$, where V are vertices (tasks, events, resources), E are relations between them (control-flow, roles, policies), and C - is the context, including time slices and process versions. This structure generalizes the process graph $G = (V, E)$ by adding the environmental changes described in $D = (\Delta T, \Delta R, \Delta P)$. Using Knowledge Graphs helps align execution data with business context and allows ML models to learn not only from event logs but also from structural and contextual features, by applying structural masking before training and inference. Future research should explore this approach in detail and evaluate its effectiveness on real-world data.

References:

1. Aalst W. M. P. van der. No Knowledge Without Processes: Process Mining as a Tool to Find Out What People and Organizations Really Do. *Proceedings of the International Joint*

Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2014). Issue January 2014. P. 11–16.

2. Aalst W. Van der Process Mining. Berlin, Heidelberg : Springer Berlin Heidelberg, 2016. 1-467 p. DOI:10.1007/978-3-662-49851-4.

3. Dumas M., Rosa M. La, Mendling J. et al. Fundamentals of Business Process Management. Berlin, Heidelberg : Springer Berlin Heidelberg, 2018. 1-527 p. DOI:10.1007/978-3-662-56509-4.

4. Evermann J., Rehse J.-R., Fettke P. Predicting process behaviour using deep learning. *Decision Support Systems*. Vol. 100, 08.2017. P. 129–140. DOI:10.1016/j.dss.2017.04.003.

5. Gottschalk S., Demidova E. EventKG: A Multilingual Event-Centric Temporal Knowledge Graph. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 10843 LNCS, 2018. P. 272–287. DOI:10.1007/978-3-319-93417-4_18.

6. Guan S., Cheng X., Bai L. et al. What is Event Knowledge Graph: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 35, Issue 7. P. 7569–7589. DOI:10.1109/TKDE.2022.3180362.

7. Hildebrandt T., Mukkamala R. R., Slaats T. Designing a Cross-Organizational Case Management System Using Dynamic Condition Response Graphs. *2011 IEEE 15th International Enterprise Distributed Object Computing Conference*(08.2011). IEEE, 2011. DOI:10.1109/EDOC.2011.35. P. 161–170.

8. Tax N., Verenich I., Rosa M. La et al. Predictive Business Process Monitoring with LSTM Networks. 2017. P. 477–492. DOI:10.1007/978-3-319-59536-8_30.

Volodymyr Kovalchuk

Master's student

Lyubov Borkovska

Candidate of Technical Sciences, Associate Professor

STATE NON-COMMERCIAL ENTERPRISE "STATE UNIVERSITY "KYIV AVIATION INSTITUTE", Ukraine

DIGITALIZATION OF INTERNAL BUSINESS PROCESSES USING INTELLIGENT CHATBOTS: DEVELOPMENT METHODOLOGY AND SOCIAL IMPACT

Abstract. The study investigates the methodology for developing and implementing intelligent chatbots to digitalize internal business processes. The paper highlights the integration of artificial intelligence technologies into corporate management systems to enhance communication, automate repetitive tasks, and improve organizational efficiency. The proposed development methodology demonstrates the practical benefits of chatbot adoption in modern business environments.

Keywords: digitalization, business processes, intelligent chatbots, artificial intelligence, automation.

1. Introduction

The digital transformation of internal business processes is becoming a key factor in increasing organizational efficiency and competitiveness. The use of intelligent chatbots represents one of the most promising tools for automating communication, document management, and employee support. Chatbots based on artificial intelligence (AI) technologies allow for faster and more accurate processing of routine requests while reducing administrative workload.

This paper presents a structured methodology for developing and integrating intelligent chatbots into enterprise systems and analyzes their social and organizational impact.

2. Development Methodology

The proposed methodology consists of six main stages:

1. Business process analysis - identifying repetitive operations suitable for automation.
2. Chatbot architecture design - defining user roles, dialogue logic, and data sources.
3. Technology selection - choosing appropriate AI frameworks such as Dialogflow, Rasa, or Microsoft Bot Framework.
4. Integration with corporate systems - ensuring data exchange with CRM, ERP, or HR systems.
5. Model training and testing - applying natural language processing and machine learning to improve accuracy.
6. Performance evaluation - measuring efficiency indicators, user satisfaction, and time savings.

7. The methodology emphasizes modular design and adaptability to various organizational contexts, allowing chatbots to evolve with business needs.

3. Results and Discussion

The introduction of chatbots into internal management systems has proven to significantly optimize communication and operational workflows. Empirical observations from pilot projects show:

1. a 15-25% increase in productivity,
2. a 30-40% reduction in administrative workload,
3. improved access to information and decision-making speed.

Chatbots also serve as a bridge between employees and digital systems, enhancing transparency and collaboration. However, successful implementation requires continuous monitoring, employee training, and ethical data management.

4. Social Impact

From a social perspective, the use of intelligent chatbots fosters a more dynamic and connected workplace. Employees benefit from reduced routine workload and improved digital literacy. At the same time, it challenges organizations to address issues related to data privacy, adaptation, and resistance to technological change.

The transition to AI-assisted business environments demands not only technical readiness but also cultural acceptance within the organization.

5. Conclusion

The proposed development methodology demonstrates that intelligent chatbots can play a crucial role in the digitalization of internal business processes. Their integration enhances operational efficiency, improves communication quality, and supports strategic decision-making. Future research should focus on combining chatbot systems with predictive analytics and cognitive AI for real-time business management.

References

1. Nurmi J., Wang X., Ahonen T. Applying CDMA technique to network-on-chip. *IEEE Trans. VLSI Syst.*, 15(10), 2007, pp. 1091–1100.
2. Cohen S., Nutt W., Sagie Y. Deciding equivalences among conjunctive aggregate queries. *J. ACM* 54, 2007. doi:10.1145/1219092.1219093.
3. Kosiur D. *Understanding Policy-Based Networking*. 2nd ed. Wiley, New York, 2001.
4. Spector A. Z. Achieving application requirements. In: S. Mullender (Ed.), *Distributed Systems*, 2nd ed., ACM Press, New York, 1990, pp. 19–33. doi:10.1145/90417.90738.
5. Douglass B. P., Harel D., Trakhtenbrot M. B. Statecharts in use: structured analysis and object-orientation. In: Rozenberg G., Vaandrager F. W. (Eds.), *Lectures on Embedded Systems*, LNCS 1494, Springer-Verlag, London, 1998, pp. 368–394.

¹ **Georgii Kozhevnikov**

PhD, Professor, National Technical University "Kharkiv Polytechnic Institute"

² **Serhii Mytrofanov**

Head of the CO "CF "BRIGHT EYES"

³ **Hanna Zavolodko**

PhD, Associate Professor, National Technical University "Kharkiv Polytechnic Institute"

^{1,3} *Department of Multimedia and Internet Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute"*

² *CO "CF "BRIGHT EYES"*

AI: RESTORING AGRICULTURAL PRODUCTION AMID WAR AND CLIMATE CHALLENGES

Abstract. In conditions of war, water scarcity and accelerated climate change, traditional agriculture is losing its ability to ensure food security and employment stability. The article discusses the concept of a compact aquaponics farm with integrated artificial intelligence for resource management. It analyses current scientific achievements in aquaponics, vertical farming and energy management. It assesses the potential for implementing the model in Ukraine as a pilot project with further scalability for regions experiencing drought, conflict and other environmental stresses.

Contemporary global challenges — war, water shortages and climate change — are sharply worsening food security. The war caused losses of over \$40 billion in the agricultural sector [10;11], while mine contamination and damaged irrigation systems further limit traditional farming [12;13]. One solution is technology-oriented small farms with efficient use of water, energy and knowledge. The proposed concept combines aquaponics, environmental management and AI for process optimisation.

Aquaponics. Classic systems at the University of the Virgin Islands yield ~5 tonnes of fish per module annually and up to three times higher basil yields compared to soil [2;3]. FAO confirms water savings of up to 90% [2].

Vertical farming. Freight Farms show high yields in container farms [4]; iFarm claims growth acceleration up to 2× [5]; and ECF Berlin demonstrates successful urban aquaponics [6].

Energy management. Cornell and industry reports confirm that AI can reduce greenhouse energy use by 25–33% [7;8], consistent with the concept's goal of a 30% reduction.

Recent studies also stress IoT and consumer aspects. Malabadi et al. [14] and Sani et al. [15] review IoT-enabled hydroponics and vertical farming for sustainability. Kiu et al. [16] report generally positive consumer views of aquaponics, though willingness to pay is limited. Anila and Daramola [17] highlight that most smart aquaponics focus on monitoring, not yield prediction. Heo et al. [18] confirm higher yields in aquaponics vs. hydroponics under controlled lighting. Owusu et al.

[19] show deep learning can predict water parameters with 99% accuracy, while Yau et al. [20] propose low-cost Arduino-based IoT for small farms.

Recent reviews show that the vast majority of existing solutions in the field of smart aquaponics (73%) focus on monitoring and controlling environmental parameters, while yield forecasting, profitability, resource optimisation and sustainability assessment are still largely undeveloped. Only 37.7% of studies included formal evaluation of the proposed systems, and very rarely involved end users for testing. This indicates a significant scientific and practical gap.

Comparative experiments between aquaponics and hydroponics demonstrate the significant advantages of the former: in greenhouse conditions, the yield of green crops and herbs in aquaponics was 60–70% higher, and the yield of spring onions was 200% higher than in traditional hydroponics. At the same time, control over light spectra and water quality remains a critical factor.

The integration of machine learning and deep learning methods opens up new horizons. For example, ConvLSTM models outperform traditional algorithms in predicting water temperature in aquaponics, achieving 96–99% accuracy. This provides a reliable basis for creating predictive control systems and minimising manual labour.

From a technical standpoint, a series of prototypes based on Arduino and IoT sensors have been developed to monitor pH, temperature, dissolved oxygen, and water levels. Such systems provide an inexpensive and reliable solution for small and medium-sized farms, demonstrating the potential for scaling.

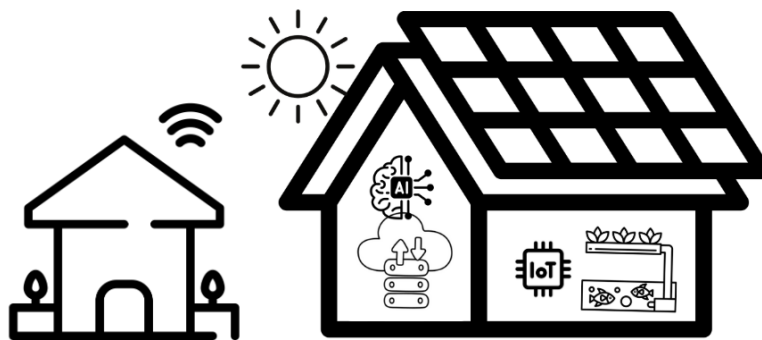


Figure 1 – Structural model of Smart Farm AI

The Figure1 illustrates the structural model of the Smart Farm AI intelligent greenhouse system, integrated with residential infrastructure and cloud-based digital services.

The architecture comprises the following subsystems:

Energy module – photovoltaic panels installed on the greenhouse roof supply renewable electricity for autonomous operation and reduce dependence on external energy networks.

Digital control module – artificial intelligence (AI) combined with cloud computing resources enables data analytics, environmental parameter forecasting, and

optimization of operating regimes, including lighting, ventilation, and water circulation.

IoT infrastructure – a distributed network of sensors and actuators provides continuous monitoring and regulation of critical parameters such as temperature, humidity, pH, electrical conductivity (EC), dissolved oxygen, and CO₂ concentration.

Biotechnological module – an aquaponic unit with a closed water cycle integrates the cultivation of hydrobionts (fish or alternative biological nutrient producers) and plants in hydroponic racks.

User interface – the residential unit with wireless connectivity indicates the capability for remote monitoring and control of farm parameters through Internet-enabled platforms.

Additionally, solar irradiation is depicted to emphasize the integration of natural sunlight with controlled LED lighting within the controlled environment agriculture (CEA) system.

Smart Farm AI's targets — 10 tonnes of fish and several tonnes of greens per 600 m², –30% energy consumption, +50% growth rate — are realistic. They are confirmed by data from UVI, FAO, Freight Farms, and modern AI energy optimisation systems. The implementation of a pilot farm in Ukraine will allow testing the modularity and scalability of the system, creating jobs in regions after demining, and forming a basis for exporting the model to countries facing water shortages and instability.

Conclusions. Smart Farm AI is not just a technology, but a human-centred solution. The Ukrainian pilot project is significant as a scientific and practical demonstration that innovations in aquaponics, artificial intelligence and vertical farming can ensure food security even in crisis conditions.

References:

1. Bibliographic reference. DSTU 8302:2015 [Electronic resource]. – Access mode: https://library.chnpu.edu.ua/download/dstu_8302.pdf
2. FAO. Small-scale aquaponic food production. Fisheries and Aquaculture Technical Paper. – Rome, 2014. – 589 p.
3. Rakocy J.E., Masser M.P., Losordo T.M. Recirculating aquaculture tank production systems: Aquaponics—integrating fish and plant culture. SRAC Publication No. 454. – 2006.
4. Freight Farms. Crop Guide [Electronic resource]. – Access mode: <https://freightfarms.com/>
5. iFarm. Urban vertical farms [Electronic resource]. – Access mode: <https://ifarm.fi/>
6. ECF Farmsystems. Berlin Aquaponics [Electronic resource]. – Access mode: <https://ecf.farm/>
7. Cornell University. AI-based greenhouse energy optimisation. – Ithaca, 2022.
8. Requirements for References [Electronic resource] // EES Journal. – Access mode: <https://ees-journal.com/downloads/Requirements%20for%20References.pdf>
9. Demographic Yearbook 'Population of Ukraine' [Electronic resource] // State Statistics Service of Ukraine. – Access mode: <https://stat.gov.ua/sites/default/files/2023-10/Демографічний%20щорічник%20%27Населення%20України%27.pdf>
10. Total losses from the war in the agricultural sector [Electronic resource] // AVM Agrocenter. – Access mode: <https://avm-ua.org/uk/post/zagalni-zbitki-vid-vijni-v-agrarnomu-sektori-vze-perevisili-40-mlrd-dolariv-zvit-kse-argocentr>

11. Damages Report as of 01.01.2024 [Electronic resource] // KSE. – Access mode: https://kse.ua/wp-content/uploads/2024/04/01.01.24_Damages_Report.pdf
12. National Mine Action Centre Ukraine. Statistics [Electronic resource]. – Access mode: <https://ua-nmac.org/uk/statistics/>
13. How landmines affect the environment [Electronic resource] // Rayon.in.ua. – Access mode: <https://eco.rayon.in.ua/topics/550513-smertelni-syurprizi-viyni-yak-minuvannya-teritorii-ukraini-vplivae-na-ekologiyu>
14. Malabadi R.B., Meti N.T., Chalannavar R.K. Greenhouse farming: Hydroponic vertical farming – IoT technologies: An updated review. *World Journal of Advanced Research and Reviews*. 2024. Vol. 21(2). P. 384–400. DOI: 10.30574/wjarr.2024.21.2.XXXX.
15. Sani S., Ogbaga C., Abdullahi H., Adamu A. Modern approaches to sustainable agriculture. *International Journal of Innovative Science and Research Technology*. 2024. Vol. 9(5). P. 1234–1245.
16. Kiu Q.-S.C., Ooi A.-L., Teoh C.-Y. Consumer perspectives towards aquaponics: Products and native fish cultivation Conference: IOP Conference Series: Earth and Environmental Science 2024. Vol. 1408. 012001. DOI: 10.1088/1755-1315/1408/1/012001.
17. Anila M., Daramola O. Applications, technologies and evaluation methods in smart aquaponics: a systematic review of the literature. *Artificial Intelligence Review*. 2025. Vol. 58. Pp. 1–30. DOI: 10.1007/s10462-025-10789-x.
18. Heo J., Beck J., Subah Z., Ryu J. H. Evaluation of crop growth between hydroponics and aquaponics with different lighting. *Frontiers in Horticulture*. 2024. Vol. 3. 1413224. DOI: 10.3389/fhort.2024.1413224.
19. Ovus R., Mayoko J.S., Lee J.Y. Predicting water temperature during the growth of fish (Nile tilapia) and plants (lettuce) in a prototype safe automated aquaponics environment using deep learning. *Open Access Journal*. 2024. Vol. 11. e11786. DOI: 10.4236/oalib.1111786.
20. Yau S.R., Jalani J., Sadun A.S., Rejab S.M., John J. Development of an aquaponics farming technology system using Arduino based on the Internet of Things. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 2025. Vol. 45(2). Pp. 11–24. DOI: 10.37934/araset.45.2.1124.

Illia Krytskyi

2nd-year Master's student in Software Engineering

Iryna Yurchuk

PhD in Physical and Mathematical sciences, Assoc. Prof. of Software Systems and Technologies Department

Taras Shevchenko National University of Kyiv

APPLICATION OF DEEP LEARNING FOR AUDIO OBJECT CLASSIFICATION AND METADATA GENERATION

Abstract. The purpose of this publication is to present a modern approach to the automatic classification of sound objects and the generation of structured metadata based on deep learning methods. The research focuses on building a hybrid neural network model that combines convolutional and recurrent layers to capture both spectral and temporal properties of audio signals. Such an approach allows the system to recognize environmental sounds, speech patterns, or technical noises and automatically produce descriptive metadata for further data analysis or indexing.

Keywords: deep learning, CNN-LSTM, sound classification, metadata generation, audio processing, artificial intelligence.

The rapid development of artificial intelligence technologies has fundamentally changed the field of multimedia analysis, especially in audio signal processing. Modern intelligent systems increasingly require the ability to automatically interpret sound events for use in environmental monitoring, smart devices, security analytics, and media indexing.

One of the essential components in such systems is metadata — structured information that describes the detected sound object by its main parameters (type, source, duration, intensity, and frequency characteristics). Traditional approaches based on manual labeling and shallow feature extraction are no longer sufficient to meet the accuracy and scalability demands of modern applications.

Therefore, this study aims to develop an architecture for sound object classification and automatic metadata generation using deep neural networks trained on diverse audio datasets.

To achieve the research goal, several existing approaches were analyzed, including classical spectral analysis methods, MFCC (Mel-Frequency Cepstral Coefficients), and advanced deep learning architectures such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer-based models.

The proposed system adopts a CNN-LSTM hybrid model, where the CNN component extracts spatial-spectral features from Mel-spectrograms, and the LSTM component captures sequential dependencies in time. This hybrid design significantly improves classification performance on dynamic sound events such as footsteps, sirens, or speech.

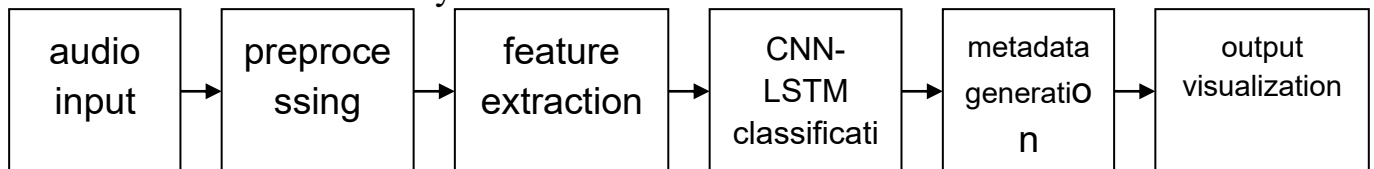
The architecture of the proposed system consists of several main stages:

- Audio preprocessing – noise reduction, normalization, and segmentation into consistent-length clips, followed by conversion into Mel-spectrograms.
- Feature extraction – computation of MFCC, spectral contrast, chroma features, and zero-crossing rates.

- Neural network classification – training a CNN-LSTM model using labeled datasets such as UrbanSound8K or ESC-50, to identify sound categories.
- Metadata generation – construction of a structured JSON document containing fields such as label, confidence, timestamp, duration, and frequency_range.
- Visualization and integration – a web-based user interface developed with React and Next.js, enabling real-time display and export of classification results.

The backend was implemented in Python, utilizing libraries such as Librosa for feature extraction, PyTorch for deep learning, and NumPy/Pandas for data handling.

The workflow of the system can be summarized as follows:



This modular design allows easy integration into higher-level systems for sound-based analytics or automated media indexing.

Experimental evaluations were performed on open-source datasets. The CNN-LSTM model achieved an average classification accuracy of 91–93% across 10 sound categories, outperforming baseline CNN and RNN models by approximately 6–8%.

The system demonstrated robust generalization for both environmental and synthetic audio, maintaining stable predictions even under noisy conditions. The generated metadata provided rich contextual information that can be directly consumed by analytical tools or stored in databases for further processing.

Practical applications include:

- automatic sound recognition in smart surveillance systems,
- environmental and industrial noise monitoring,
- medical and bioacoustic signal analysis,
- multimedia content indexing and recommendation systems.

Conclusion. The developed deep learning-based approach for sound object classification and metadata generation provides an efficient, scalable, and accurate solution for real-time audio analysis. The integration of CNN-LSTM architecture ensures a high level of adaptability and precision in identifying sound events, while the automatic generation of metadata enables efficient organization and retrieval of audio information.

Such technologies can serve as a foundation for intelligent audio-driven ecosystems and facilitate the development of future AI-based applications in smart environments, security, healthcare, and digital media.

References:

1. Choi, K., Fazekas, G., & Sandler, M. (2017). *Convolutional recurrent neural networks for music classification*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
2. Piczak, K. J. (2015). *Environmental sound classification with convolutional neural networks*. IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP).
3. Salamon, J., Jacoby, C., & Bello, J. P. (2014). *A dataset and taxonomy for urban sound research*. ACM Multimedia Conference.
4. Hershey, S., et al. (2017). *CNN architectures for large-scale audio classification*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
5. Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. Y., & Sainath, T. N. (2019). *Deep learning for audio signal processing*. IEEE Journal of Selected Topics in Signal Processing.

Dmytro Masiuk

PhD student

Taras Shevchenko national university of Kyiv

CHALLENGES OF DATA ANONYMIZATION IN EXPERIMENTAL RESEARCH

Introduction

In the modern financial environment, the ability to leverage data for developing intelligent risk detection methods is a huge ethical challenge. However, this does not make it less of a necessity. That is something I need to solve for my research as well. The quality and quantity of the data I would be able to employ for my research hinges on how good would be my anonymization proposal. Anonymization is the process of transforming datasets to prevent identification of individuals or organizations. It has become of paramount importance, in order to ensure privacy and compliance in experiments involving sensitive financial information. As regulatory frameworks such as the General Data Protection Regulation (GDPR) in the European Union and similar local laws worldwide are on the rise, ethical approaches to anonymization are no longer optional but essential to achieve legal compliance and to conduct responsible research.

Main challenges

Risks of re-identification. Advances in machine learning and big data analytics have increased the likelihood of re-identification, especially when anonymized data is combined with external datasets [1]. This raises critical questions about whether traditional anonymization techniques remain sufficient in high-risk financial contexts and can be still used to full extent.

Data privacy versus data quality. Over-anonymization will degrade data quality, sometimes severely, limiting the ability of algorithms to detect necessary patterns in order to locate fraudulent transactions efficiently. Ethical practice requires balancing privacy protections with the need for accurate, actionable models. Finding this balance is proving to be even more tricky as the demand for anonymous data strengthens.

Bias and representation. Removing key demographic or behavioral values will unintentionally bias models, particularly in fraud detection or scoring, potentially worsening their performance as well as their interpretability, which is crucial nowadays.

Recent breakthroughs

A novel technique named SEAL[2], which stands for Self-Refining Small Language Model Anonymizers, trains smaller language models (SLMs) to anonymize text effectively, without relying on expensive or external large models like ChatGPT, version 4. These SLMs are capable of not only anonymizing, but also critiquing their own outputs, that way iteratively improving privacy. What is even more impressive,

8-billion parameter models are able to achieve privacy–utility trade-offs comparable to GPT-4 and can even surpass it in some scenarios.

For example, on the synthetic dataset SynthPAI which consists of personal profiles with text comments, an 8B SLM trained using SEAL achieves a privacy-utility balance on par with GPT-4 anonymizer. With further self-refinement, it even surpasses GPT-4 in privacy—while retaining high readability [2]. This method employs a technique called “adversarial distillation”. Firstly, LLM anonymizer processes original text. Secondly, another model tries to guess private parameters that were previously anonymized. Thirdly, yet another model measures readability, semantic preservation and privacy based on previous results. Basically, these models select a “trajectory” of sorts, which aims to increase privacy, as well as maintain readability and preservation. Afterwards those models are being fine-tuned controllably in order to prefer anonymization with stronger privacy and better utility, refining its outputs iteratively. What is also important, this method is capable of context-aware anonymization. That means it effectively hides not just names or numbers, but hints and patterns that can be exploited by sophisticated models. In my opinion this quality is critical for keeping transactional data safe.

Conclusion

This is a method worth trying out in practice, and I am eager to do so with my data and see the results. I would also like to try various other methods and/or their combinations, as well as compare them against each other, in order find one that gives best results. I will lay my findings out in my future paper on the same topic.

References:

1. Ohm, P. “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization” (2010).
2. Kyuyoung Kim. “Self-Refining Language Model Anonymizers via Adversarial Distillation” (2025).

Oleksandr Mazurets

Ph.D. in Engineering Science, Associate Professor of Computer Science Department

Oleksandr Ovcharuk

Postgraduate student

Roman Vit

Postgraduate student

Khmelnytskyi National University, Ukraine

PRACTICE INVESTIGATION OF NEURAL NETWORK DETECTING OF CONSTRUCTION WASTE BY PHOTOS

Abstract. This paper presents a practice-oriented investigation of neural architectures for visual recognition of construction and demolition (C&D) waste in realistic conditions. We study a two-stage pipeline that first localizes candidate fragments on a scene and then assigns material-specific labels, emphasizing data curation, restrained augmentation, and evaluation protocols aligned with production constraints. Results indicate that coupling a single-shot detector with per-class residual classifiers improves recall on visually confusable materials while preserving precision, offering a pragmatic path toward reliable sorting-line deployment.

Rapid urban growth and post-reconstruction have intensified C&D waste streams and exposed the limits of manual inspection and rule-based vision under dirt, weathering, highlights, and occlusion. Deep networks are robust when data mirror operating conditions, but field viability depends more on cross-scene stability and the ability to separate near-identical textures (e.g., light concrete vs. limestone, weathered wood vs. fiberboard) [1]. This study adopts a practice-minded design with minimal assumptions about image capture, conservative class-balance expectations, and metrics tuned to costly error modes.

The recognition system operates in two stages: a single-shot detector first localizes fragments and suggests coarse material types, while a set of one-vs-rest residual classifiers refines these predictions. The final label combines the most confident binary output with the detector's hypothesis to resolve ambiguities, reducing cross-class interference and enhancing separability for confusable categories. The dataset merged open C&D sources with real-world fragment photos, ensuring diversity in background, lighting, contamination, and occlusion to match deployment conditions. Augmentation remained moderate – limited geometric and photometric changes – to mimic camera variability without distorting material identity and to balance minority classes [2].

Implementation favored reproducibility and portability. The detector used moderate input resolution to balance thin-fragment sensitivity with throughput; residual classifiers consumed aspect-ratio-preserving crops normalized to architecture statistics. Early stopping and light regularization limited overfitting on scarce classes. At inference, a calibrated consensus accepted a binary label when its posterior exceeded a margin and did not contradict the detector; ties defaulted to the detector to preserve spatial context.

The study evaluated the accuracy gains from adding specialized models, the

effect of augmentation breadth, and robustness under distribution shifts. The base detector showed high precision but missed thin or occluded fragments. Incorporating binary specialists improved recall without increasing false positives, indicating sensitivity to micro-textures and edge geometry. Moderate augmentation – rotations, slight affine shifts, and controlled color variations – stabilized minority classes without raising calibration error, confirming that limited data enrichment is effective..

Robustness was tested on hold-out scenes from different cameras and backgrounds. Although absolute scores dipped with illumination and scale shifts, the two-stage pipeline preserved inter-class ordering and a favorable precision–recall balance versus a monolithic baseline. Errors clustered around boundary crops truncating discriminative features, heavy soiling altering apparent albedo, and boxes containing mixed materials. Arbitration mitigated the first two but not the third, indicating value in instance segmentation or multi-label recognition when warranted.

System-wise, the hybrid design balances accuracy and maintainability. Detectors efficiently propose spatial hypotheses; binary specialists trained with hard negatives add discriminative sharpness where needed. Modularity simplifies evolution: adding or redefining a class requires only the relevant specialist and threshold updates, not full re-training. Data quality remains the dominant lever: coverage across lighting, contamination, and scale governs generalization more than incremental architectural tweaks. Continuous dataset stewardship – sampling from production, rapid relabeling of failure clusters, cautious pseudo-labels – and lightweight domain adaptation (batch-norm recalibration or shallow adapters) provide predictable improvements without disruptive retraining.

In sum, a two-stage neural pipeline for photo-based C&D waste recognition proves viable in realistic settings. Pairing scene-level detection with per-class residual refinement and a simple arbitration rule yields recall gains on confusable materials while preserving precision, reducing misroutes and stabilizing downstream processing. Future work will emphasize multi-label handling of mixed fragments, instance-level segmentation to reduce boundary truncation, and light-touch domain adaptation for rapid cross-site transfer, supporting circular-economy operations through more reliable visual triage.

References:

1. M. Molchanova, V. Didur, O. Mazurets, O. Sobko, O. Zakharkevich, Method for Construction and Demolition Waste Classification Using Two-Factor Neural Network Image Analysis, in: CEUR Workshop Proceedings, 3970, 2025, pp. 168–182.
2. O. Mushtyn, O. Sobko, M. Molchanova, O. Mazurets, Convolutional Neural Network Architecture for Image-Based Architectural Style Recognition, in: Evolving Science: Theories, Discoveries and Practical Outcomes, Proceedings of the 4th International Scientific and Practical Conference, Zurich, Switzerland, 2025, pp. 130–143.

Taras Melnyk

Bachelor in Software Engineering, master's degree student of the Faculty of Computer Science and Technology

Yevhen Tatarynov

PhD in Math, associate professor of the Faculty of Computer Science and Technology

State University "Kyiv Aviation Institute"

APPLICATION OF LARGE LANGUAGE MODELS FOR ANALYZING THE COMPLIANCE OF PROGRAM CODE WITH SOLID PRINCIPLES

Abstract. This paper assesses the effectiveness of modern large language models (LLMs) in identifying compliance with SOLID principles in C# code. Twenty code samples were analyzed using several OpenAI models. Accuracy, precision, recall, and F1-score were used as evaluation metrics. Results show that while faster models provide higher formal accuracy, more advanced ones deliver deeper architectural reasoning. Future work should focus on improved prompting, LLM + RAG integration, specialization for code analysis, and code smell detection.

Keywords: SOLID, principle, quality, large language model, GPT, C#, review, architecture.

In modern software development, increasing attention is paid to the quality of software systems and adherence to architectural principles. Automating code evaluation using large language models (LLMs) appears to be a promising approach for improving code cleanliness and enhancing the development process [1].

Today, several software design principles exist to define the correct direction for building scalable, maintainable, and high-quality systems. Among these, the SOLID principles are considered the most effective and widely adopted [2].

The goal of this study is to present the results of evaluating how effectively modern large language models can identify violations or compliance with SOLID principles in program source code.

For this study, 20 C# code samples were selected that either comply with or violate SOLID principles. The dataset includes simple code fragments as well as archived full-scale .NET reference projects. Each example was manually assessed for SOLID compliance.

The following OpenAI models were used for evaluation: GPT-4o, GPT-4.5, GPT-4.1, and o4-mini-high [3]. Two prompts were developed: one for a code snippet and one for a full project archive. All queries were executed in anonymous, temporary sessions to avoid contextual memory effects.

In cases of ambiguous responses ("Partially"), results were determined based on the quality of reasoning. Model outputs were compared against predetermined ground truth values. Four evaluation categories were used:

1. True Positive (TP): the LLM correctly indicated that the principle is followed;

2. True Negative (TN): the LLM correctly indicated that the principle is violated;

3. False Positive (FP): the LLM indicated compliance, though the principle is violated;

4. False Negative (FN): the LLM indicated violation, though the principle is followed.

Based on these values, the following metrics were calculated:

1. Accuracy – overall correctness of the model;
2. Precision – proportion of correct "compliant" predictions;
3. Recall – how many actual compliant cases were correctly detected;
4. F1-Score – harmonic mean of precision and recall, balancing FP and FN rates.

Table 1 presents the metric values obtained for each evaluated model.

Model	Accuracy	Precision	Recall	F1-score
GPT-4o	87%	97%	85%	91%
GPT-4.5	83%	98%	82%	89%
GPT-4.1	91%	93%	95%	94%
o4-mini-high	67%	59%	94%	72%

Table 1 – Quantitative results of the study

GPT-4o responded quickly but analyzed only a tiny portion of the input. GPT-4.5 performed a much deeper analysis and handled a significantly larger context, providing more comprehensive conclusions. GPT-4.1 behaved similarly to GPT-4o – fast but rather superficial.

In terms of argumentation and analysis, o4-mini-high performed on par with GPT-4.5, if not better, while being considerably faster. However, it appears to be "overtrained" – it attempts to perfect the code even where this is unnecessary, which negatively affects its evaluation metrics.

GPT-4.1 achieved the highest formal metrics, while GPT-4.5 and o4-mini-high showed better reasoning quality and depth of architectural insight. This allows us to conclude that the choice of model depends on objectives: if speed and accuracy are priorities, GPT-4o or GPT-4.1 should be preferred; if high-quality, in-depth code auditing is required, GPT-4.5 or o4-mini-high are more suitable.

Based on the results obtained, several directions can be proposed for improving the effectiveness of using LLMs for SOLID compliance and code quality verification:

1. Prompt engineering

Testing more structured prompts, for example, asking separately about each SOLID principle.

2. LLM + RAG architecture

Creating a system combining LLMs with Retrieval-Augmented Generation to supply additional external knowledge [4].

3. Domain-specific models

Fine-tuning models dedicated to code quality analysis. Such models, though less universal, can more effectively detect common anti-patterns.

4. Detection of Code Smells

Focusing on specific symptoms of poor design (code smells), many of which directly relate to SOLID. This approach is more reliable due to its reliance on measurable code characteristics (size, duplication, dependency count, hierarchy depth, etc.).

In summary, the study confirms that LLMs can serve as an efficient tool for automating architectural compliance checks; however, they should be used carefully and complemented by traditional techniques currently. Future research should focus on expanding the capabilities of such models.

References:

1. Martins, G.F., Firmino, E.C.M., De Mello, V.P. (2024). The Use of Large Language Model in Code Review Automation: An Examination of Enforcing SOLID Principles. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2024. Lecture Notes in Computer Science(), vol 14736. Springer, Cham. doi:10.1007/978-3-031-60615-1_6.
2. Robert C. Martin, Design Principles and Design Patterns, 2000. URL: http://objectmentor.com/resources/articles/Principles_and_Patterns.pdf.
3. OpenAI ChatGPT, 2025. URL: <https://chatgpt.com>.
4. What is RAG (Retrieval-Augmented Generation)?, 2025. URL: <https://aws.amazon.com/what-is/retrieval-augmented-generation>.

Viktoriia Moiseienko

4th year bachelor of Computer Science student, Faculty of Information Technologies
vika.moiseienko@knu.ua

Olena Fedusenko

Candidate of Engineering Sciences (Ph. D.)

fedusenko@knu.ua

Taras Shevchenko National University of Kyiv

FORECASTING FOOD PRICES USING ARIMAX MODEL

Abstract. This paper explores the use of the ARIMAX model for food price forecasting. The approach extends the traditional ARIMA framework by incorporating external macroeconomic factors such as inflation, crop yield, and fuel costs. This allows modeling the influence of external economic conditions on food price dynamics, enhancing the interpretability and flexibility of the forecasting process.

Keywords: food prices; ARIMAX; forecasting; time series; macroeconomic factors

Food prices are among the most volatile economic measures and are determined by seasonal impact, global trade, and domestic production costs. Therefore, it's very important to develop precise food price forecasting models for successful market planning.

Several studies have applied statistical and machine learning methods to forecast food prices. For example, some authors used ARIMA models for agricultural commodity price prediction, while others extended their analysis to include seasonal effects with SARIMA [3]. Neural-network based methods have also been explored, demonstrating improved predictions but often lacking interpretability [4].

However, traditional time series models often fail to account for macroeconomic and external shocks that significantly influence price dynamics.

Recent advances in forecasting show that hybrid models like ARIMAX can both utilize past prices as well as macroeconomic variables beyond the price series [1]. The purpose of this study is to explore how effective the ARIMAX model can be for forecasting food prices and how it compares to the standard ARIMA baseline.

Traditional time series models like ARIMA (AutoRegressive Integrated Moving Average) are strong for univariate data but they neglect external shocks and macroeconomic influences. To overcome this, the ARIMAX (ARIMA with eXogenous inputs) model was utilized, its general form contains one or more external variables and therefore comes into use to model price movements under the impact of forces outside the historical price data itself.

Historical prices of selected food items were collected from official Ukrainian statistical data sources [2]. The following exogenous variables (X_t) were utilized in the model to enhance predictive accuracy: crop yield, inflation, fuel costs, market concentration index, consumer demand, seasonality indices.

The ARIMAX model extends ARIMA by incorporating these predictors, following the general form [1]:

$$Y_t = c + \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^m \beta_k X_{t-k} + \epsilon_t$$

Model forecasts the price Y_t by combining the influence of its own past values ($\varphi_i Y_{t-i}$) and past prediction errors ($\theta_j \epsilon_{t-j}$) with the direct impact of external economic factors X_{t-k} (the exogenous component). The coefficients β_k quantify the effect of these external factors, such as inflation or fuel costs, on the final price forecast.

This work was able to utilize the ARIMAX approach effectively to generate robust, data-driven projections of food commodity prices. The results validate the belief that the inclusion of suitable macroeconomic variables has a significant effect on the forecast ability of standard time series models.

References:

1. GeeksforGeeks. What Is an ARIMAX Model? - GeeksforGeeks. GeeksforGeeks. URL: <https://www.geeksforgeeks.org/artificial-intelligence/what-is-an-arimax-model/>.
2. Держстат України. Новини. 2025. URL: <https://ukrstat.gov.ua/>.
3. Food price index prediction using time series models: A study of Cereals, Millets and Pulses / Santosh Kumar Majhi et al. URL: <https://assets-eu.researchsquare.com/files/rs-2999898/v1/e814577c-a3d3-44d8-a096-dd028cf55575.pdf?c=1714040925>
4. Department of Agricultural Economics Aristotle University of Thessaloniki 54124 Thessaloniki Greece. Predicting Prices of Staple Crops Using Machine Learning: A Systematic Review of Studies on Wheat, Corn, and Rice. MDPI. URL: <https://www.mdpi.com/2071-1050/17/12/5456>.

Oksana Moisiuk

Master's Degree Student, Department of Technology Management

Anna Kolomiets

PhD, associate professor of the department of technologies management

Taras Shevchenko National University of Kyiv

DEVELOPMENT OF AN AI-POWERED APPLICATION FOR SMART DINING DISCOVERY AND PERSONALIZED TABLE RESERVATION

Abstract. Modern urban lifestyles demand efficiency, personalization, and digital assistance in everyday activities — including dining. This paper presents a mobile application project titled DineHub, which leverages artificial intelligence to enhance restaurant discovery, automate table reservations, and personalize dining experiences. The proposed system integrates machine learning, predictive analytics, and natural language processing to recommend restaurants, predict crowd levels, and adapt to user preferences.

Keywords: artificial intelligence, machine learning, mobile application, recommender systems, dining management, dining experience.

Introduction. The dining industry has a rapid digital transformation, yet most existing restaurant booking systems remain limited to static reservation functionalities. Platforms such as OpenTable, TheFork, and Expireenza provide basic booking interfaces but lack real-time personalization, predictive capacity management, or AI-based feedback interpretation. Consequently, users often face choice overload and restaurants struggle with uneven table utilization and low customer retention [6]. Therefore, it becomes essential to design a next-generation mobile application that unifies AI-powered personalization, predictive analytics, and seamless user experience into one ecosystem.

The aim of this work is to create DineHub, a next-generation mobile platform that reimagines how people choose restaurants and reserve tables. The core idea of the project is to merge artificial intelligence with user-centered design, creating a dining experience that feels personal, intelligent, and effortless. Rather than acting as a simple booking tool, DineHub is intended to function as a digital companion that learns from user preferences, adapts to individual habits, and offers relevant suggestions in real time [2, 5].

To achieve this purpose, the system integrates advanced data-driven methods capable of recognizing behavioral patterns and predicting future choices. Machine learning models analyze user activity, restaurant data, and contextual factors such as location and time of day to deliver refined, accurate recommendations. Predictive components forecast seat availability and expected waiting times, helping users plan their visits while supporting restaurants in maintaining optimal occupancy levels [4]. Natural language processing is employed to interpret user feedback and sentiment, allowing the platform to evolve dynamically as more data is collected [1, 3].

The DineHub platform is deployed within the AWS ecosystem to ensure flexibility, security, and scalability. The client interface is developed with Flutter, providing a responsive cross-platform experience across Android, iOS, and web environments. The backend operates on Spring Boot microservices. Core AI

capabilities are powered by AWS Bedrock, which provides managed access to advanced foundation models for recommendation, personalization, and natural language processing.

This combination of cloud-native architecture, AI-as-a-service infrastructure, and lightweight frontend design ensures that DineHub remains scalable, intelligent, and user-centered

Results. The deployment of the DineHub platform confirmed the potential of artificial intelligence to redefine digital dining experiences. The system seamlessly unites personalized recommendation algorithms, sentiment-aware feedback analysis, and adaptive reservation management within a single intelligent environment.

Pilot testing revealed a measurable improvement in the speed and accuracy of restaurant selection, as well as higher satisfaction with recommendation relevance. Users emphasized the platform's ability to adapt dynamically to their changing habits and preferences, creating a sense of intuitive personalization. The integrated machine learning models accurately forecasted booking demand, while natural language processing refined restaurant rankings in real time based on evolving sentiment patterns.

Conclusions. The research demonstrates that DineHub effectively applies AI to deliver a more intuitive, efficient, and personalized dining experience. Through continuous learning from user behavior and feedback, DineHub evolves dynamically, offering increasingly relevant suggestions and enhancing the overall interaction between diners and restaurants.

References:

1. Deloitte Insights, How AI is Revolutionizing Restaurants, 2025. URL: <https://www.deloitte.com/us/en/insights/industry/retail-distribution/ai-in-restaurants.html>
2. SevenRooms, Restaurant Trends and Diner Expectations, 2024. URL: https://go.sevenrooms.com/rs/519-YNM-008/images/US_Restaurant_Trends_Report_2024_FINAL.pdf
3. Deloitte US, Building restaurant resilience, 2024. URL: <https://www.deloitte.com/us/en/Industries/consumer/articles/future-of-restaurants-and-ai.html>
4. Emerging Insights, AI and the Restaurant Industry in 2025, 2025. URL: <https://emerging.com/insights/emerging/ai-and-the-restaurant-industry-in-2025>
5. Restaurant Technology News, Research: 70% of restaurants either actively using or piloting AI to boost loyalty programs, 2025. URL: <https://restauranttechnologynews.com/2025/06/research-70-of-restaurants-either-actively-using-or-piloting-ai-to-boost-loyalty-programs/>
6. M. G. Dancausa Millán & M. G. Millán Vázquez de la Torre. An Economic Perspective on the Implementation of Artificial Intelligence in the Restaurant Sector. Administrative Sciences, Vol. 14, No. 9, 2024. DOI: 10.3390/admsci14090214

Denys Muzyka

Bachelor in Software Engineering, Master Student

Stanislav Marchenko

PhD Student at Department of Automated Systems Software

Andrii Yarmilko

PhD in Information Technologies, Docent, Associate Professor at Department of Automated Systems Software

Bohdan Khmelnytskyi National University of Cherkasy

POLYMODEL VIDEO STREAM COMPRESSION IN COMPUTER VISION SYSTEMS

Abstract. The paper considers approaches to improving the efficiency of image processing in applied computer vision systems through software-based maximization of informative image features. A polymodel compressor combining various filters and algorithms for adaptive preservation of key features during data compression is proposed. The developed approach ensures an optimal balance between processing speed, data volume, and image quality, increasing the accuracy of classification and segmentation systems.

Keywords: computer vision, image preprocessing, video stream compression, polymodel compression, informative features, filters, segmentation.

In computer vision applications such as real-time industrial systems, classification sensor systems, and medical imaging, where image processing accuracy and speed are crucial, software-based maximization of informative features of raster images is widely used [1]. Appropriate methods and software tools are required for the preprocessing of visual input data, particularly for image classification and segmentation tasks [2; 3]. The expected result of implementing such tools is the minimization of the volume of processed data while simultaneously increasing their informativeness. The scientific and technical complexity of this problem is caused by anomalies and the difficulties in determining the structure of large volumes of visual data. Nevertheless, solving this problem would stimulate the further expansion of visual analysis in applied technologies [4].

Thus, research in the field of software-based maximization of informative features of raster images is relevant in the context of overcoming the challenges of computer vision technologies related to the growing volume of visual data and heterogeneity in their structure. An obvious task of image preprocessing is to improve the overall image quality to obtain more accurate results. However, during processing, it is also possible to impart specific properties to an image that are required by a particular application [5]. For example, reducing the overall size of all dataset images can significantly save computing time and resources. On the other hand, certain preprocessing methods can improve the efficiency of working with more complex patterns. However, when applying specific filters or other feature maximization methods to raw data, there is a likelihood of unpredictable anomalies that can affect the result. Therefore, preprocessing of video data should be performed even in the absence of apparent image defects.

Video stream compression is one of the key technologies used to reduce data volume [4]. In the processes of storing and transmitting video data, it is essential for minimizing resource consumption and increasing the speed of information processing. One of the main problems that arises during compression is the loss of important image details. Despite the overall optimization of most common video

processing algorithms for size reduction, their operation often results in the loss of details such as object edges, textures, and low-contrast elements. As a result, identification quality in classification and segmentation tasks deteriorates, and information about small image details is lost [6].

As a solution to this problem, a polymodel compressor is proposed, providing an integrated use of various tools for detecting and preserving essential image characteristics in the video stream, which is critical for applied computer vision systems [2]. The proposed application can be configured for specific computer vision tasks, ensuring the preservation of the most important information for each particular use case. This is achieved by configuring and combining filters, their selective application to individual images or regions, and real-time image quality control with the ability to adjust filter parameters accordingly [3].

Particular attention is paid to the problem of losing essential details of the original image, such as object edges, textures, and low-contrast details, which can significantly affect the efficiency of computer vision systems in various application areas, including facial recognition, traffic monitoring, surveillance systems, sports analysis, and medical diagnostics. Traditional compression methods optimized for the human eye often do not consider the specific needs of computer vision algorithms, highlighting the necessity of developing new compression methods such as polymodel compression [4]. Methods such as thresholding, erosion, dilation, opening, closing, and normalization help reduce defects, equalize tones, and minimize the impact of incorrect data [7]. The use of multiple models of filters and image processing allows combining the advantages of different methods, adapting and integrating approaches for specific needs, and optimizing images for computer vision [3].

The ability to reproduce video with filters that search for informative features ensures a high level of preservation of crucial information, which is essential for accurate analysis and effective operation of computer vision systems. In general, the issues discussed in the study demonstrate the importance of input data quality for achieving high accuracy and efficiency in computer vision systems. The development of new processing and compression methods that ensure the preservation of key image details is critical for the further evolution and improvement of these technologies [4; 7].

References:

1. How features are represented in a raster – ArcMap | Documentation. URL: <https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/how-features-are-represented-in-a-raster.htm>.
2. Haralick R. M., Shanmugam K., Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973. SMC-3(6). P. 610–621. DOI: 10.1109/tsmc.1973.4309314.
3. OpenCV: OpenCV-Python Tutorials. OpenCV documentation index. URL: https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html.
4. Ghanbari M. *Standard Codecs: Image Compression to Advanced Video Coding*. 3rd ed. Institution of Engineering & Technology, 2011. <https://doi.org/10.1049/PBTE054E>.
5. Python interface to Tcl/Tk. Python documentation. URL: <https://docs.python.org/uk/3/library/tkinter.html>.
6. Gonzalez R. C., Woods R. E. *Digital Image Processing*. 4th ed. Pearson, 2018.
7. Soille P. *Morphological Image Analysis: Principles and Applications*. 2nd ed. Springer, 2004.

Oleksii Paharskyi

Master in Computer Science, PhD Student, Faculty of Information Technologies

Georhii Gaina

PhD in Technical Sciences, Professor, Professor of the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

LARGE LANGUAGE MODEL MULTI-AGENT SYSTEM FOR PROGRAMMING

Abstract. This paper explores a multi-agent system to address such issues of variable code quality and a lack of reliable verification, aiming to improve the reliability of LLM-based code generation. We focus on an architecture composed of a Programmer Agent, a Tester Agent, and an Executor Agent that interact through an iterative 'test-driven refinement' process. Our results from experiments on 11 modern LLMs show that this multi-agent approach improves code generation quality in most cases. These findings demonstrate the potential of multi-agent systems to create more robust intelligent developer assistants, addressing key limitations of single-model approaches.

Keywords: multi-agent systems, large language models, code generation, prompt engineering, pass@1, test-driven refinement.

The rapid development of information technologies and the increasing volume of software require new development approaches. AI-based systems, particularly large language models (LLMs), are becoming key tools for automating routine operations and improving development quality.

Popular tools like GitHub Copilot [1] and ChatGPT [2] demonstrate significant AI potential in code generation. However, they have substantial limitations, including: variable quality and correctness of generated code, data privacy concerns with cloud services, and a lack of specific context for complex tasks.

The purpose of this work is to research and develop a multi-agent system based on LLMs that would solve these problems by implementing an iterative approach to code generation and verification, ensuring an increase in its quality and reliability.

Methodology and System Architecture

The proposed system is based on a multi-agent architecture that decomposes the complex task of code writing into three specialized roles:

1. **Programmer Agent:** Responsible for generating the initial code and its subsequent improvement based on feedback.
2. **Tester Agent:** Independently generates a set of test cases to verify the code's correctness.
3. **Executor Agent:** A Python script that runs the generated tests and returns detailed error reports.

The workflow is based on the "test-driven refinement" principle [5]. The Programmer Agent generates code; the Tester Agent creates tests; the Executor Agent runs them. If tests fail, error messages and failure details are passed to the Programmer Agent, which enters an iterative code refinement loop.

Advanced prompt engineering techniques are used for effective agent management: "Chain-of-Thought" for the Programmer Agent (forces the model to plan the solution in detail) and role-playing with examples (few-shot) for the Tester Agent (helps generate relevant and edge-case tests).

The system's effectiveness was evaluated on 11 modern LLMs, including StarCoder2, CodeLlama (34B), GPT-4o-mini, GPT-o3-mini, Claude 3.7 Sonnet, Gemini 2.5 Flash, and DeepSeek-R1. Four recognized benchmarks for Python code generation were used: HumanEval [3], MBPP [4], APPS, and DS-1000. The primary quality metric chosen was **pass@1**, which evaluates functional correctness on the first attempt.

The results (Table 1) demonstrate that the proposed multi-agent approach consistently improves the pass@1 score for most models on all benchmarks.

Model	HumanEval	HumanEval (multiagent)	MBPP	MBPP (multiagent)	APPS	APPS (multiagent)	DS-1000	DS-1000 (multiagent)
StarCoder (15B)	18.9	34.1	26.9	43.6	15.2	20.3	18.5	25.3
StarCoder2 (15B)	46.3	52.2	66.2	70.7	29.5	35.4	33.8	41.5
CodeLlama (34B)	44.6	51.8	46.1	69.3	35.1	40.6	40.5	48.6
GPT-4o-mini	57.9	54.7	77.4	91.4	50.3	58.7	55.1	62.8
GPT-4.1-mini	60.5	70.2	79.2	88.2	53.1	60.7	58.9	65.8
GPT-o3-mini	89.4	92.4	92.1	96.3	65.8	70.8	70.3	75.9
Claude 3.5 Sonnet	78.1	83.2	85.3	88.9	58.2	63.6	63.7	68.3
Claude 3.7 Sonnet	82.4	88.1	88.9	94.5	61.5	67.5	67.2	71.7
Gemini 2.0 Flash-lite	70.2	75.8	80.5	84.2	54.3	60.6	59.8	68.1
Gemini 2.5 Flash	74.3	78.4	83.1	91.0	56.7	62.1	62.4	73.8
DeepSeek-R1 (671B)	73.78	77	82.5	90.8	60.1	65.8	65.9	74.1

Table 1 – Evaluation results

Key findings from the experiments:

- **Approach Effectiveness:** The multi-agent system improved pass@1 results for all open-source models (StarCoder, StarCoder2, CodeLlama), showing the greatest increase for models with lower baseline scores.
- **Model Dependency:** A direct correlation was confirmed between model size, its specialization in code (CodeLlama), and generation quality.
- **"Diminishing Returns" Effect:** For the most powerful "thinking" models (GPT-o3-mini, Claude 3.7 Sonnet, DeepSeek-R1), which already have high baseline scores, the gain from the multi-agent approach was less significant. This suggests these models may internally perform similar verification steps.

These findings demonstrate the potential of MAS for creating more reliable intelligent developer assistants. The practical value lies in the possibility of increasing developer productivity and deployment flexibility, particularly local deployment, to ensure codebase confidentiality.

References:

1. The AI editor for everyone, GitHub. – <https://github.com/features/copilot>.
2. Introducing ChatGPT, OpenAI, 2022 – <https://openai.com/index/chatgpt/>.
3. Chen, M., Tworek, J., Jun, H., et al. Evaluating Large Language Models Trained on Code // arXiv. – 2021. – arXiv:2107.03374.
4. Austin, J., Odena, A., Nye, M., et al. Program Synthesis with Large Language Models // arXiv. – 2021. – arXiv:2108.07732v1.
5. Shinn, N., Tang, F. F., Liu, Y., et al. Test-Driven Development for Code Generation // arXiv. – 2024. – arXiv:2402.13521v1.
6. Ding, N., Zhou, Y., Liu, G., et al. AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation // arXiv. – 2023. – arXiv:2312.13010.
7. Ding, N., Chen, Y., Wang, R., et al. MapCoder: Multi-Agent Code Generation for Competitive Problem Solving // arXiv. – 2024. – arXiv:2405.11403.

Andrii Pavlov

Ph.D., Senior researcher

Institute of Information Technologies and Systems, Kyiv 03187, Ukraine

A THEORY-BASED DECISION FRAMEWORK FOR ENCODING CATEGORICAL VARIABLES IN REGRESSION AND CLASSIFICATION PROBLEMS

Abstract. This work, for the first time, states the main problem of encoding information to transform categorical variables into numeric form. We describe and analyze the primary target-agnostic encoding methods to demonstrate their accuracy at solving the stated problem. A new classification of categorical variables that considers data relations represented by semantic overlap and availability of an attribute is proposed. We analyze the most widely used target-based encoders to find the most robust ones. Based on the analyses provided, we offer a new decision framework for encoding categorical variables in regression and classification problems grounded in theoretical foundations.

Key words: Categorical Variable, Encoding Information Problem, Theoretical approach.

Introduction. Many datasets in real-world problems, in addition to numeric variables, contain categorical variables. Since most machine learning (ML) models can't process categorical variables directly, there is a need to transform them into numeric representations. There are a lot of methods that can be applied to this problem. All of them are divided into target-agnostic and target-based methods [1, 2]. A question arises: which method to apply in which situation? A commonly used way is to apply them all and select the best performer as the result. Another approach is to develop a general guide [3] or framework [4] based on the results of applying those methods to a substantial set of datasets [3-8]. The main downside of this approach is that it does not account for the underlying theory behind the methods, thereby neglecting the original problem they are intended to solve. Thus, the problem of selecting the appropriate encoding method (encoder) is still relevant. This work is focused on developing an approach to encoding categorical variables based on the theory behind them and the main problem of encoding information: transforming the information from one space to another so that it is preserved.

Problem statement. Any categorical variable carries information meaningful for the modeling problem. In this work, we consider categorical variables where each value in the vector denotes a single entity. Depending on the type of categorical variable (ordinal or nominal), the carried information differs. An ordinal variable defines the ordinal relation between its values, while a nominal variable does not exhibit any relations between them. In this context, the main problem of encoding information is to transform the information represented by the relations among categories (groups) of a categorical variable from string space to numeric space, preserving it in accordance with the modeling purpose. There are two sources of information loss when transforming it: loss in volume and loss in accuracy. In this work, we aim to develop a framework to encode categorical variables that:

- a) is in concordance with the main problem of encoding information;
- b) accounts for the information loss due to the transformation;

- c) handles categorical data represented by semantic overlap and availability of an attribute;
- d) takes into account the representation level of encoded information;
- e) considers the encoders' application assumptions and the ML problem type (regression, classification);
- f) accounts for the current theoretical findings in the development of decision frameworks for encoding categorical variables.

Extended classification of categorical variables. Besides the data relations represented by ordinal and nominal types of categorical variables, groups of a categorical variable can exhibit semantic closeness. Another type of relations that groups can have is presence/absence of an attribute in the data. To cover these types, we offer a new classification shown in Figure 1.

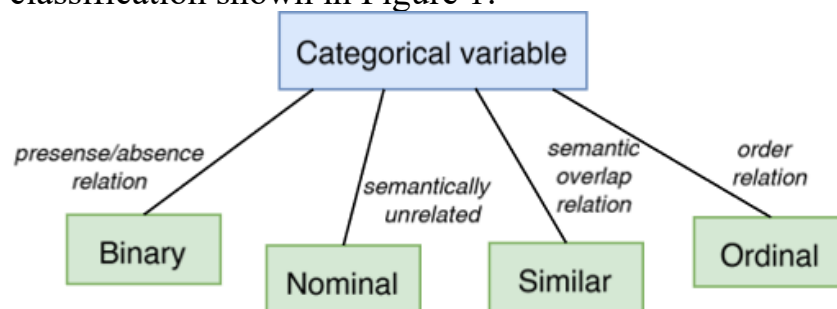


Figure 1 – An extended classification of categorical variables

The methodology of the research. Since target-based methods use information from the target, the tasks a)-d) refer to the target-agnostic methods. The methodology we use to fulfill a) and b) is to analyze the primary target-agnostic encoders, such as One-hot, Ordinal, Similarity, Hash, Min-hash, Count, and Frequency encoders, and show how accurate their transformations are and how well they maintain the original volume of the encoder information. The result of this analysis is shown in Figure 2.

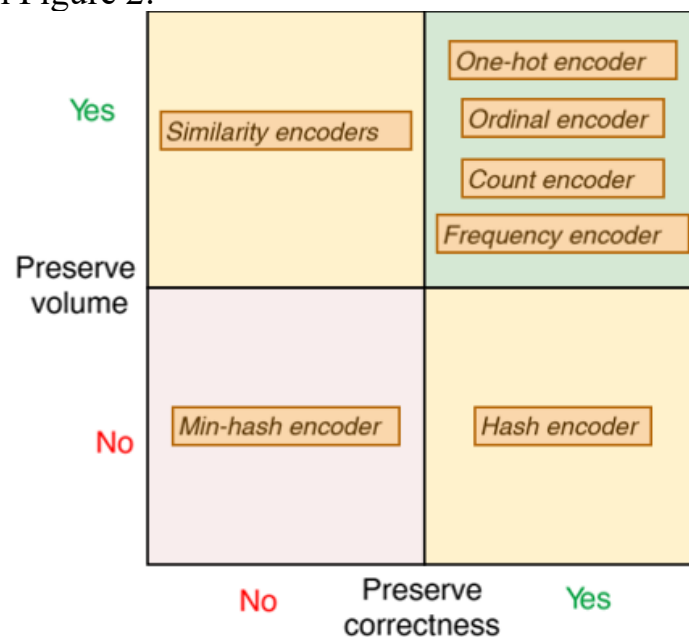


Figure 2 – The quality of solving the main problem of encoding information using target-agnostic methods.

Application of target-based (target) encoders is also relevant, in addition to target-agnostic ones, in supervised learning. Thus, there is a need to analyze these to identify the differences between them and outline their application domains. Our analysis showed that they historically evolve from the simplest one, Mean target encoder, to the most sophisticated one, CatBoost encoder. Each of the methods had its drawbacks, which the next generation of them was intended to eliminate. As a result of the analysis, we classified them according to the inherent approach they use to eliminate the drawbacks into: no approach (Weight of evidence, Mean target, GLMM), Bayes approach (S-shrink, M-estimate, James-Stain, Quantile), cross-validation approach (Leave-one-out, K-fold, CV-GLMM, CatBoost). Based on the overall analysis, we propose a new approach to encoding categorical variables, as illustrated in Figure 3.

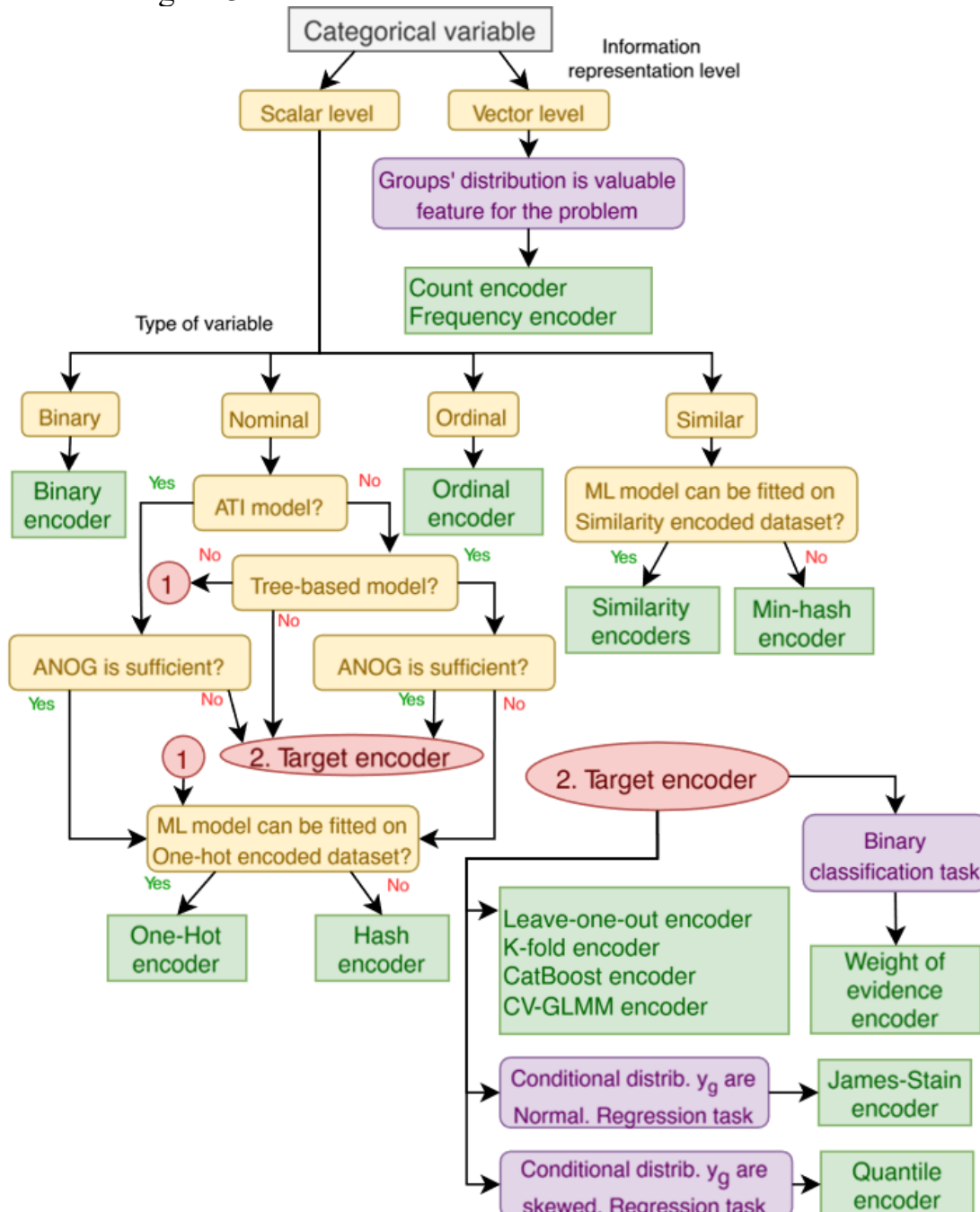


Figure 3 – A decision framework to encoding categorical variables.

The block ‘ML model can be fitted ...’ implies that a researcher should make a decision: accept the information losses or learn the ML model for much longer using the result of One-hot or a similarity encoder. The framework accounts for the theoretical results in [3]: encoding of a nominal variable depends on the ML model type and the Average Number of Observations per Group (ANOG). According to our analysis, K-fold target encoder, Leave-one-out target encoder, CatBoost and CV-GLMM encoder are the most robust ones. The usage of the other encoders is based on their specific assumptions and conditions, which were discovered as a result of our analysis.

Results and Conclusion. In this research we: state the problem of encoding information for transforming categorical variables into numeric form; classify target-agnostic methods based on their ability to maintain the encoded information; offered an extended classification for categorical variables that accounts for semantic overlap and availability of an attribute in the data; analyze target-based methods and classify them based on the primary approach they use, and indicate the most robust from them; offered a new theory-based decision framework for encoding categorical variables in regression and classification problems. In conclusion, we recommend using the offered framework for encoding categorical variables in practice.

References:

1. Pargent, Florian, et al. "Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features." *Computational Statistics* 37.5 (2022): 2671-2692.
2. Breskuvienė, Dalia, and Gintautas Dzemyda. "Categorical feature encoding techniques for improved classifier performance when dealing with imbalanced data of fraudulent transactions." *International Journal of Computers Communications & Control* 18.3 (2023).
3. Zhu, Wenbin, Runwen Qiu, and Ying Fu. "Comparative study on the performance of categorical variable encoders in classification and regression tasks." *arXiv preprint arXiv:2401.09682* (2024).
4. Bolikulov, Furkat, et al. "Effective methods of categorical data encoding for artificial intelligence algorithms." *Mathematics* 12.16 (2024): 2553.
5. Matteucci, Federico, Vadim Arzamasov, and Klemens Böhm. "A benchmark of categorical encoders for binary classification." *Advances in neural information processing systems* 36 (2023): 54855-54875.
6. Seca, Diogo, and João Mendes-Moreira. "Benchmark of encoders of nominal features for regression." *World Conference on Information Systems and Technologies*. Cham: Springer International Publishing, 2021.
7. Prokopev, V.: Mean (likelihood) encodings: A comprehensive study, <https://www.kaggle.com/code/vprokopev/mean-likelihood-encodings-a-comprehensive-study>, (2018), last accessed 2025/09/15.
8. Potdar, Kedar, Taher S. Pardawala, and Chinmay D. Pai. "A comparative study of categorical variable encoding techniques for neural network classifiers." *International journal of computer applications* 175.4(2017): 7-9.

Denys Popov

Master's Degree Student, Department of Technology Management

Anna Kolomiets

PhD, associate professor of the department of technologies management

Taras Shevchenko National University of Kyiv

DEVELOPMENT OF AN AI-DRIVEN APPLICATION FOR NUTRITION MONITORING AND HEALTHY HABIT FORMATION

Abstract. Unbalanced nutrition is a leading health risk factor, causing obesity and chronic diseases. This paper presents a mobile application project that uses AI to analyze meal photos, calculate calories and nutrients, and provide personalized recommendations. The proposed solution combines AI-based analysis and gamification to promote sustainable healthy habits and increase users' nutritional awareness.

Keywords: artificial intelligence, mobile application, machine learning, nutrition tracking, information technology.

Introduction. The relevance of the problem of unbalanced nutrition and its impact on human health is global. According to international studies, overweight and nutrient deficiencies cause millions of premature deaths each year [1]. Despite the popularity of digital nutrition solutions (MyFitnessPal, Calorie AI, YAZIO), most of them are characterized by limited functionality, low accuracy in food recognition, and insufficient personalization, which reduces their long-term effectiveness [4, 5]. This justifies the need to develop an innovative tool to support healthy eating habits using modern information technologies.

The aim of this work is to develop an intelligent mobile application that integrates image analysis technologies, personalized recommendations, and elements of gamification. The main objectives are as follows: to integrate machine learning methods (LLM models) for accurate meal recognition from photos [3]; to apply mathematical modeling (Harris–Benedict and Mifflin–St Jeor formulas) for calorie calculation; to design a food quality index (Nutri-Score, IQF) as a comprehensive indicator of nutritional value [6]; to develop the application architecture and database for efficient data storage and processing; and to implement user motivation mechanisms through gamification [4, 5].

An interdisciplinary approach was applied, combining domain analysis, mathematical modeling, algorithmic design, and system architecture development.

The client side of the application is implemented using the Flutter framework, which ensures cross-platform compatibility for iOS and Android [2]. Backend services are built in Python using gRPC for high-performance communication. The developed logical database model provides structured storage of information about users, nutrition history, meals, and their ingredients.

For the mathematical component, algorithms for calculating calories and macronutrients were applied, along with the integrated IQF indicator combining NutriScore, nutrient balance, fiber, and salt levels [6]. To enhance accuracy, machine learning models were used, including convolutional neural networks (CNN) for visual recognition and large language models (LLM) for text-based meal analysis [3].

Results. A prototype of the mobile application *Calorythm* has been developed, providing the following features: accurate recognition of meals based on photos and text descriptions; automatic calculation of caloric value and nutrients; personalized recommendations for improving dietary habits; storage of nutrition history and visualization of user progress; and gamification elements in the form of a virtual pet whose condition depends on the user's adherence to dietary goals.

Prototype testing demonstrated high recognition accuracy, positive user feedback, and the feasibility of a subscription-based monetization model.

The proposed approach eliminates key shortcomings of existing solutions. The use of LLM models ensures higher accuracy and personalization compared to *MyFitnessPal* and *YAZIO* [3]. Integration of gamification elements increases user engagement, which is a crucial factor in the long-term development of healthy habits. Unlike most existing solutions, the application supports comprehensive nutritional value analysis and provides scalability through integration with fitness trackers.

Conclusions. The developed AI-powered mobile application is an innovative tool for promoting healthy eating habits. Combining image analysis, mathematical modeling, and gamification ensures both scientific novelty and practical value. The prototype proved the effectiveness of the approach; future work will expand the food database and integrate health monitoring devices, confirming its potential for commercial use in digital healthcare.

References:

1. Global Nutrition Report, *Progress Towards the Global Nutrition Targets*, 2024. URL: <https://globalnutritionreport.org/progress-towards-the-global-nutrition-targets>.
2. Flutter Documentation, *Architecture Design Patterns*, 2025. URL: <https://docs.flutter.dev/app-architecture/design-patterns>.
3. Z. Zeng, M. Song, D. Wang, et al., Research on food image recognition based on improved CNN, in: *IEEE International Conference on Information Technology and Management Engineering*, 2023. doi:10.1109/ICITME58763.2023.10230230.
4. M. B. Bayer, L. Biron, J. Haase, et al., Gamification in digital health: a systematic review, *IEEE Journal of Biomedical and Health Informatics*, 2023. doi:10.1109/JBHI.2023.3256789.
5. MDPI, Mobile healthcare applications and gamification for sustained health maintenance, 2024. URL: <https://www.mdpi.com/2071-1050/9/5/772>.
6. AGROLAB GROUP, Adjustment of the Nutri-Score algorithm, 2024. URL: <https://www.agrolab.com/en/news/food-news/4842-nutriscore-update-radar-01-2024-en.html>.

Vladyslav Pylypenko

PhD student at the Department of Computer Science

Vladyslava Skidan

Ph.D., Associate Professor, Department of Information and Computer Technologies

Antonina Volivach

Ph.D., Associate Professor, Department of Information and Computer Technologies

*Kyiv National University of Technologies and Design 1, 2, 3***PREDICTING STUDENTS' SUCCESS AND GRADATION OF B USING MACHINE LEARNING**

Modern educational systems use digital technologies that allow collecting large amounts of data on students' learning activities, forming a multidimensional set of educational data. This opens up opportunities for the application of machine learning methods to predict their learning outcomes. An important factor is the ability to predict the level of academic success above average with a grade of "B" (82–89 points), as this is a step towards a high level of "A" (90–100 points). In the study, data were collected and processed from the university's electronic journal, the Moodle platform, and from a developed video player plugin for collecting data on interaction with video materials [1]. The dataset of 4013 user sample records was divided into training and test samples, which were distributed in the ratio of 803/3210. The training sample contained 3210, and the test sample contained 803. The data were obtained for 2024 and 2025 years of using this plugin in the Moodle environment for first-year undergraduate students of the Department of Information and Computer Technologies of the Kyiv National University of Technologies and Design [2]. As part of data preparation, a correlation matrix was formed based on 10 features that characterize the behavioral activity of students in Moodle courses and interaction with video content. The constructed confusion matrices are presented in Figure 1.

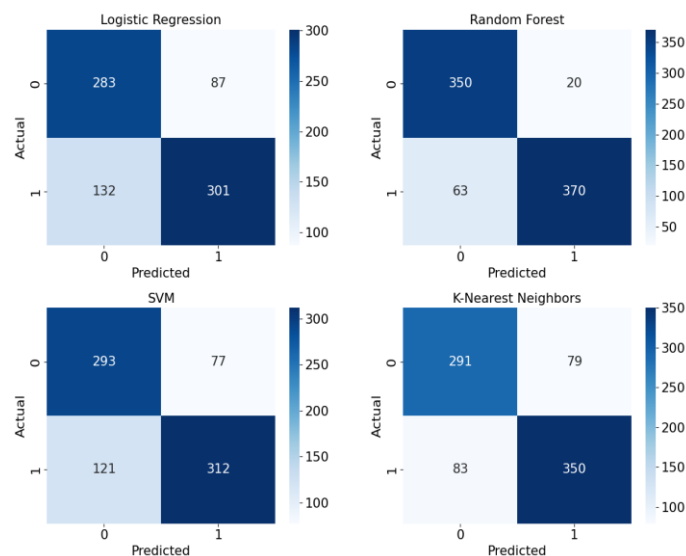


Figure 1 – Confusion matrices for the created models

The constructed ROC-curves for each model are presented in Figure 2.

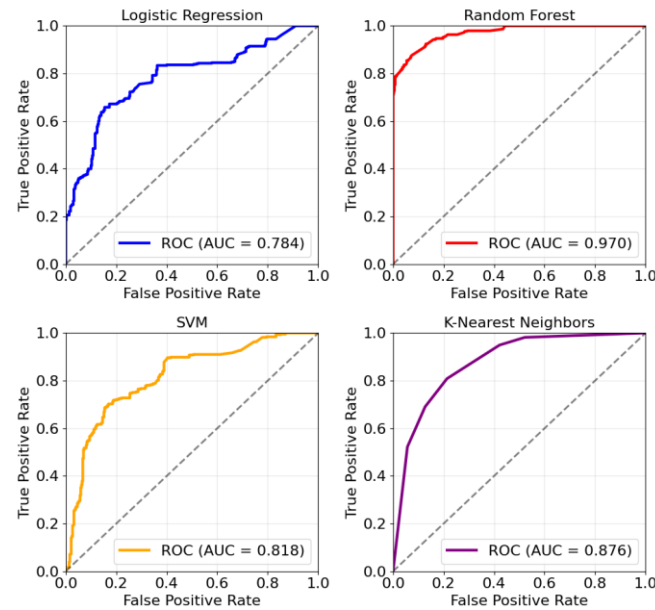


Figure 2 – ROC curves for the constructed prediction models

The obtained results of metrics characterizing the accuracy and efficiency of the model are presented in Table 2.

Table 2
Results of metrics characterizing the accuracy and efficiency of models

Algorithm	Accuracy	Balanced Accuracy	F1 Score	Sensitivity	Specificity	Precision	AUC
LR	0.72	0.73	0.73	0.69	0.7	0.78	0.78
Random Forest	0.89	0.90	0.90	0.86	0.95	0.95	0.97
SVM	0.75	0.76	0.76	0.720	0.72	0.80	0.82
KNN	0.80	0.79	0.81	0.80	0.78	0.82	0.87

Conclusions. The study showed that machine learning methods are effective for predicting student success. The best results among the four algorithms were shown by Random Forest. Its accuracy was 0.89, AUC-0.97, which indicates a high ability to distinguish classes. Random forest best balanced the errors of the first and second kind, which makes it suitable for practical use.

References:

1. Пилипенко В., Стаценко В. А ПЛАГІН ДЛЯ ЗБОРУ ДАНИХ ВЗАЄМОДІЇ КОРИСТУВАЧІВ MOODLE З ВІДЕО МАТЕРІАЛАМИ. Наука і техніка сьогодні. 2025. 1(42). Ст. 1318–1330. URL: [https://doi.org/10.52058/2786-6025-2025-1\(42\)-1318-1330](https://doi.org/10.52058/2786-6025-2025-1(42)-1318-1330).
2. Пилипенко В. ПРОГНОЗУВАННЯ ВИСОКОГО РІВНЯ АКАДЕМІЧНОЇ УСПІШНОСТІ СТУДЕНТІВ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ Наука і техніка сьогодні. 2025. 1(42). Ст. 1634–1649. URL: [https://doi.org/10.52058/2786-6025-2025-8\(49\)-1634-1649](https://doi.org/10.52058/2786-6025-2025-8(49)-1634-1649)

Oleksandr Redenskyi

2nd year Master's student, Department of Software Systems and Technologies

Kseniia Dukhnovska

PhD in Technical sciences, Assoc. Prof. of Software Systems and Technologies

Department

Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

HEURISTIC OPTIMIZATIONS FOR THE TRANSPORTATION PROBLEM IN URBAN CONDITIONS

Abstract. A method for accelerating the solution of the classical transportation problem in dense urban environments is presented. The approach leverages heuristic distance estimation aligned with city road grids and spatial decomposition of the service area into regions separated by major highways. The software stack employs Redis for fast geospatial indexing and lookups, and standard transportation solvers based on the Least-Cost initial solution and the MODI (potential) method for iterative refinement.

Keywords: transportation problem; urban logistics; route cost estimation; region-based decomposition; Least-Cost method; MODI method; Redis geospatial

Introduction. Urban logistics systems orchestrate the movement of goods from suppliers to consumers under tight time and infrastructure constraints. Information Logistics Systems (ILS) integrate software, hardware, and procedures to monitor flows, plan capacity, and coordinate transportation while interfacing with ERP/CRM platforms [1]. In modern city settings, the ability to compute feasible, cost-efficient allocations rapidly is crucial for last-mile scenarios, where solver latency directly impacts service quality and operational cost.

Background. The transportation problem (TP) is a special case of linear programming in which shipments from multiple sources must satisfy demands at multiple sinks at minimum cost. Although classical solvers exist, practical deployments face large, dynamic instances and must respond under real-time constraints. This work focuses on method and system design choices that leverage urban structure to reduce solve time without materially degrading allocation quality.

Let there be m suppliers with supplies $a_1, \dots, a_m$ and n consumers with demands $b_1, \dots, b_n$, and per-unit transportation costs c_{ij} . The objective is to find flows x_{ij} that satisfy supply/demand constraints and minimize total cost. Balanced instances ($\sum_i a_i = \sum_j b_j$) are particularly convenient; unbalanced instances admit standard transformations by adding dummy nodes.

Two well-established techniques are used operationally in this system: Least-Cost method to construct an initial basic feasible solution by prioritizing lowest-tariff cells, and MODI method to test optimality and iteratively improve the solution via reduced-cost cycles. Degeneracy is handled by ε -allocations to complete the basis ($(m+n-1)$ basic variables) and prevent ambiguous potentials.

Efficient access to locations, regions, and spatial predicates is a bottleneck in city-scale planning [2]. Redis is employed as an in-memory datastore with geospatial

capabilities. Redis supports point storage and fast queries, while RediSearch allows indexing polygonal regions (WKT) and evaluating predicates such as WITHIN, CONTAINS, and INTERSECTS [3]. This supports fast mapping of suppliers/consumers to regions and high-throughput lookups needed by the solver.

Main part. Exact road-network distances are expensive to compute repeatedly. In dense grids, shortest-path distances often approximate straight-line metrics. Therefore the costs are computed using a grid-consistent metric (Manhattan distance) as a surrogate for road travel. This reduces computational overhead for cost matrix construction while reflecting the anisotropy of urban movement along orthogonal roadways. The heuristic preserves relative ranking of many OD pairs, supporting high-quality initial allocations with substantially less compute.

Major highways and arterial corridors create de facto barriers and funnels in city mobility. In this work the city is partitioned into regions delimited by such features. The TP is then solved within each region first, substantially shrinking each subproblem. After intra-region demands and supplies are satisfied and corresponding rows/columns removed, a residual cross-region problem over remaining nodes is solved. This decomposition lowers effective problem dimensionality, improving asymptotic and realized runtime while respecting urban structure.

Unbalanced cases are transformed by adding dummy supply or demand to preserve algorithmic simplicity in MODI. ϵ -allocations are inserted to complete the basis and avoid cycles that would otherwise prevent unique potential assignment. The dominant costs arise from cost matrix construction and MODI iterations. Region decomposition reduces both by shrinking matrices early; Manhattan distances further reduce per-entry computation. Redis in-memory storage accelerates repeated spatial lookups and avoids disk I/O latency.

Empirically, replacing exact road distances with Manhattan distance and applying region-first decomposition yields faster solve times compared with monolithic formulations backed by disk-based stores. Using Redis alone provides some latency reduction, and combining Redis with decomposition delivers the largest improvement as instance size grows. While minor deviations from the true minimal-cost allocation can occur due to the distance surrogate, this is favorable for interactive, city-scale logistics planning where responsiveness is paramount.

Conclusion. A practical method and system for rapid logistics flow allocation in urban settings is described. By aligning cost estimation with road-grid structure and decomposing by arterial-defined regions, the approach achieves near-optimal allocations with markedly reduced computation. The architecture, which consists of Redis-backed geospatial access, heuristic cost matrices, and classical TP solvers, offers a balanced pathway from theory to deployable decision support for last-mile and intra-city logistics.

References

1. Sustainable urban logistics: Concepts, methods and information systems – Springer, 2013
2. Sheffi, Y. (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods / Sheffi, Y. – Prentice-Hal, 1984
3. Geospatial | Docs <https://redis.io/docs/latest/develop/interact/search-and-query/advanced-concepts/geo/>

Maksym Rudenko

Postgraduate Student

Mykyta Myronenko

Ph. D., Assistant Professor of the Department of Computer Sciences

Sumy State University

INFORMATION-EXTREME MACHINE LEARNING OF AN AUTONOMOUS ON-BOARD SYSTEM FOR MILITARY EQUIPMENT RECOGNITION

Abstract. The aim of this research is to improve the accuracy of machine learning within an intelligent on-board system of an autonomous unmanned aerial vehicle (UAV) designed for military equipment recognition. To enhance the functional efficiency of the intelligent on-board system, an algorithm of parallel-sequential optimization of selection levels is proposed.

Keywords: autonomous unmanned aerial vehicle, intelligent on-board system, military equipment, information-extreme machine learning, selection level, parallel-sequential optimization.

One of the most demanded areas of development and application of unmanned aerial vehicles (UAVs) in the modern world is the military domain. The rapid progress in technologies such as electronics, robotics, 3D printing, and satellite navigation has already transformed UAVs into highly effective means of engagement that, in terms of their importance in modern armed conflicts, are comparable to traditional types of weapons — aviation, artillery, and armored forces. This transformation significantly alters the approaches to military strategy and tactics. However, as technologies evolve, so do the means of electronic warfare (EW) aimed at countering UAVs. Therefore, there arises a need to develop an intelligent on-board system capable of target acquisition and autonomous homing through the application of computer vision technologies based on machine learning algorithms and pattern recognition methods.

A well-known technology used for learning and pattern recognition is the convolutional neural network (CNN). The key feature of CNNs lies in their use of convolutional filters that extract specific features and structural characteristics of images, which are then transmitted to a multilayer neural network. The main disadvantages of artificial neural networks developed within the framework of the structural bionic approach include their inflexibility during retraining as the number of classes increases, as well as the need for large datasets of military equipment images. Therefore, a promising direction for implementing an intelligent on-board system for recognizing military equipment of an autonomous UAV is the application of the principles and methods of machine learning developed within the framework of the information-extreme intellectual technology (IEI-technology) [1]. The learning algorithm based on IEIT involves a multi-iterative optimization procedure using an information criterion of functional efficiency (CFE) of the spatio-temporal parameters of functioning that directly affect the system's functional efficiency [2]. One of these parameters is the selection level, which defines the coordinates of binary reference or averaged realization vectors of patterns that determine the geometric

centers of class containers in the Hamming binary space. The selection level allows the geometric centers of class recognition containers to be aligned with the empirical centers of pattern scattering and, according to the maximum-distance principle of pattern recognition theory, to maximize the interclass code distance for a given alphabet.

The elements of the averaged binary realization vectors $\{x_{m,i} \mid m = \overline{1, M}, i = \overline{1, N}\}$, m – the number of classes and n – the number of realizations per class, are defined by the formula:

$$x_{m,i} = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_{m,i}^{(j)} > \rho_m; \\ 0, & \text{else,} \end{cases} \quad (1)$$

where $x_{m,i}^{(j)}$ – the binary training matrix [1], ρ_m – the selection level of the coordinates of the averaged vector. In practice, the default selection level $\rho_m = 0,5$, reduces the functional efficiency of the intelligent system's learning process. Therefore, it is reasonable to apply an algorithm of parallel-sequential optimization of selection levels. The parallel optimization stage defines an initial optimal system of selection levels, while the sequential stage determines the optimal selection levels specific to each element of the averaged reference realization vectors. The general optimization formula for the selection levels ρ^* can be represented as follows:

$$\rho^* = \arg \max_{G_\rho} \{ \max_{G_\delta} \{ \max_{G_\Omega \cup G_d} \bar{E}^* \} \}, \quad (2)$$

where G_ρ – the domain of permissible values of the parameter ρ , G_δ – the permissible domain of the control-tolerance field parameter δ , G_Ω – the domain of permissible recognition features of classes, G_d – the permissible domain of class container radii d in the binary space, $\bar{E}^*$ – the optimal averaged value of the CFE for all recognition classes.

Thus, the optimization process of both the selection levels and other training parameters of the intelligent on-board system within the IEIT framework consists in finding the global maximum of the averaged informational functional-efficiency criterion, computed within its definition domain.

References:

1. Dovbysh A.S., Budnyk M.M., Piatachenko V. Yu., Myronenko M. I. Information-Extreme Machine Learning of On-Board Vehicle Recognition System. *Cybernetics and Systems Analysis*, 2020, 56(4), pp 534-543. DOI:10.1007/s10559-020-00269-y.
2. Igor Naumenko. Information-extreme machine training of on-board recognition system with optimization of RGB-component digital images / Igor Naumenko, Mykyta Myronenko, Taras Savchenko // *Radioelectronic and Computer Systems*. –No 4. (98) – 2021. – P. 59-70. – DOI: 10.32620/reks.2021.4.05.

Oleksii Skukis

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, research fellow

A BRIEF ANALYSIS OF TOPOLOGIES IN PARTICLE SWARM OPTIMIZATION ALGORITHMS

Abstract. A brief analysis of topologies in particle swarm optimization algorithms is presented. It is noted that the successful choice of topology is of great importance for the successful application of PSO, since it determines how the balance between exploitation and exploration is achieved.

Particle Swarm Optimization (PSO) is a metaheuristic paradigm of global optimization that has gained popularity due to its simplicity of application to uncontrolled complex multidimensional problems that cannot be solved using traditional deterministic algorithms. The canonical particle swarm optimizer is based on the swarm behavior and social cooperation of flocks of birds, schools of fish, and is largely based on the evolutionary behavior of these organisms. PSO is widely used to solve poorly structured optimization problems of continuous, discrete, constrained, and unconstrained functions [1].

At the same time, PSO is a bio-inspired algorithm for solving optimization problems, where “particles” move in the search space, updating their positions based on their personal best experience and the best experience of their neighbors or the entire swarm. Topology determines the structure of connections between particles, i.e., who exchanges information about the best positions with whom [2]. This critically affects the balance between exploring space and using known solutions, the speed of convergence, and the probability of falling into local minima.

The swarm topology determines how particles interact and exchange information, which is a key factor affecting the performance of the algorithm.

There are various topologies, but the most common are global topology (Gbest) and local topology (Lbest).

In global topology, each individual particle is connected to all others and updates its velocity and position using the best position found in the entire swarm. This ensures fast convergence of algorithms but causes a tendency to get stuck in local minima prematurely.

In local topology, particles are divided into subgroups, in which each particle communicates only with its immediate neighbors. Updates are based on the best position found within its neighborhood. This helps maintain diversity and better explore the search space, but can lead to slower convergence.

In addition to the basic ones, there are other topologies that provide various compromises between convergence speed and research capability [3–5]. Among them is the ring topology. This is a type of local topology, but the particles are arranged in a circle. Each particle has two neighbors (one on the left and one on the right). Star topology is one central particle (or several) that communicates with all the others, and the other particles interact only with it. Hierarchical topology is particles organized into hierarchical groups, where information exchange occurs at different levels.

Clusters – the swarm is divided into groups, with local leaders in each. Tree – a hierarchical structure similar to a binary tree. Random topology – connections are generated randomly with a certain average degree, ensuring the connectivity of the graph. Adaptive/dynamic – the topology changes during iterations.

The choice of topology is crucial for the success of PSO, as it determines the balance between exploitation and exploration, where exploitation is the use of already found optimal information to improve the solution, and exploration is the search for new, potentially better areas in the search space.

The analysis of topologies usually focuses on aspects such as the impact on convergence, the ability to avoid local optima, the complexity of implementation, and the impact on performance for different types of tasks.

Topology affects the speed of information propagation. High connectivity typically accelerates convergence but reduces swarm diversity, increasing the risk of local optima. Low connectivity promotes better exploration but slows down the process. Research shows that the optimal topology depends on the type of task, swarm size, computational budget, and PSO variant.

References:

1. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995.
2. Kennedy, J.; Mendes, R. Population structure and particle swarm performance. In Proceedings of the 2002 Congress on Evolutionary Computation, CEC'02, Honolulu, HI, USA, 12–17 May 2002.
3. Kennedy, J. Small Worlds and Mega-Minds: Effects of Neighbourhood Topology on Particle Swarm Performance. In Proceedings of the IEEE Congress on Evolutionary Computation, Washington, DC, USA, 6–9 July 1999; Volume 3, pp. 1931–1938.
4. Kennedy, J.; Mendes, R. Population Structure and Particle Swarm. In Proceedings of the IEEE Congress on Evolutionary Computation, Honolulu, HI, USA, 12–17 May 2002; pp. 1671–1676.
5. Mendes, R.; Kennedy, J.; Neves, J. Watch thy Neighbour or How the Swarm can Learn from its Environment. In Proceedings of the IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, 26 April 2003; pp. 88–94.

Maksym Steblivskyi

Master's Degree Student, Department of Technology Management

Ihor Miroshnychenko

PhD, Associate Professor of the Department of Technology Management

Taras Shevchenko National University of Kyiv

UNCOVERING SANCTIONS EVASION VIA THIRD COUNTRIES: A NETWORK ANALYSIS OF RUSSIAN IMPORT DATA

Abstract. Economic sanctions have become a primary tool to constrain Russian military aggression against Ukraine. However, Russia continues to acquire restricted goods through intermediary states. Although network analysis is a powerful approach for detecting relationships between entities, it remains underutilized in the study of sanctions evasion. This research applies the metric of betweenness centrality to Russian import data to identify key transit countries involved in the re-export of dual-use goods.

Keywords: sanctions evasion, import, network analysis, betweenness centrality

Introduction. Since 2022, the coalition of sanctioning partners – the European Union, the United States, the United Kingdom, and Japan – has introduced the Common High Priority Items List (CHPL), which specifies 50 dual-use products prohibited for export to Russia [1]. These items include microelectronics, bearings, and machinery used in Russian weapons systems. Despite these restrictions, Russia continues to procure such goods indirectly through “third” countries that have not yet joined the sanctions coalition [2]. Uncovering such evasion routes is essential for strengthening sanctions enforcement and ensuring Ukraine’s national security.

Data and Methods. This study analyzed Russian import customs data from January to March 2025. Shipment records were filtered to include only products manufactured in the EU, the US, the UK, and Japan. Furthermore, only goods corresponding to the 50 Harmonized System (HS) codes listed in the CHPL were retained. Each record contained the following attributes: origin country, where the product was manufactured; departure country, from which it was exported to Russia; and destination country – Russia itself. A directed graph was constructed using the Python library NetworkX, where each node represents a country, and edges represent the shipments of CHPL goods between countries during the study period. Origin countries were placed on the outer ring, while potential transit countries occupied the inner ring of the network (Figure 1). The betweenness centrality for each node was computed using the formula:

$$C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (1)$$

where σ_{st} denotes the number of shortest paths between nodes s and t , and $\sigma_{st}(v)$ represents the number of those paths passing through node v [3]. Additionally, the total shipment value per country in USD was calculated and logarithmically scaled to determine the visual size of each node in the graph.

Results and Discussion. The analysis reveals several key jurisdictions acting as intermediaries in the flow of sanctioned goods to Russia. Countries such as

Turkey, the UAE, China, India, and Hong Kong exhibit the highest betweenness centrality scores, indicating their pivotal roles in facilitating re-exports (Figure 1). These findings confirm that Russia relies on specific logistical hubs to bypass trade restrictions. These intermediaries serve as crucial nodes in the supply chain of restricted technologies.

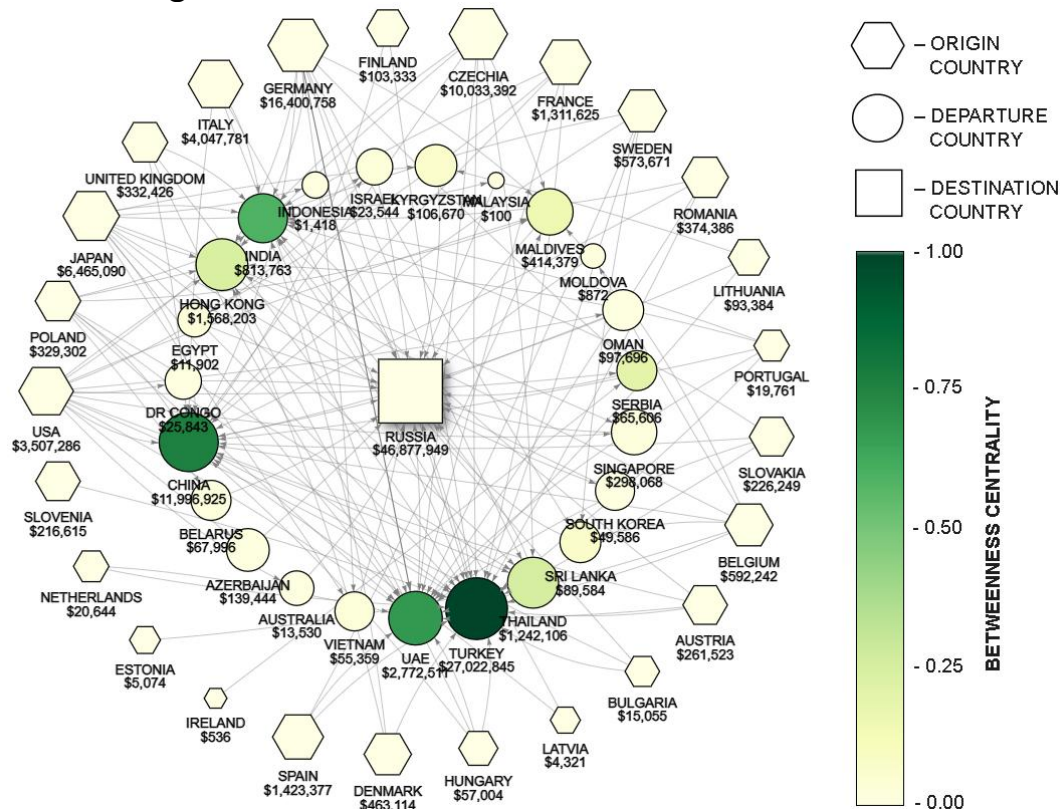


Figure 1 – Network of countries supplying CHPL-listed products to Russia. Node size reflects total shipment value; color reflects normalized betweenness centrality

Conclusions. Future research should extend this network-based framework to the company level, applying betweenness centrality to identify distinct suppliers and intermediary firms involved in sanctions evasion. Comparison of customs data across different time periods could also reveal how Russia's trade networks evolve in response to imposed sanctions, offering insight into the dynamics of sanctions circumvention.

References:

1. U.S. Department of Commerce, Bureau of Industry and Security. Common High Priority Items List, 2024. URL: <https://www.bis.gov/licensing/country-guidance/common-high-priority-items-list-chpl>.
2. Bilousova, O., Hilgenstock, B., Ribakova, E., Shapoval N., Vlasyuk, A., Vlasiuk, V. Challenges of export controls enforcement. How Russia continues to import components for its military production. KSE Institute, 2024, pp. 55.
3. Freeman, L.C., Borgatti, S.P., White, D.R. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 1991, 13, 141–154. [https://doi.org/10.1016/0378-8733\(91\)90017-N](https://doi.org/10.1016/0378-8733(91)90017-N).

Maksym Suprunenko

master's degree student, department of technologies management

Ihor Myroshnychenko

PhD, associate professor of the department of technologies management

Taras Shevchenko National University of Kyiv

PHOTOREALISTIC SYNTHETIC DATA GENERATION FOR ROBUST UKRAINIAN FISCAL RECEIPT OCR

Abstract. The development of high-accuracy Optical Character Recognition (OCR) systems for specialized, low-resource documents, such as Ukrainian fiscal receipts, is severely hindered by data scarcity and degradation challenges inherent to thermal printing.[1] This paper presents a novel methodology for procedurally generating a massive, perfectly annotated synthetic dataset of Ukrainian fiscal receipts using the Blender 3D engine. Our key innovation lies in utilizing Physically-Based Rendering (PBR) to model realistic thermal paper degradation (fading and low contrast) and applying procedural geometry techniques for simulating complex, non-rigid distortions like crumpling and curling. The pipeline leverages the Blender Python API (bpy) to implement Domain Randomization (DR) of extrinsic factors (illumination, camera viewpoint) and intrinsic factors (material decay). This approach automatically produces pixel-perfect ground truth, including character-level bounding boxes and Key Information Extraction (KIE) labels. By strictly adhering to current and prospective Ukrainian fiscal regulations and generating targeted adversarial corruption examples, this framework effectively minimizes the Simulation-to-Real (Sim2Real) domain gap, providing a scalable, cost-effective, and high-fidelity resource to train robust document intelligence models.

Keywords: Synthetic Data, OCR, Key Information Extraction (KIE), Blender, Domain Randomization, Cyrillic.

Introduction. The demand for highly accurate document intelligence in the Ukrainian FinTech and accounting sectors requires OCR performance exceeding 99% accuracy to mitigate financial risks associated with manual data entry.[2] This challenge is compounded by two factors: data scarcity—a lack of large, high-quality public datasets for Ukrainian financial documents [3]—and physical degradation—the rapid fading and low contrast of text printed on leuco-dye-based thermal paper.

Synthetic data generation, through 3D rendering and procedural scripting, offers a scalable solution to bypass the costly and time-consuming manual annotation process while ensuring perfect ground truth annotations. Crucially, the 3D approach allows for meticulous control over physical realism, which is non-trivial to replicate using simple 2D augmentations. [4] Our work focuses on building a photorealistic pipeline that specifically models the unique regulatory and physical challenges of Ukrainian fiscal receipts.

Methodology. The generation framework is implemented entirely within the Blender 3D environment, using its native Python API (bpy) to automate scene creation, data injection, randomization, and ground truth extraction. The scene setup comprises three primary components:

Domain Randomization and Ground Truth: To ensure robustness and Sim2Real transfer capability [1], Domain Randomization (DR) systematically varies scene parameters:

Evaluation Metrics. The resulting synthetic dataset will be used to train specialized OCR and KIE models. Efficacy will be measured using standard metrics across three key tasks:

1. Text Recognition: Character Error Rate (CER) and Word Error Rate (WER).
2. Object Detection: Mean Average Precision (mAP) for locating text lines and key fields.
3. Key Information Extraction (KIE): F1 score (Precision/Recall) for structured data extraction (e.g., Total Amount, Date, TINs).

Conclusion and Future Work. To counter data scarcity and the degradation of thermal paper, our targeted synthetic approach utilizes a Blender-based pipeline. By employing photorealistic PBR and Domain Randomization, we generate an infinite, perfectly annotated dataset. This system minimizes the Sim2Real gap, offering a crucial, high-fidelity resource for robust Document Intelligence solutions.

Future Work will proceed in three sequential phases:

1. Dataset Generation: Systematically generating an initial dataset of 100,000 high-resolution synthetic images with full character-level and KIE annotations.
2. Model Training: Training state-of-the-art transformer-based OCR (e.g., TrOCR) and KIE models (e.g., LayoutLM) using the generated synthetic data.
3. Real-World Validation: Conducting a rigorous Sim2Real validation by evaluating the trained models on a small, manually annotated set of real-world Ukrainian fiscal receipts to confirm performance gains over models trained on general-purpose data.

References:

1. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). URL: <https://arxiv.org/abs/1703.06907>
2. Gorbunova, E., et al. (2023). DocUnet: Document Image Unwarping via A Global Shape-aware Representation. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). URL: <https://arxiv.org/abs/2303.11603>
3. Kim, G., et al. (2022). OCR-free Document Understanding Transformer. Proceedings of the European Conference on Computer Vision (ECCV). URL: <https://arxiv.org/abs/2112.10685>
4. Borisyuk, F., et al. (2022). Rosetta: A Robust Scened Text Recognition System with a Unifying Visual-Language Model. Proceedings of the AAAI Conference on Artificial Intelligence. URL: <https://arxiv.org/abs/2203.02749>

Yehor Symonov

Master, Researcher, Department of Methods and Technological Tools for Building Intellectual Software Systems

Sergii Yershov

Doctor of Physical and Mathematical Sciences, Senior Researcher, Head of the Department of Methods and Technological Tools for Building Intelligent Software Systems

Institute of Cybernetics named after V.M. Glushkov, NAS of Ukraine

THE METHODOLOGY FOR CONSTRUCTING SYNTHETIC ANOMALIES FOR VALIDATING ML PROCESSES

We present a principled methodology for constructing synthetic anomalies to validate machine-learning pipelines. The approach formalizes generation as minimizing distributional divergence under constrained deviation and introduces a locality-preserving realism criterion. A unified framework combines latent- and feature-space perturbations with modality-aware generators and integrates them into cross-validation, stress testing, and robustness assessment. We also outline a metric suite for realism, impact, and diversity, and demonstrate empirical feasibility within standard ML operations.

Keywords: synthetic anomalies, anomaly detection, robustness evaluation, variational autoencoders (VAE), generative adversarial networks (GAN), feature-space perturbation, stress testing, cross-validation, multimodal data, MLOps.

Theoretical Foundations of Synthetic Anomaly Modeling in Machine Learning Systems. In modern machine learning, synthetic anomaly modeling is crucial for validating and improving model robustness. Since real anomalies are rare and data distributions are highly imbalanced, generating synthetic anomalies that reflect real statistical and structural patterns is essential for reliable anomaly detection [1].

Following the generalized classification, four major groups of methods are distinguished: statistical, heuristic, generative-adversarial (GAN-based), and hybrid. Statistical approaches rely on estimating the parameters of the base data distribution

$p_{data}(x)$ and generating synthetic samples x^* that violate density equilibrium in selected regions of the feature space. Heuristic methods employ rule-based or stochastic perturbations of attributes. Generative-adversarial models, structured

around the interaction between a generator $G(x)$ and a discriminator $D(x)$, enable the

creation of realistic deviations from $p_{data}(x)$ through adversarial learning. Hybrid schemes combine statistical and neural mechanisms, achieving a balance between accuracy, interpretability, and controllability of the generation process.

Anomalies are typically classified as point (individual outliers), contextual (deviant in specific contexts), and collective (groups violating expected data patterns).

The construction of synthetic anomalies can be formalized as an optimization problem of minimizing informational divergence between real and synthetic data distributions while maintaining a controlled degree of deviation in the feature space.

Let $\mathcal{X} \subset \mathbb{R}^d$, $p_{data}(x)$ – empirical distribution, $p_{syn}(x; \theta)$ – parameterized synthetic distribution. Then, the generation task is defined as:

$$\min_{\theta} \mathbb{E}_{x \sim p_{data}} \left[D(p_{data}(x), p_{syn}(x; \theta)) \right] \quad s.t. \quad \mathbb{E}[\Delta(x, x^*)] \geq \varepsilon, \quad (1)$$

where $D(\cdot)$ denotes a divergence measure (e.g., Kullback–Leibler, Jensen–Shannon, or Wasserstein), $\Delta(\cdot)$ is the anomaly degree function, and ε represents the threshold of acceptable deviation. The generator must therefore produce samples that are statistically plausible yet informationally significant for validation purposes.

The criterion of synthetic anomaly realism can be formulated as a condition of local structural preservation: $\|\nabla_{\mathcal{N}(x)} p_{syn}(x^*) - \nabla_{\mathcal{N}(x)} p_{data}(x)\|_2 \leq \delta$, where $\mathcal{N}(x)$

is the local neighborhood in the feature space, and δ defines the permissible distortion threshold. This ensures controlled generation of deviations that are relevant for assessing model robustness [2].

Algorithmic Strategies for the Creation and Integration of Synthetic Anomalies into ML Validation Processes. Effective use of synthetic anomalies in ML validation demands a systematic approach combining generative modeling, optimization, and adaptive integration to construct controlled deviations for assessing model robustness and generalization across data types.

A fundamental tool for synthetic anomaly generation is the Variational Autoencoder (VAE), which performs a stochastic mapping of data from the feature space $\mathcal{X}$ into a latent space $\mathcal{Z}$. During reconstruction from the latent representation z , the model approximates the conditional distribution $p_{\theta}(x|z)$, enabling the formation of controlled perturbations: $z^* = z + \eta \cdot \epsilon$, $\epsilon \sim \mathcal{N}(0,1)$, where the parameter η defines the degree of “anomalousness.” This latent perturbation mechanism generates synthetic examples that preserve the topology of the original data while deviating beyond the learned manifold, thus enabling the assessment of model resilience to rare or extreme inputs.

A complementary approach employs Generative Adversarial Networks (GANs), where synthetic anomalies are generated through adversarial interaction

between a generator $G(z)$ and a discriminator $D(x)$. To control the magnitude of deviations, modified architectures such as Conditional GAN (cGAN) or Adversarial Autoencoders (AAE) are applied [3]. These incorporate additional constraints into the loss function:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z; \alpha)))], \quad (2)$$

where the control parameter α determines the intensity of synthetic perturbation. Hence, the GAN-based framework operates as a stochastic generator of controlled anomalies, balancing diversity and realism.

The feature-space perturbation method treats anomaly synthesis as a local transformation problem within the feature space.

For an input vector $x = (x_1, \dots, x_d)$, the perturbed sample is generated as:

$$x_i^* = x_i + \gamma_i \cdot f_i(x), \quad i = 1, \dots, d, \quad (3)$$

where γ_i is the feature sensitivity coefficient, and $f_i(x)$ represents the local perturbation function derived from gradient-based or information-theoretic relevance scores. This method allows the targeted modification of critical attributes, producing synthetic combinations that may not naturally occur in the training distribution but are informative for stress-testing model stability.

Synthetic anomalies are integrated into validation via k-fold augmentation, stress tests, and robustness evaluations: in extended k-fold CV each fold receives a controlled anomaly subset to gauge sensitivity, while stress tests expose models to

anomalies with increasing deviation η_t to estimate robustness thresholds under escalating perturbations. A modular, modality-aware generation setup—RNN/Transformer-attention for time series, GraphVAE/GraphGAN for graphs, conv-VAE for images, and seq2seq or diffusion for text—preserves both universality and domain fidelity. The unified framework combines latent- and feature-level generation, supports multimodal data, and adaptively injects synthetic samples into ML validation; quality and utility are measured by SARS (realism), VII (validation impact), and ADR (diversity).

Experimental Evaluation of the Methodology and Its Application for Enhancing ML Process Reliability. The experiments validate the proposed synthetic anomaly generation method and assess its effect on model stability and generalization. All tests were run in Google Colab Pro+ with an NVIDIA Tesla T4 GPU, 4-core Xeon CPU, and 52 GB RAM, ensuring reproducibility, GPU acceleration, and compatibility with TensorFlow, PyTorch, and Scikit-learn.

Three representative datasets of different modalities were selected to evaluate the proposed approach: KDDCup99 (network intrusion detection), NAB (Numenta Anomaly Benchmark), CIFAR-10 (images)

All datasets were normalized to the $[0, 1]$ range. Structured features were transformed using RobustScaler to mitigate outlier effects. For time-series data, a sliding window of 32 observations with 50% overlap was applied, while image data were rescaled to 64×64 pixels.

Two types of generative architectures were implemented to produce synthetic anomalies:

- Variational Autoencoder (VAE) – consisting of two convolutional layers in the encoder, a latent space of 32 dimensions, and a loss function:

$$\mathcal{L} = \mathbb{E}_{q_{\phi}(z|x)} [\|x - \hat{x}\|_2^2] + \beta KL(q_{\phi}(z|x) || p(z)),$$

with $\beta = 0.5$ controlling the stochastic regularization.

- Conditional Generative Adversarial Network (cGAN) – employing a generator with three convolutional blocks and a discriminator with 0.3 dropout regularization. Training was performed for 50 epochs, with a batch size of 128

and learning rate 2×10^{-4} using the Adam optimizer.

Generated anomalies were incorporated into the validation pipelines of several models: Random Forest, LightGBM, LSTM, and CNN (for CIFAR-10). A 3-fold cross-validation was used throughout, with fixed random seed = 42 for reproducibility.

Table 1 presents the averaged results comparing baseline models (without anomalies) and models enhanced with synthetic anomalies (VAE/GAN-based).

Table 1. The averaged results comparing baseline models

Dataset	Model	F1-score	AUC	RI	Δ VDS
KDDCup99	LightGBM	0.934 to 0.956	0.972 to 0.985	0.88 to 0.94	-21%
NAB	LSTM	0.812 to 0.846	0.867 to 0.891	0.79 to 0.85	-18%
CIFAR-10	CNN	0.901 to 0.927	0.942 to 0.959	0.84 to 0.90	-15%

The results show that integrating synthetic anomalies significantly improves validation stability and generalization. Average increases of 5–12% in F1-score and 3–8% in AUC were observed. The Robustness Index (RI) improved by 7–10%, confirming reduced sensitivity to irregular or noisy inputs.

To ensure statistical validity, a two-sample t-test and Wilcoxon test were applied between baseline and modified models; in all cases, $p < 0.05$, confirming the reliability of the observed improvements. Average generation latency on the GPU-T4 platform was approximately 1.2×10^{-3} seconds per sample for VAE and 3.8×10^{-3} seconds for GAN, indicating computational feasibility for MLOps integration.

Summary and Practical Implications. The experimental results confirm that the proposed methodology enhances ML model validation without incurring excessive computational overhead. The optimal synthetic anomaly ratio was empirically determined to be 5–8% of the training dataset, balancing realism and efficiency. The combined use of VAE and GAN architectures yielded the highest robustness gains (up to 10%).

The results confirm the practicality of synthetic anomalies for enhancing reliability in industrial ML systems-especially in cybersecurity, finance, and fault detection and demonstrate the feasibility of their integration into real-world MLOps and continuous validation pipelines.

References:

1. H. Kim, C. Lee, Enhancing anomaly detection via generating diversified and hard-to-distinguish synthetic anomalies, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), Association for Computing Machinery, New York, NY, USA, 2024, pp. 1089–1098. doi:10.1145/3627673.3679623.
2. F. Croce, M. Hein, Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks, in: Proceedings of the International Conference on Machine Learning (ICML), 2020.
3. M. Krichen, Generative adversarial networks, in: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, Delhi, India, 2023, pp. 1–7. doi:10.1109/ICCCNT56998.2023.10306417.

Danylo Taranov

Master Student, Department of Software Engineering,

Oleh Moroz

PhD, Associate Professor, Department of Software Engineering,

Kyiv Aviation Institute

ARTIFICIAL INTELLIGENCE AS A DIGITAL TWIN FOR OPTIMIZING PROJECT MANAGEMENT

Abstract. The paper presents the concept of an AI agent as a digital twin of an IT project, which integrates data from version control systems (Git), task trackers (Jira), and CI/CD to forecast key metrics and simulate management scenarios. The developed architecture enables proactive risk management and resource optimization.

Keywords: digital twin, artificial intelligence, project management, machine learning, simulation, intelligent agent, optimization.

1. Introduction

In the context of high competition and the dynamism of the IT market, the effectiveness of software development project management is a critical success factor. Traditional approaches based on static plans often prove to be reactive and unable to adapt quickly to changing requirements and technical debt. The concept of the Digital Twin (DT), which has evolved from modeling physical systems to reflecting business processes, opens up opportunities for creating dynamic project management models. This paper examines an approach to creating an AI agent as an intelligent digital twin of an IT project management system.

2. Purpose and novelty

The purpose of this work is to develop the concept of a digital twin for an IT project, capable of forecasting, simulation, and providing recommendations for management decisions based on data analysis from real development systems.

The scientific novelty lies in developing an approach that goes beyond traditional passive monitoring. The uniqueness of the proposed solution consists of:

1. Integrating data from heterogeneous sources (Git, Jira, CI/CD) into a unified analytical model, which allows for obtaining context-dependent forecasts that consider not only task metrics but also code activity.
2. Applying a simulation-based "what-if" approach for the predictive assessment of management scenarios (e.g., changes in team composition, release scope, task priorities) before their actual implementation.
3. Developing the concept of an AI agent that functions not just as an analytical tool, but as an active, predictive element for real-time decision support.

3. Concept

The proposed AI agent architecture ensures integration with common development environments:

Second, it achieves optimization of resource allocation. The ability to simulate "what-if" scenarios provides managers with a tool to make informed decisions regarding changes in priorities, team reorganization, or engaging additional specialists, minimizing the negative impact on overall progress.

Third, it increases the transparency of the development process. A single, integrated model that analyzes data from all systems becomes the "single source of truth," reducing ambiguity in communications and project status assessments.

5. Conclusion

An architecture for an AI agent as an intelligent digital twin of an IT project is proposed, which provides integrated development lifecycle management, scenario simulation, and data-driven decision-making.

The main challenges are the complexity of integration with various corporate systems, the need for high-quality data for training models, and issues of trust in forecasts and the ethical use of AI mechanisms.

The scientific and practical significance of the work lies in creating a foundation for a new generation of intelligent project management systems that combine artificial intelligence, data analytics, and simulation modeling. Prospects for further research include integration with DevOps platforms for analyzing infrastructure metrics and the development of Explainable AI (XAI) mechanisms.

References:

1. Robin K., , Judith M., Andreas W., Jingxi Z.: Digital Twins for Software Engineering Processes, (2025) URL: https://judithmichael.github.io/downloads/preprints/KMWZ25_ICSE_Nier_SoftwareEngineeringProcess_DTs.pdf
2. Oluwakemi F., Godwin O., Okiomah E., Titilayo P. та інші, Improving Workforce Productivity through Data-Driven Metrics: Insights from Agile Teams, (2024) URL: https://www.allmultidisciplinaryjournal.com/uploads/archives/20250318185955_MGE-2025-2-084.1.pdf
3. Miguel A., Juan A.: Digital Twins in Software Engineering—A Systematic Literature Review and Vision, (2024) URL: https://www.researchgate.net/publication/377688712_Digital_Twins_in_Software_Engineering-A_Systematic_Literature_Review_and_Vision

Ivan Vakulov

Master's student, group TSHI-11

Georgy Gaina

Candidate of Technical Sciences, Professor, Professor at the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

INTELLIGENT TOOLS FOR DEVELOPER SUPPORT: PROSPECTS FOR THE DEVELOPMENT OF IDE EXTENSIONS BASED ON LLM

Abstract. The paper examines the use of large language models (LLMs) in IDE extensions for programmer support. A prototype for Visual Studio Code was developed with functions of code generation, explanation, and review. Comparative experiments with GPT-4.1-nano and Gemini 2.0 Flash showed that GPT-4.1-nano ensures accuracy and structured code, while Gemini provides higher speed. Results confirm the potential of LLM integration for improving software development efficiency.

Keywords: artificial intelligence, large language models, code generation, Visual Studio Code, intelligent tools.

In contemporary software engineering, the increasing complexity of software systems necessitates the development of new tools to support programmers. Traditional IDE functionalities such as autocompletion, debugging, and static analysis remain useful but no longer meet the rising demands for speed and code quality. In this context, large language models (LLMs) demonstrate significant potential, as they can perform tasks previously regarded as exclusively human competencies: generating code from textual descriptions, explaining logic, identifying errors, and even suggesting optimizations.

As part of this study, a prototype extension for Visual Studio Code was developed to integrate LLM capabilities into the everyday workflow of programmers. The system supports several scenarios, including automatic function generation, explanation of selected fragments, review with detection of stylistic and logical errors, and answering user questions about code. The module architecture follows the principle of functional decomposition: individual subsystems are responsible for generation, analysis, and interaction with the user, while all operations are carried out through model APIs.

The experimental component involved two providers, OpenAI (GPT-4.1-nano) and Google (Gemini 2.0 Flash). More than thirty test tasks were used to evaluate response quality across parameters such as syntactic correctness, semantic accuracy, style, universality, and creativity. Results confirmed that GPT-4.1-nano produces more structured and well-typed code aligned with clean code principles, while Gemini 2.0 Flash operates almost twice as fast, providing concise yet correct output. This distribution of strengths demonstrates the models' complementarity: GPT is better suited for complex analytical tasks, whereas Gemini is more effective for fast and routine operations.

Practical validation of the module in the VS Code environment revealed that response quality strongly depends on the precision of query formulation, underscoring the importance of prompt engineering as a key factor in effective LLM use. Furthermore, the study results demonstrated the feasibility of employing LLMs for code documentation and review automation, which reduces team workload and enhances process standardization.

Future development of such tools is closely linked to the integration of automated testing, enabling models not only to generate code but also to produce and execute tests in real time. Another promising direction is the deployment of adaptive agents capable of autonomously selecting the optimal model and generation parameters depending on task context. Analytical dashboards also hold considerable potential, allowing the tracking of model performance and user interaction outcomes. In perspective, this will contribute to the emergence of fully-fledged “intelligent development environments” that act not only as instruments but as active participants in the programming process.

IDE extensions powered by large language models open new horizons for software engineering. They not only automate routine tasks but also improve code quality through review and explanation functions. The comparative study confirmed that GPT-4.1-nano is more effective for complex analytical tasks, while Gemini 2.0 Flash provides high speed and concise solutions. Further progress in this area will enable the creation of multifunctional intelligent programming assistants integrated into testing and CI/CD pipelines, supporting adaptive strategies and delivering a new level of efficiency in software development.

References:

1. Vaswani, A., Shazeer, N., Parmar, N. et al. Attention is All You Need. Advances in Neural Information Processing Systems, 2017.
2. Chen, M., Tworek, J., Jun, H. et al. Evaluating Large Language Models Trained on Code. arXiv preprint arXiv:2107.03374, 2021.
3. OpenAI. GPT-4 Technical Report. OpenAI, 2023. URL: <https://openai.com/research/gpt-4>
4. Google DeepMind. Gemini Technical Overview. DeepMind, 2023. URL: <https://deepmind.google/technologies/gemini/>
5. Microsoft. Visual Studio Code Extension API. Microsoft, 2024. URL: <https://code.visualstudio.com/api>

Mykhailo Voitovych

Postgraduate student

Oleksii Bychkov

Doctor of Science in Engineering, Professor

Taras Shevchenko National University of Kyiv

HYBRID AUTOMATA-GUIDED MULTI-AGENT REINFORCEMENT LEARNING ARCHITECTURE FOR MULTI-PHASE UAV SWARM MISSIONS

Abstract. Unmanned aerial vehicle (UAV) swarms often face missions that involve multiple distinct phases such as broad-area search, target tracking, and threat evasion. This article describes a hybrid decision-making scheme that couples a supervisory hybrid automaton with multi-agent deep deterministic policy gradient (MADDPG) to coordinate UAV swarms across multiple mission modes (Search, Track, Safe). The automaton handles discrete mode switching on event triggers (e.g., target found/lost), while MADDPG learns continuous, mode-conditioned control.

Keywords: multi-agent reinforcement learning; UAV swarms; hybrid automata; MADDPG

Real UAV swarm missions decompose into distinct behaviors—wide-area search, persistent tracking, and contingency safe actions [5]. Reinforcement learning (RL) provides a powerful framework for UAV swarms to learn coordinated behaviors through trial-and-error interaction with the environment [1].

In multi-agent settings of RL, each UAV acts as an intelligent agent that observes its surroundings and collaboratively learns a policy to maximize a shared mission reward. Prior work shows MARL effectiveness for multi-UAV search, tracking, and formations in dynamic environments [1, 4]. One algorithm in particular, Multi-Agent Deep Deterministic Policy Gradient (MADDPG), has proven successful for continuous control with centralized training, decentralized execution: a centralized critic uses global information during learning; decentralized actors run onboard with local observations in flight [1, 3, 4].

A monolithic policy must balance conflicting objectives and vast state–action spaces, which often leads to unstable training and brittle execution. Structuring behavior into discrete modes reduces complexity and aligns control with mission logic, improving sample efficiency and reliability [1, 3, 5, 6]. A hybrid automaton is a formal model that combines discrete states (modes) with continuous dynamics (control laws within each mode) which could extend MADDPG in ability not to train one policy to handle complex scenarios [2].

The proposed hybrid automata + MADDPG approach (Figure 1) integrates discrete mission logic with continuous multi-agent reinforcement learning to enable robust UAV swarm coordination. The hybrid automaton supervises the swarm by switching between three operational modes: Search, Track, and Safe. Transitions are event-driven, triggered by conditions such as target detection, loss, or jamming. Once a mode is active, control is delegated to MADDPG controllers. In this framework, decentralized actors generate actions for each UAV based on local observations, while a centralized critic evaluates joint state–action pairs during training, supported

by a global replay buffer for stable updates. This architecture follows the CTDE paradigm, ensuring policies can be trained centrally but executed independently onboard each drone.

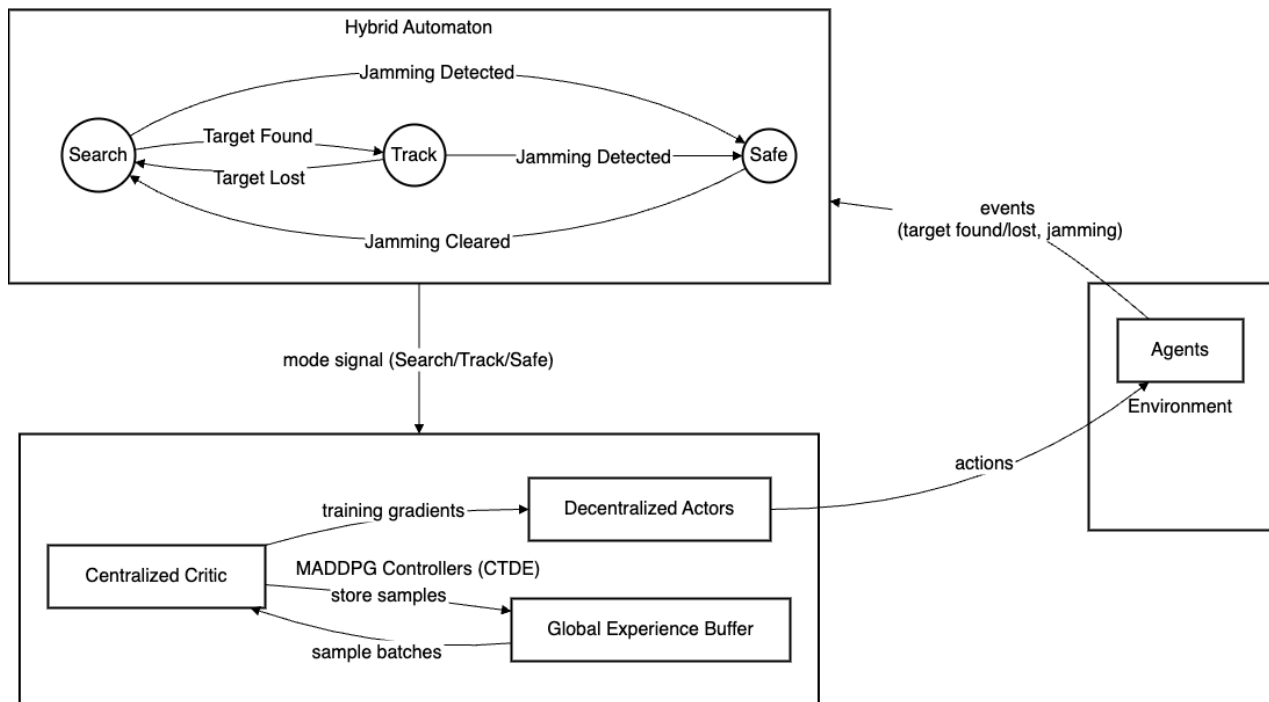


Figure 1 - Hybrid Automata + MADDPG high level architecture

References:

1. Y. Cai, F. Jiang, J. Wang, X. Shen, Graph-based multi-agent reinforcement learning for collaborative UAV search and target tracking in dynamic environments, *Neurocomputing* 500 (2022) 751–764.
2. J. E. Sierra-García, M. Santos, Intelligent control of a UAV with a cable-suspended load using a neural network estimator, *Expert Systems with Applications* 185 (2021) 115380.
3. K. Zhang, Z. Yang, T. Başar, Multi-agent reinforcement learning: A selective overview of theories and algorithms, in: *Handbook of Reinforcement Learning and Control*, Springer, Cham, 2021.
4. J. Xu, G. Liang, Enhanced Q learning and deep reinforcement learning for unmanned combat intelligence planning in adversarial environments, *Scientific Reports* 15(1) (2025) 28364. doi:10.1038/s41598-025-13752-3.
5. Z. Feng, X. Na, S. Hai, Q. Sun, J. Shi, Deep reinforcement learning for UAV target search and continuous tracking in complex environments with Gaussian process regression and prior policy embedding, *Electronics* 14(7) (2025) 1330. doi:10.3390/electronics14071330.
6. M. Liu, J. Wei, K. Liu, A two-stage target search and tracking method for UAV based on deep reinforcement learning, *Drones* 8(10) (2024) 544. doi:10.3390/drones8100544.

Eduard Zakabluk

Oleksandr Lapin

Sumy State University, 116, Kharkivska Street, Sumy, 40007, Ukraine

INFORMATION-EXTREME MACHINE LEARNING OF AN ONBOARD GEOINFORMATION SYSTEM FOR SEMANTIC SEGMENTATION OF AGRICULTURAL LAND

Abstract. The study proposes a method of information-extreme synthesis for performing semantic segmentation of agricultural land images using an onboard geographic information system based on data obtained from unmanned aerial vehicles, aiming to ensure high recognition accuracy.

Keywords: machine learning, information-extreme technology, geographic information system, image segmentation

Introduction. Accurate assessment of crop conditions plays a vital role in improving agricultural productivity. The combined application of unmanned aerial vehicles (UAVs) and artificial intelligence (AI) enables rapid and cost-efficient agromonitoring, facilitating automated crop identification and the delineation of cultivation areas to support evidence-based decision-making by farmers and governmental agencies. UAVs offer superior mobility, affordability, and data acquisition speed, while ensuring minimal disturbance to vegetation, thus representing an optimal solution for crop surveillance. This study introduces the design of a UAV system based on information-extreme technology, aimed at processing digital imagery of agricultural territories and identifying potential crop cultivation zones.

Related works. Although significant advancements have been made in neural network architectures, they continue to face challenges associated with high-dimensional feature spaces and complex multi-class classification problems. Methods grounded in the geometric interpretation of decision rules have demonstrated superior performance [1]. In this context, the information-extreme intelligent technology (IEI-technology) is considered a promising alternative to neural-like structures, as it is characterized by high flexibility, unification, and interpretability, which allows controlling the machine learning process to achieve high accuracy.

Methods and materials. The objective of the information-extreme synthesis of a learning-enabled geoinformation system is to maximize its functional efficiency, defined by the system's information capacity, which in turn directly influences the reliability of its classification outcomes. To train the machine learning model, four image fragments were extracted from a UAV-acquired digital image of agricultural land (Figure 1).

Each fragment measured 50×50 pixels and corresponded to a separate crop field. On the basis of these fragments, an alphabet of recognition classes was established : class X_1^0 - field No. 1, class X_2^0 - field No. 2, class X_3^0 - field No. 1:

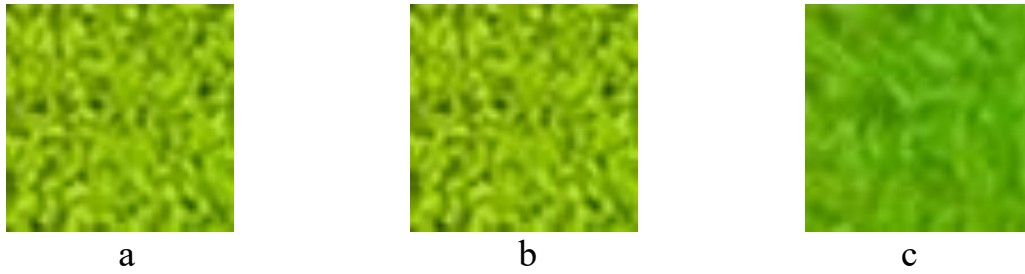


Figure 1: Images of regions of interest frames: a – class X_1^0 ; b – class X_2^0 ; c – class X_3^0 .

Results

As a result of the information-extreme machine learning of the second depth level, the optimal radii of the recognition class containers were obtained according to the modified Kullback criterion [2] (Fig. 2). In the figure, the working area for determining the information criterion function used to optimize the machine learning parameters is shown in dark color.

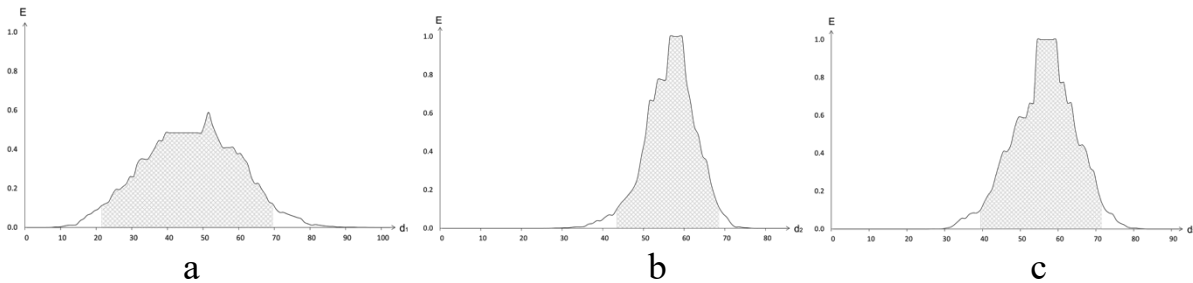


Figure 2: Graphs of the dependence of the normalized information criterion on the radii of recognition class containers: a – recognition class X_1^0 ; b – recognition class X_2^0 ; recognition class X_3^0 .

The optimal radii of recognition class containers equal: $d_1^* = 51$ code units for class X_1^0 at the value $E_1 = 0.59$; $d_2^* = 53$ for class X_2^0 at the value $E_2 = 1.0$; $d_3^* = 56$ for class X_3^0 at the value $E_3 = 1.0$. Using these optimal geometric parameters, a set of decision rules was constructed.

The experimental results demonstrated that the proposed approach validates the applicability of information-extreme methods within onboard geoinformation systems for crop monitoring and establishes a practical basis for semantic segmentation under real UAV operating conditions.

References

1. K. R Muller., S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernelbased learning algorithms, IEEE Transactions on Neural networks 12 (2001): 181-202.
2. Naumenko I., Myronenko M., Savchenko T., Piatachenko V. Information-Extreme Machine Learning of an On-board Ground Object Recognition System with a Choice of a Base Recognition Class. CEUR Workshop Proceedings 2022 CEUR-WS.org

Olha Zalutska

Assistant of Computer Science Department

Oleksandr Mazurets

Ph.D. in Engineering Science, Associate Professor of Computer Science Department

Maryna Molchanova

Ph.D. Senior Lecturer of Computer Science Department

Khmelnyskyi National University, Ukraine

EFFICIENCY ANALYSIS OF WRECKING WASTE CLASSIFICATION USING NEURAL NETWORK

Abstract. This study examines the efficiency of a neural approach to identifying and classifying wrecking waste in photographic data. Building on a two-factor design that couples scene-level detection with per-class binary refinement, we quantify where accuracy is earned, where it is lost, and which components: architecture, data composition, augmentation, and arbitration – most strongly govern the outcome. Using a composite dataset of ten material classes, we report that the hybrid method attains up to 97.8% classification accuracy, with substantial gains over a detector-only baseline on heterogenous or texture-confusable categories. Detector metrics such as $mAP50 \approx 0.746$ and $mAP50-95 \approx 0.669$ confirm reliable localization, while binary residual classifiers close the gap in label assignment, lifting macro F_1 from ≈ 0.751 to ≈ 0.974 . These findings indicate that efficiency, measured as correct material routing per unit inference cost, is maximized when detection and classification are decoupled yet reconciled through a calibrated arbitration rule.

C&D waste is visually complex due to overlapping fragments, contamination, and shared texture–color statistics [1]. End-to-end detectors with multi-class heads often show cross-class interference and unstable false positives, whereas a split pipeline – with detection followed by one-vs-rest specialist classifiers – improves stability and adaptability [2].

This study compares the efficiency and robustness of a two-stage approach against a detector-only system. The pipeline starts with YOLOv11 trained on a curated dataset of ten C&D classes (brick, concrete, foam, general waste, gypsum board, pipes, plastic, stone, tile, wood). Detector outputs (640×640 inputs) provide bounding boxes and coarse class hypotheses. Cropped fragments are then processed by binary ResNet50/50V2 classifiers trained per class with restrained augmentation – rotations, affine jitter, and mild photometric variation simulating realistic conditions. During inference, a calibrated consensus validates the binary decision if its posterior exceeds a threshold and aligns with the detector; otherwise, the detector’s label prevails. Efficiency is evaluated at detection (localization, coverage) and classification (Accuracy, Precision, Recall, F_1 , macro-averaged) levels.

The image corpus integrates an object-detection dataset collected at an industrial sorting facility together with stacked-waste and RGB-D supplements to improve coverage of concrete, brick, wood, and related classes. This composite was split into detector training/validation sets and, separately, into per-class binary subdatasets.

YOLOv11 achieves robust localization with overall $mAP50 \approx 0.746$ and $mAP50-95 \approx 0.669$, as well as mean IoU near 0.897, indicating that box proposals are

generally well placed even under heterogeneous backgrounds. Classwise performance varies: categories with distinctive macro-texture and geometry (brick, concrete, gypsum board, tile) maintain high Precision–Recall pairs, whereas visually heterogeneous or small-scale classes (foam, pipes, plastic) degrade, a pattern typical of detectors operating under partial occlusion and surface contamination. These detector characteristics are desirable for throughput few missed regions and reliable boxes yet they leave residual label noise that impacts downstream routing costs.

The binary ResNet stage yields decisive improvements in attribution. On a stratified evaluation, the detector’s macro metrics centered near Accuracy ≈ 0.601 and F1 ≈ 0.751 . After refinement, macro Accuracy rises to ≈ 0.975 and macro F1 to ≈ 0.974 , with the largest deltas on historically weak classes. For foam, F1 improves from ≈ 0.544 to ≈ 0.988 ; for pipes, from ≈ 0.450 to ≈ 0.959 ; for plastic, from ≈ 0.455 to ≈ 0.937 . Even for strong classes such as brick and concrete, residual classifiers consolidate Recall without materially harming Precision. Aggregate best-case accuracy over the ten classes reaches $\approx 97.8\%$, evidencing that the detector’s spatial prior plus specialist refinement is more sample-efficient than increasing detector capacity alone.

Training-time sweeps over epochs reveal a typical U-shaped risk curve: gains accrue through 10-15 epochs for most classes, while further epochs occasionally erode minority-class generalization, especially for categories with higher intra-class variability such as plastic or pipes. Comparisons with MobileNetV2 show that compact backbones can approach ResNet performance on simpler classes, but deeper residual models better capture micro-texture cues that separate confusable materials. From an efficiency standpoint, this suggests a mixed strategy: retain MobileNet-scale specialists for easy, high-frequency categories and reserve ResNet-scale specialists for the hard tail to maximize accuracy per compute budget.

References:

1. M. Molchanova, V. Didur, O. Mazurets, O. Sobko, O. Zakharkevich, Method for Construction and Demolition Waste Classification Using Two-Factor Neural Network Image Analysis, in: CEUR Workshop Proceedings, 3970, 2025, pp. 168–182.
2. O. Hladun, O. Zalutskaya, V. Klimenko, O. Mazurets, Research on the Effectiveness of Classifying the Remains of Destroyed Buildings Using MobileNetV3 Neural Network Architecture, in: Innovations in Science: From Theoretical Foundations to Practical Impact, Proceedings of the 1st International Scientific and Practical Conference, Antwerp, Belgium, 2025, pp. 158–162.

Mariia Zheludkova

Master's degree student of the faculty of information technology

Oleh Zaritskyi

Doctor of Engineering Science, professor (associate) of the faculty of information technology, Taras Shevchenko National University of Kyiv

Taras Shevchenko National University of Kyiv

SOCIAL AND ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE

Abstract. The paper examines the social and ethical implications of artificial intelligence, focusing on its influence on privacy, employment, democracy, and human dignity. It analyses current international regulatory initiatives, including the OECD and UNESCO recommendations, as well as the EU Artificial Intelligence Act and the Council of Europe Framework Convention. The study highlights the need for comprehensive legal frameworks to ensure transparency, accountability, and ethical governance of AI technologies.

Keywords: AI, ChatGPT, privacy, ethics, regulation, discrimination.

Artificial intelligence (AI) is defined as the most significant technological breakthrough of the 21st century. Its rapid advancement is striking: ChatGPT, launched in November 2022, had already reached more than 700 million active weekly users by July 2025 – approximately 10% of the global population – collectively sending 2.5 billion messages per day, or 29,000 messages per second [1]. Statistical research shows that users most frequently engage with this technology for non-work-related queries – nearly 70% of all interactions – and this proportion continues to grow alongside the expanding user base [1]. In 78% of requests, the most common topics include practical guidance, information retrieval, and content creation (writing) [1].

Given these figures, societies are becoming increasingly concerned about the social and ethical implications of AI technologies, ranging from data bias and privacy violations to their influence on employment, democracy, and human dignity [2]. Consequently, the issue of legal regulation of these processes has become highly relevant in the modern context [2].

The first step in this direction was taken with the publication of policy recommendations by the Organisation for Economic Co-operation and Development (OECD) in 2019 and the United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2021. Both institutions promoted AI as an innovative technology whose core value-based principles should include the following [3, 4]:

- AI should serve humanity and contribute to sustainable development while safeguarding human rights and fundamental freedoms;
- biased decisions must be avoided by ensuring diversity and inclusiveness;
- users must understand how the system operates;
- AI systems should be technically robust, verified, transparent, secure, and environmentally responsible;
- developers and operators must be accountable for the impact of their systems.

However, the main criticism of these documents was that they remained declarative and lacked binding legal force.

In 2024, the Council of Europe Framework Convention on Artificial Intelligence and the European Union Artificial Intelligence Act were adopted to address this shortcoming. These instruments legally formalized the aforementioned recommendations. Among other provisions, the EU AI Act classified systems according to their level of risk [5, 6]:

- those posing an unacceptable risk must be prohibited and blocked within the European Union (pending possible exceptions) if they exhibit the following characteristics: cognitive-behavioural manipulation of individuals or vulnerable groups, social scoring, biometric identification and categorization of persons, or similar real-time detection systems;
- those posing a high risk, as they may adversely affect safety and fundamental human rights, are divided into two categories: systems covered by EU product safety legislation, and AI systems operating in specific high-impact domains (such as critical infrastructure, workforce management, or border control), which must be registered in an EU database;
- those with a limited risk, for which the primary requirement concerns transparency – users must be informed that they are interacting with AI;
- those with a minimal risk, which are not regulated – for example, video games or spam filters.

Nevertheless, these instruments still lack comprehensive and systematic measures addressing copyright protection, disinformation, labour market impacts, and transparency regarding corporate data training practices.

Thus, while technological progress cannot – and should not – be halted, it must be carefully regulated and guided in ways that align with the constructive needs of society. Unfortunately, at present, no complete and effective regulatory framework exists to govern AI systems comprehensively.

References:

1. A. Chatterji, T. Cunningham, D. Deming, Z. Hitzig, C. Ong, C. Shan, K. Wadman, How people use ChatGPT, NBER Working Paper No.34255, September 2025, JEL No. J01, O3, O4
2. O. Zaritskyi, Index of Artificial Intelligence Systems Ethics, Information Technology and Implementation (IT&I-2022), Kyiv, Ukraine, November 30 - December 02, 2022, pp. 169-174.
3. OECD, Recommendation of the Council on Artificial Intelligence, 2019. URL: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
4. UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2022. URL: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000381137&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_75c9fb6b-92a6-4982-b772-79f540c9fc39%3F_%3D381137eng.pdf&updateUrl=updateUrl6019&ark=/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi&fullScreen=true&locale=en#1517_21_EN_SHS_int.indd%3A.8918%3A4
5. Council of Europe, Framework convention on artificial intelligence and human rights, democracy and rule of law, 2024. URL: <https://rm.coe.int/ai-convention-brochure/1680afaeba>
6. Official Journal of European Union, Artificial Intelligence Act, 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

Vadym Ziuziun

PhD, Associate Professor of the Department of Technologies Management

Andrii Utvenko

PhD student of the Department of Technologies Management

Taras Shevchenko National University of Kyiv

APPLICATION OF REINFORCEMENT LEARNING TO OPTIMIZE INPUT PARAMETERS IN ORDER TO ACHIEVE PREDEFINED PERFORMANCE METRICS DURING BENCHMARKING OF A HARDWARE–SOFTWARE SYSTEM

Abstract. The growing complexity and versatility of contemporary software–hardware systems demand advanced methods for parameter selection and configuration optimization, particularly in the context of benchmarking tasks. Traditional manual approaches are often inefficient and may fail to deliver optimal results in dynamic and multifactorial environments. Reinforcement learning emerges as a promising paradigm for building autonomous agents capable of intelligently searching for and selecting input parameters, progressively improving system performance through iterative interaction with the environment.

Keywords: reinforcement learning, hardware–software system, optimization, benchmarking, adaptability, autonomy, hyperparameter optimization.

Modern hardware–software systems feature high levels of technical complexity, incorporating numerous functional subsystems and software modules that must operate efficiently in a diverse range of conditions. One of the central challenges in their testing and deployment is the effective selection of input parameters to achieve desired indicators such as performance, energy efficiency, or stability. This challenge is further amplified by the variability of real-world scenarios, the dynamic nature of operational environments, and the multifactorial influence of both internal and external factors.

The application of reinforcement learning (RL) offers new opportunities for the automation and intelligent adaptation of this process. RL agents learn iteratively by interacting with the system: by adjusting configuration parameters, they observe changes in target metrics, receive rewards or penalties according to outcomes, and gradually formulate optimal decision-making policies. Unlike traditional approaches, RL is able to incorporate long-term effects of actions and supports the development of self-learning capabilities in dynamic or partially unknown settings.

Specialized benchmarking libraries and modeling platforms, such as algorithm configuration environments and automated optimization tools, are highly valuable in these tasks. Advanced search methods for multi-dimensional parameter spaces, multi-agent frameworks, and automatic adjustment of agent strategies based on accumulated experience play a decisive role in improving system outcomes. Progress has been made in the development of frameworks for hyperparameter optimization, which can be integrated with RL to further accelerate the search for the most effective configurations.

Modern hardware–software systems present a significant engineering challenge due to their multifaceted structure, the integration of diverse software modules, and

their operation in highly variable conditions. A critical issue in their design and evaluation is the effective selection and tuning of input parameters to meet predefined objectives, which often include performance, energy efficiency, and operational stability. Traditional approaches relying on manual configuration or static heuristics frequently fall short in adapting to the dynamic and high-dimensional environments inherent to these systems. Reinforcement learning (RL) offers a transformative framework by enabling agents to learn optimal configuration policies through trial-and-error interaction with the target system, thereby autonomously balancing multiple conflicting objectives. RL agents iteratively adjust parameters based on observed feedback in the form of performance metrics, effectively constructing decision policies that improve over time and are capable of responding to changing conditions and operational uncertainties. The use of dedicated benchmarking libraries and frameworks such as DACBench [1], coupled with modern hyperparameter optimization platforms like Optuna [2] and Auto-WEKA [3], facilitates the systematic evaluation of RL-driven configuration strategies and supports robust and scalable solution development. Moreover, the integration of multi-agent configurations and dynamic parameter control mechanisms enhances the flexibility and scalability of these approaches for complex, large-scale, and distributed environments. This paradigm reduces human intervention, accelerates the search for optimal configurations, and increases the reliability and adaptability of software-hardware systems operating in heterogeneous and evolving contexts.

A reinforcement learning-centered strategy dramatically increases the scalability of testing procedures, ensures robustness to variable external factors, and decreases dependence on subjective human decision making. Such solutions demonstrate substantial potential for adaptation across a wide spectrum of technical problems – from engineering complex information systems to distributed computing platforms. Nevertheless, the successful deployment of RL-based approaches requires in-depth understanding of practical aspects of environment configuration, as well as ongoing evaluation of the performance, stability, and extensibility of known algorithms in various operational contexts.

References:

1. N. Awad, T. Eimer, F. Hutter, M. Lindauer. Multi-agent Dynamic Algorithm Configuration. arXiv preprint. 2022. arXiv:2210.06835.
2. T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'19, Anchorage, AK, USA, 2019. 2623–2631. DOI: 10.1145/3292500.3330701.
3. L. Kotthoff, C. Thornton, H. Hoos, F. Hutter, K. Leyton-Brown. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. Journal of Machine Learning Research. 2017. 18: 16–261. <https://www.jmlr.org/papers/volume18/16-261/16-261.pdf>.

Oksana Zolotukhina

PhD in Technical Sciences, Associate Professor, Associate Professor of the Department of Intelligent Technologies

Oleg Ilarionov

PhD in Technical Sciences, Associate Professor, Associate Professor of the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

FEATURES OF BUILDING A RECOMMENDATION SYSTEM FOR ELECTIVE COURSES

Abstract. The article presents the features of the design of the architecture of a recommender system to support the selection of elective subjects in higher education institutions. A four-layer model is proposed, which includes data integration, constraint checking, recommendation generation, and user interaction, providing personalized and informed course selection.

Keywords: recommendation system, elective courses, system architecture, learning personalization, decision support.

Recommendation systems are a modern tool for intelligent decision support designed to effectively navigate users in large amounts of information. Their main task is to reduce information overload and provide personalized access to content, goods or services [1]. Thanks to the use of such systems, the user receives individually relevant offers, which allows optimizing the search process, reducing time and increasing the convenience of interaction with information or commercial platforms. An important result of the functioning of recommendation algorithms is not only facilitating the choice, but also expanding the user's capabilities by discovering new areas of content or products that may meet his interests, even if they were not previously explicitly formulated. In the field of education, recommendation systems are increasingly used as a tool to support the educational process, academic counseling and improving student success [2]. Recommender systems in education, on the one hand, aim to improve the student experience, reduce indecision when choosing courses or materials, and increase learning productivity; on the other hand, they are becoming an important tool for educational institutions to improve academic advising, reduce student dropout rates, and improve the overall quality of educational programs.

One of the key tasks is to help students choose courses or educational resources that best match their academic goals, level of knowledge, and current situation (e.g., semester, topics of disciplines already completed). Research [3] demonstrates that models that take into account courses taken in the same semester (session group) when building a recommendation work better than traditional approaches based on popularity or associations. To promote courses related to the core, the authors also propose post-processing to adjust recommendations. Research [4] uses the GRU4Rec and NARM models to predict the next actions of the user, considering only the current session of interactions with educational materials. Such approaches are especially useful in online courses, where the student can work with the system

anonymously or with a minimal history of interactions. In the article [5], the application of collaborative filtering to recommend elective courses to students is considered. The authors propose a methodology based on the use of information about prerequisites and average GPA, applying the singular value decomposition (SVD) algorithm to determine the similarity between students. Experimental results on real data show the high efficiency of the model, which allows to form personalized and accurate recommendations for choosing courses. It should be noted that educational systems often require explanations for recommendations: it is useful for the student to know why the proposed course or material is considered relevant - this increases trust in the system and helps in the conscious formation of educational trajectories. In the article [6] it is noted that one of the weak features of many systems is the lack of transparency and explanations, which can reduce user motivation.

Elective courses are courses chosen by students from a catalog approved by the university. Their purpose is to form an individual educational trajectory: a student can deepen knowledge in a professional field, expand interdisciplinary training, or develop additional competencies [7]. In Ukraine, elective courses have a special specificity associated with national education standards, the autonomy of higher education institutions, and the characteristics of academic culture. The choice is made within the framework of the educational program and meets the requirements of the European Credit Transfer System (ECTS), which operates in Ukrainian higher education institutions. The university usually forms a limited list of available electives depending on staffing and resource availability.

The development of recommendations for the selection of elective courses for students of higher education institutions in Ukraine should take into account the existing system of restrictions regulated by the procedure for choosing disciplines. In particular, this:

- academic restrictions – only those disciplines that are allowed in the student's curriculum for their year of study and educational program can be recommended;

- study load restrictions – the total number of credits that a student studies during a semester or academic year is fixed, therefore the recommendation of disciplines must take into account situations of exceeding or falling short of the credit norm;

- restrictions of the structural and logical scheme – studying disciplines usually requires possession of a certain set of knowledge and skills (this can be confirmed by studying prerequisite disciplines or, for example, by the results of non-formal or informal education), the recommendation must take into account the student's existing knowledge and skills required to study the recommended discipline;

- restrictions on the number of students and the teaching load of teachers – study groups formed to study elective courses must be economically profitable for the educational institution (for this purpose, a lower limit on the size of the group is established) and can be formed only taking into account the limitations on the teaching load of the teachers of the discipline.

Taking into account the above restrictions, the general architecture of the

recommender system for elective courses can be represented by the diagram in Fig. 1 and includes:

- external systems that serve as a data source;
- Data Integration Layer;
- Restriction check Layer;
- Recommendation building Layer;
- Output recommendation & Interaction Layer.

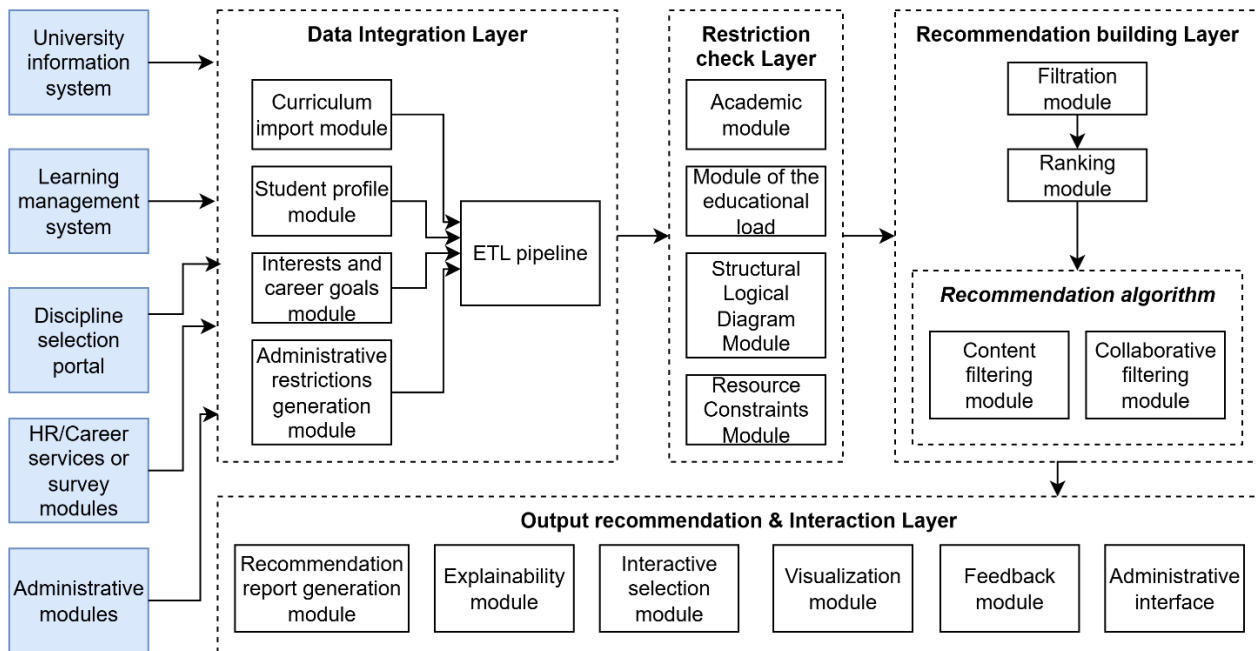


Figure 1 – General architecture of a recommender system for elective courses

The Data Integration Layer collects, unifies, and prepares data from various sources: the university's information system, LMS, administrative modules, and survey platforms. At this stage, data about students, curricula, courses, prerequisites, faculty workloads, and group restrictions are normalized, cleaned, and transformed into a standardized format suitable for further analytics. It should be noted that in the practical implementation of the system, it is necessary to take into account the features of external systems, since this directly affects the structure and formats of data that need to be processed on the Data Integration Layer. The Restriction Check Layer includes modules to check the compliance of proposed courses with academic, structural-logical, and resource constraints, ensuring that recommendations do not conflict with the curriculum, prerequisites, and allowable course load. Recommendation Building Layer implements the basic logic of recommendation generation: combines content and collaborative filtering, analyzes the student's profile, interests, achievements, and similarity with other users, forming a personalized list of elective courses. The Output Recommendation & Interaction Layer is responsible for generating explained recommendations, interactive interaction with the user and feedback: the student receives a clear explanation of the reasons for the choice, can compare options, view the workload and schedule, and the system takes into account his actions to further improve the recommendations.

The proposed architecture of the recommendation system for the selection of elective courses provides coordinated work between different data sources and constraint checking mechanisms, which makes the recommendation process more accurate and manageable. Due to the modular design, the system can be integrated into the existing IT infrastructure of the university and gradually expanded. This approach allows you to automate part of the administrative processes and support students in making informed decisions when choosing disciplines, while adhering to educational and organizational requirements.

References:

1. Ko H, Lee S, Park Y, Choi A. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*. 2022; 11(1):141. <https://doi.org/10.3390/electronics11010141>
2. Zhang, Q., Lu, J. & Zhang, G. (2021). Recommender Systems in E-learning. *Journal of Smart Environments and Green Computing*, 1(2), 76-89.
3. Md Akib Zabeed Khan, Agoritsa Polyzou. Session-based Methods for Course Recommendation. (2024). *Journal of Educational Data Mining*, 16(1), 164-196. <https://doi.org/10.5281/zenodo.11384740>.
4. Thai-Nghe, Nguyen & Sang, Pham. (2022). A Session-Based Recommender System for Learning Resources. *Communications in Computer and Information Science*. 10.1007/978-981-19-8069-5_51.
5. Rahman Md. Mijanur, Oaishi Chelsea, Mahmud Jubaer & Sadat Md. (2025). Elective Course Recommendation System Using SVD Algorithm. 10.1007/978-981-97-8096-9_34.
6. Souabi, S., Retbi, A., Khalidi Idrissi, M., & Bennani, S. (2021). Recommendation Systems on E-Learning and Social Learning: A Systematic Review. *Electronic Journal of e-Learning*. <https://doi.org/10.34190/ejel.19.5.2482>
7. Khoruzha, L. Electiveness as a guiding principle of adult education. *Pedagogical Education: Theory and Practice. Psychology. Pedagogy*. 35 (1) (2021): 6-11.

I. Yarmolai, aspirant

ORCID 0009-0001-7013-5790

Military Institute of Taras Shevchenko National University of Kyiv

CONSTRUCTION OF A MATHEMATICAL MODEL OF SEISMOACOUSTIC MONITORING FOR REMOTE EXPLORATION

Abstract. The construction of a mathematical model of seismoacoustic monitoring of explosive fields for remote exploration is considered. As is known, seismoacoustic monitoring of blast fields is used for remote sensing and is a set of routine observations, whereby the mode of the observations themselves and the spectral parameters of the object under study depend on the research task at hand. As shown in the article, the reliability of explosive signal classification is improved through the use of remote sensing information technologies based on seismoacoustic monitoring. To achieve the research goal, it is necessary to construct a mathematical model of a continuous explosive field signal that would reflect the most important aspects of the explosive field signal monitoring process. In the process of constructing such a model, it is necessary to take into account both the parameters describing the process itself and the parameters of interference and natural background noise, as well as the characteristics of the transmission function of the medium.

Keywords: seismoacoustic monitoring, parametric mathematical model, seismoacoustic analysis, explosive fields, seismoacoustic signal, seismoacoustic signal model.

Introduction. Seismoacoustic monitoring of blast fields for remote sensing is a set of regime observations, and the regime of the observations themselves and the spectral parameters of the object under study depend on the research task. Increasing the reliability of blast signal classification is achieved through the use of information technologies in remote sensing based on seismoacoustic monitoring. The term "regime observations" used above provides a clear procedure.

First of all, it is necessary to build a mathematical model of the inseparable signal of the blast field, which would reflect the most significant moments of the process of monitoring blast field signals. In the process of building such a model, it is necessary to take into account both the parameters that describe the process itself and the parameters of the interference and natural noise background, as well as the specifics of the environment's transfer function. To monitor blast fields, it is necessary to collect statistics for various signals of blast fields and transfer functions of the environments in which the signal propagates. This will provide a priori ideas about both the explosive field at the study points and the explosive field signals themselves. A priori knowledge of the random process of interference will enable a significant reduction in its influence on estimating the explosive field signal under study. This weakening is achieved by optimizing the processing procedures that take into account the a priori statistics of the random interference process. The work takes into account the influence of the instability of the research process parameters and optimizes the procedure for processing observed data according to criteria that consider the characteristics of the natural background interference. The process of monitoring explosive fields is reduced to estimating the informative parameters of parametric mathematical models of separate and inseparable signals of the explosive field, the superposition of which forms the explosive field itself. The set of all

informative parameters of each explosive field signal forms a vector of these parameters in n -dimensional Euclidean space. Each of these vectors uniquely characterizes the signal, the parameters of the mathematical model of which are this vector. Thus, the classification of resolved and inseparable signal explosion fields is reduced to the classification of these vectors in n -dimensional Euclidean space.

The purpose of this article is to construct a mathematical model of an indistinguishable explosive field signal that would reflect the most important aspects of the explosive field signal monitoring process. In the process of constructing such a model, it is necessary to take into account both the parameters describing the process itself and the parameters of interference and natural background noise, as well as the characteristics of the transmission function of the medium [1].

Main part. Mathematical model of indivisible signals. The model of the separate signal of the explosive field is a superposition of solutions to a second-order differential equation, which describes the superposition of oscillators that enter at different times, each with its own amplitude and natural frequency. The model of the explosive field we will present is the model of separate signals in the following way:

$$M(t, \lambda) = \sum_{i=0}^I \Phi(t - \lambda_{0+4i}) \lambda_{1+4i} \left[e^{-\lambda_{2+4i}(t - \lambda_{0+4i})} \sin[\lambda_{3+4i}(t - \lambda_{0+4i})] \right]; \lambda = \lambda_{p+4i} \quad i=\overline{0, I}; p=\overline{0, 4}, \quad (1)$$

here λ – vector of free model parameters; I – number of pilmoles that are in superposition; p – corresponding submodel number; $\Phi(t)$ – Heaviside function; i – submodel serial number; λ_{0+4i} – signal arrival time; λ_{1+4i} – signal amplitude; λ_{2+4i} – signal attenuation coefficient; λ_{3+4i} – signal frequency;

The optimal estimation of signal parameters involves determining the vector of free parameters that minimizes the value of the criterion of agreement between the model and the observation data. Such a model gives good agreement in the case of modeling a linear system of oscillating objects and thus takes into account the oscillatory nature of explosive signals.

As a criterion of agreement, we choose the value of the norm L_2 (the integral of the square of the deviation of the model from the observed data $y(t)$) or L_1 (the integral of the modulus of the deviation of the model from the observed data $y(t)$). In the first case, the criterion $F(\lambda)$ takes the form $F(\lambda) = \int_T [y(t) - M(t, \lambda)]^2 dt$.

$$\text{In another case: } F(\lambda) = \int_T |y(t) - M(t, \lambda)| dt$$

The optimal estimate of free parameters λ^* is the point in the parameter space that minimizes $F(\lambda)$:

$$F(\lambda^*) = \min_{\lambda \in \Lambda} F(\lambda), \quad (2)$$

$y(t)$ – analytical approximation vector of a function of observed data values;

Λ – the set of all possible values of a vector λ .

1. These parameters enter the model nonlinearly. As a result, criterion (2) is a surface with many local extrema. This type of criterion has determined the approach to finding the global minimum. The points in the vicinity of which local extrema are

located are selected by the random search method (Monte Carlo) according to a priori distributions [2]. Therefore, the number of calculation cycles must be sufficiently large to ensure the correctness of the result, which becomes increasingly complex with a large model dimensionality.

To find the local minimum of the criterion, we need to calculate the time derivatives $\frac{\partial F(\lambda)}{\partial \lambda_k}$, $k = \overline{0, K}$, and, equating them to zero, create a system of equations that has the form:

$$\frac{\partial F(\lambda)}{\partial \lambda_k} = \int_T [y(t) - M(t, \lambda)] \frac{\partial M(t, \lambda)}{\partial \lambda_k} dt = 0, \quad k = \overline{0, K}$$

Система приводится:

$$\int_T [y(t) D(M(t, \lambda))] dt = \int_T [M(t, \lambda) D(M(t, \lambda))] dt,$$

here $D(M(t, \lambda))$ – a vector composed of functions, each of which is a derivative of the model with respect to all corresponding components of the vector λ .

For model (1), this vector has the form:

$$D(M(t, \lambda)) = \left\{ \frac{\partial M(t, \lambda)}{\partial \lambda_{p+i}} \right\}; \quad p = \overline{0, 3}; \quad i = \overline{0, I}, \quad (3)$$

The relevance of the variational approach in solving geophysical problems can be traced, for example, in [3].

The search for the minimum of criterion (2) is carried out using the Levenberg-Marquardt algorithm [4,5], for a priori randomly selected points in the space of free parameters of model (1).

Model (1) can be generalized to the case of a flow of inseparable signals, i.e., when the signal carriers intersect. Then the model of a flow of inseparable signals takes the form:

$$M(t, \lambda) = \sum_{s=1}^S \sum_{i=0}^I \Phi(t - \lambda_{0+i,s}) \lambda_{1+i,s} \left[e^{-\lambda_{2+i,s}(t - \lambda_{0+i,s})} \sin[\lambda_{3+i,s}(t - \lambda_{0+i,s})] \right], \quad (4)$$

here S – the number of signals that overlap in the group; λ – matrix of free model parameters; I – number of pilmoles that are in superposition; p – corresponding submodel number; $\Phi(t)$ – Heaviside function; i – submodel serial number. $\lambda_{0+4i,s}$ – start time of the S -th signal; $\lambda_{1+4i,s}$ – amplitude of the S -th signal; $\lambda_{2+4i,s}$ – attenuation coefficient of the S -th signal; $\lambda_{3+4i,s}$ – frequency of the S -th signal.

That is, the model (4) consists of S separate discrete signals. In the general case, S is a random variable. In the matrix of free parameters of the model $\Lambda = \lambda_{k,s}$; $k = \overline{0, K}$; $s = \overline{1, S}$; $K = 4I$; , the column with the number s is the vector of parameters of the s -th signal.

In this parameter vector mod $k, 4$ – is the number of the $\lambda_0, \lambda_1, \lambda_2, \lambda_3$ parameters. This harmonic has the number ant $k, 4$. We adopt the following notation:

$\text{mod } k, 4$ - is the remainder of the number divided k by 4, and $\text{ant } k, 4$ - is the integer part of the division k by 4.

The above example of model (1) can be considered a special case of the more general model (4), in the case when $\lambda_{0+i,s}$ does not depend on i .

The optimal estimate of the matrix of free parameters of the model (4) Λ is determined by the values of the free parameters that give the minimum value of criterion (2) on the set of local minima of this criterion. The procedure for obtaining the optimal estimate consists in determining the points of local extrema closest to each of them for the set of M matrices Λ , and then selecting the global one on the set of local minima [6].

The point that gives this minimum criterion in the $(K+1) \cdot S$ dimensional space is chosen as optimal for the free parameters of the model signal. In such a procedure, convergence to the optimal solution is ensured by probability with increasing number M . In this case, we can assume that we have obtained as a result of the solution a superposition of separate signals M (1) that began at different times.

Conclusions. A new mathematical model of the explosive seismic field is presented, which accounts for various types of signals in the explosive field. A new mathematical model of non-separable signals of the explosive field is presented. A mathematical apparatus for solving this model is provided. It is worth noting that these algorithms were developed for automated seismic acoustic monitoring systems.

To assess the adequacy of the model for non-separable signals in the explosive field, simulation modeling of non-separable signals was carried out within the framework of an improved methodology for seismoacoustic monitoring, based on seismoacoustic analysis.

References:

1. Abramowitz M. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables / M. Abramowitz, I. A. Stegun, // 9th printing. New York: Dover, 1972.
2. Mostovoy V.S., Toliupa S.V., Shevchenko A.M. Optimization of the dynamic parameters of an object in a mathematical model of seismo-acoustic monitoring of natural and engineering objects. Science-intensive technologies. – Kyiv: NAU, 2023. – No. 2 (58). – P. 412-422. <https://doi.org/10.18372/2310-5461.58.17647>.
3. Mostovoy V.S. Optimal detection of signals against a background of microseismic noise. Proceedings of the National Academy of Sciences of Ukraine. – 2008. – No. 1. – P. 106-110.
4. Marquardt D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters / D. Marquardt // SIAM Journal on Applied Mathematics, 1963, p. 431–441.
5. Plessix R.-E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications - Geophys. J. Int. (2006) 167, 495
5. Mostovyy V., Mostovyy S. Estimation of parameters of seismic waves. Transactions of NAS of Ukraine. 2014, Vol. 2, pp. 118-123.
6. Stein S., Wyssession M. E. An Introduction to Seismology, Earthquakes, and Earth Structure. London: Wiley-Blackwell, 2002. P. 400.

Iryna Yurchuk

PhD in Physical and Mathematical sciences,

Assoc. Prof. of Software Systems and Technologies Department

Golouh Stanislav

4th year bachelor, student of the Faculty of Information Technologies,

Software Systems and Technologies Department

Shulga Vladyslav

4th year bachelor, student of the Faculty of Information Technologies,

Software Systems and Technologies Department

Pishta Ihor

4th year bachelor, student of the Faculty of Information Technologies,

Software Systems and Technologies Department

Taras Shevchenko National University of Kyiv

ENHANCING THE RELIABILITY OF UAV DETECTION THROUGH MULTIMODAL SENSOR DATA FUSION

Abstract. The rapid proliferation of unmanned aerial vehicles (UAVs) poses significant challenges for security and defense systems. Reliable detection of UAVs is complicated by their small size, low radar visibility, and ability to fly at low altitudes. Systems relying on a single sensor often prove ineffective under diverse conditions. This paper proposes and analyzes an architecture for multimodal sensor data fusion that integrates information from visual, thermal, and acoustic sensors. The goal is to develop an optimized model that significantly improves detection accuracy, reduces false alarms, and enhances overall system reliability in real-world operational environments.

Introduction. The increasing availability and capabilities of UAVs have intensified the demand for automated systems for their detection and tracking. Traditional identification methods that rely on a single sensor type (e.g., radar or optical cameras) can be unreliable and prone to errors due to weather conditions, background noise, or complex terrain. Advances in multisensor data fusion technologies [1] offer a powerful solution to this problem.

By combining data from different modalities—such as visible spectrum, thermal (infrared) radiation, and acoustic signals [3]—the system can compensate for the weaknesses of individual sensors. For example, an optical sensor may be ineffective at night or in fog, but a thermal sensor can detect the heat signature of a drone. Similarly, acoustic sensors can identify the unique sound patterns of UAV propellers [5], even when the drone is visually obscured. This approach has already proven effective in other complex domains, such as intelligent transportation systems [2].

Objective of the Study. The objective of this research is to develop and optimize a multimodal data fusion model for accurate UAV detection and classification. The study aims to improve recognition accuracy, reduce classification errors, and enhance the model's reliability for potential applications in defense, surveillance, and security systems.

System Architecture and Methods. The success of a detection system largely depends on its data collection and processing architecture. This study proposes a comprehensive architecture that utilizes multiple observation technologies [4].

The initial model employed separate architectures for each sensor input. For visible spectrum data, a Convolutional Neural Network (CNN) was used for object detection. For acoustic data, a 1D CNN was implemented for audio signal classification. While this approach provided baseline classification capabilities, it lacked sufficient reliability. For instance, the camera-based model could not detect UAVs at night, and the acoustic model produced many false alarms in noisy urban environments.

To improve performance, the final model was enhanced with a **feature-level data fusion** architecture. High-level features extracted from individual CNNs (for visual and thermal data) and from the 1D CNN (for acoustic data) were concatenated into a single feature vector. This vector was then fed into a series of dense (fully connected) layers for final classification (UAV / non-UAV). This approach allows the model to learn complex correlations between different modalities [3, 5].

Results and Discussion. To evaluate effectiveness, a comparative analysis was conducted between the accuracy of baseline (single-sensor) models and the final optimized (multimodal) model. The dataset simulated various conditions, including daylight with visual obstructions and nighttime scenarios.

Single-sensor approaches demonstrated inconsistent performance. The visual sensor (camera) achieved acceptable accuracy under ideal conditions, but its performance dropped significantly under poor lighting or adverse weather. The acoustic sensor remained relatively stable but had lower overall accuracy due to background noise [5].

In contrast, the final optimized **fused model** achieved the highest accuracy, outperforming each individual modality. The system maintained high detection accuracy across all scenarios. This confirms that data fusion enables the system to compensate for the weaknesses of individual sensors by leveraging the strengths of others, which is crucial for reliable detection [1, 2].

Conclusions. The developed multimodal neural network effectively addresses the challenge of reliable UAV detection from multisensor data streams, showing substantial improvements in accuracy compared to single-sensor systems. The combination of carefully selected sensors (visual, thermal, and acoustic) [3, 5] and a well-designed data fusion architecture proved essential for achieving high performance.

In the context of defense and surveillance, the ability to accurately and reliably classify UAVs under varying conditions plays a crucial role in enhancing situational awareness and decision-making. The proposed method contributes to automated recognition systems, improving operational efficiency and precision in security missions [4]. Moreover, this approach not only addresses current challenges in UAV

classification but also lays the groundwork for future advancements in applying AI to security systems.

Future research may focus on further optimizing the model for real-time performance [3], exploring more advanced fusion methods, and expanding the dataset to include more diverse and complex scenarios.

References:

1. Hall D. L., McMullen S. A. H. *Handbook of Multisensor Data Fusion*. – CRC Press, 2001.
2. Wei C., Qin Z., Zhang Z., Wu G., Barth M. J. *Integrating Multi-Modal Sensors: A Review of Fusion Techniques for Intelligent Vehicles*. – arXiv, 2025.
3. Svanström F., Englund C., Alonso-Fernandez F. *Real-Time Drone Detection and Tracking With Visible, Thermal and Acoustic Sensors*. – Semantic Scholar, 2020.
4. Shi X., Yang C., Xie W., Liang C., Shi Z., Chen J. *Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges*. – IEEE, 2019.
5. Frid A., Ben-Shimol Y., Manor E., Greenberg S. *Drones Detection Using a Fusion of RF and Acoustic Features and Deep Neural Networks*. – MDPI, 2024.

SECTION 2: CYBERSPACE PROTECTION TECHNOLOGIES

Stepan Bilan

Ph.D., associate professor

Taras Shevchenko National University of Kyiv

INTERNAL SECURITY OF ROBOTIC SWARM BEHAVIOR DURING MESSAGE EXCHANGE

Abstract. In modern robotics, increasing attention is being paid to the efficient construction of intelligent swarms. In this case, high-quality information exchange within a group of robots is of great importance. The correct behavior of the swarm is determined by the complete interaction of all robots in the group based on the embedded algorithms and software. Encryption is used to securely exchange information between robots. The survivability of the robot swarm depends on timely data exchange in real time. For this purpose, the best option is to use stream ciphers. However, if data exchange occurs between a large number of robots in a group without central control, then the best option is to use asynchronous stream ciphers. The paper examines the organization of data exchange between robots based on generators of pseudo-random bit sequences, which are implemented on the basis of cellular automata with active and heterogeneous cells, which makes it possible to organize reliable encryption within the swarm.

Keywords: robot swarm, asynchronous stream cipher, cellular automata, active cell, heterogeneous cell, pseudorandom bit sequence generator.

Introduction

In modern distributed intelligent systems, information exchange plays an important role, as the quality of this information determines the reliability of the entire system and its individual components. Such systems include robot swarms and IoT systems. Swarm robotics is characterized by the fact that robots operate by interacting with each other and do not have a central control device. Groups of robots that form a swarm are aimed at solving more complex problems than single robots. At the same time, there is a need for full interaction in a group between robots and their information support. Swarm robotics is based on collective behavior algorithms that exploit the behavior of colonies of animals, insects, and other biological organisms. At the same time, swarm robotics has identified a number of challenges in creating an artificial intelligent swarm, the main one being reliable information exchange, which allows for reliable behavior and high information security. The paper addresses the problem of increasing the reliability of information exchange between robots in a group by means of reliable encryption based on an asynchronous stream cipher, which is formed using cellular automata (CA) with active and heterogeneous cells.

To implement reliable information exchange between the robots of the group, CA with active cells (CAAC) and heterogeneous cells (CAHC) are used [1, 2]. Based on such CAs, pseudo-random bit sequence generators (PRBSG) are implemented, which are capable of forming several key sequences for the formation of various stream ciphers. At the same time, they have a repetition period of pseudo-random bit sequences of long duration. In KAAK, each active cell produces a pseudo-random bit sequence at its output, which is the key for generating a stream cipher for each robot in the group. Active cells transmit active signals to neighboring cells and transfer

them into an active state at the next evolutionary step. The larger the size of the CAAC, the more active cells can be present in it. Each active cell implements a local state function and a local transition function. The number of active cells depends on the number of robots in the group. In this arrangement, each robot in the group generates an asynchronous stream cipher to exchange encrypted messages with the corresponding robot in the group. Each robot is assigned its own unique stream cipher in real time.

CAHC, in addition to cells that perform one local transition function, contains cells that perform another local transition function. There are usually fewer heterogeneous cells than homogeneous ones. The output that forms an asynchronous stream cipher can come from any CAHC cell. These can be either homogeneous or heterogeneous cells.

Each robot communicates with other robots through a single, shared PRBSG with multiple outputs. This increases the resistance to attacks on the encryption keys generated by the robots. For each robot, asynchronous codes are generated in parallel in real time. An attack on a single robot does not disrupt the operation of the entire swarm, and the robots themselves analyze the behavior of neighboring robots in the swarm and disable information exchange with the malfunctioning robot whose functioning is disrupted. Using a single PRBSG restricts other robots from accessing the cipher of a single robot. In fact, many asynchronous ciphers are formed in the swarm simultaneously, which prevents an attacker from determining the operating principles and tasks facing the swarm as a whole.

The described approach is characterized by the fact that all robots in the group know the active cells of the other robots, since the same generator is used for all robots. To improve the reliability of a group of robots, it is necessary to divide one generator into many, each of which performs encryption only for one slave robot.

During the operation of a robot group in real time, robots can be added with the addition of a new active cell, and robots and the corresponding active cells can be removed from the group. Having more active cells than robots in a group increases the hacking resistance of a generator built on a cellular automata with active cells. In this case, the CAHC may contain a small number of non-uniform cells, which does not affect the duration of the repetition of the formed bit sequence. The implementation of pseudo-random bit sequence generators based on CAAC and CAHC allows the formation of a high-quality pseudo-random bit sequence due to the high flexibility in changing states. The proposed generator model, unlike generators on shift registers with linear feedback and other generators, does not allow determining statistical relationships between the ciphers generated by the CA for each robot.

References

1. S.M.Bilan, M.M.Bilan, R.L. Motornyuk. New Methods and Paradigms for Modeling Dynamic Processes Based on Cellular Automata, IGI-Global. (2020), P. 200.
2. Stepan Bilan. Formation Methods, Models, and Hardware Implementation of Pseudorandom Number Generators: Emerging Research and Opportunities.- (2017).- IGI Global, USA.- P. 301.

Ivan Bilokon

Candidate of Technical Science (PhD)

Yaroslava Fylenko

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Sofiia Denichenko

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

AUTOMATED OSINT DATA COLLECTION AND ANALYSIS FOR CYBER RISK ASSESSMENT

Abstract. This paper examines automated methods of collecting publicly available data using OSINT (Open Source Intelligence) technologies, with a focus on cybersecurity applications. The study highlights the use of machine learning to enhance the efficiency and scalability of data gathering processes. Additionally, it explores how OSINT data can be leveraged to detect and analyze information leaks, assess associated risks, and support proactive threat mitigation strategies.

Keywords: OSINT, cyber risk assessment, data leaks, machine learning, automated data collection, cybersecurity, threat intelligence.

The modern digital landscape provides vast opportunities for collecting and analyzing publicly available information. OSINT (Open Source Intelligence) refers to the systematic and lawful gathering, verification, and processing of data from open sources. It is widely applied in national security, journalism, business analytics, and cyber intelligence.

Typical OSINT sources include websites, social media, forums, government registries, open databases, news platforms, satellite imagery, and multimedia content. A key advantage of OSINT is its ability to integrate diverse data streams into a unified analytical view.

OSINT collection methods fall into technical, analytical, and automated categories. Technical approaches involve domain and IP analysis, DNS records, WHOIS data, open ports, and metadata. Analytical techniques include content analysis, social media intelligence (SOCMINT), geospatial intelligence (GEOINT), and media intelligence (IMINT). Automated collection, such as web scraping, uses tools like Maltego, Shodan, SpiderFoot, and Recon-ng to efficiently process large volumes of data.

One of the most critical OSINT applications is detecting data leaks and assessing cyber risks. Leaks of sensitive information pose serious threats, and OSINT enables timely identification of such incidents. Public data may reveal traces of leaks across forums, dark web sources, or social media, helping estimate their scope and impact.

To improve analysis accuracy, machine learning algorithms are used to classify leak types, detect anomalies, assess risks, and generate response strategies. Techniques like clustering help identify information sources, while metadata analysis

reveals the time and location of leaks. Integrating OSINT with ML models supports early warning systems and proactive threat mitigation

The global OSINT technology market is expected to grow significantly – from \$8.3 billion in 2024 to \$64.9 billion by 2033. The most dynamic growth is projected in areas such as Big Data software, video analytics, and text analytics, driven by advances in artificial intelligence and the increasing role of OSINT in cybersecurity. Other domains, including social media analytics, security analytics, web analysis, and geospatial analytics, also show steady expansion.

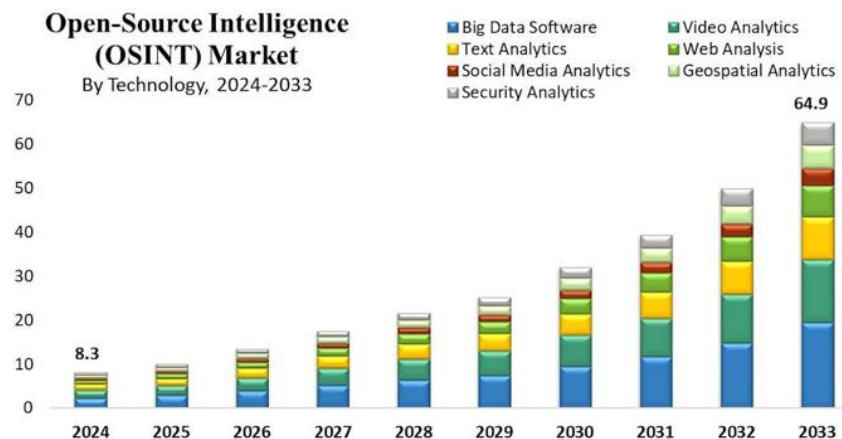


Figure 1 – Trends in the Global OSINT Technology Sector

In conclusion, OSINT combined with machine learning provides a powerful framework for data collection, analysis, and risk evaluation, enabling organizations to effectively counter modern cyber threats, particularly those related to data breaches and information leaks.

References:

1. Bazzell M. Open source intelligence techniques: Resources for searching and analyzing online information. 2018. 461 p.
2. Research D. M. Open-Source Intelligence (OSINT) Market is expected to reach a revenue of USD 64.9 Bn by 2033, at 25.6% CAGR: Dimension Market Research. *GlobeNewswire News Room*. URL: <https://www.globenewswire.com/news-release/2024/04/15/2863048/0/en/Open-Source-Intelligence-OSINT-Market-is-expected-to-reach-a-revenue-of-USD-64-9-Bn-by-2033-at-25-6-CAGR-Dimension-Market-Research.html>.
3. Rajamäki J., McMenamin S., Tiitta K. The Importance of Open-Source Intelligence in Assessing Risks of Cyberattacks. *Information & Security: An International Journal*. 2024. Vol. 55, no. 1. P. 15–31. URL: <https://doi.org/10.11610/isij.5511>.
4. Automated OSINT Techniques for Digital Asset Discovery and Cyber Risk Assessment / T. Babenko et al. *Computers*. 2025. Vol. 14, no. 10. P. 430. URL: <https://doi.org/10.3390/computers14100430>.

Nykyta Biriukov

Bachelor in cybersecurity, student

Yurii Shcheblanin

Candidate of technical sciences, senior researcher, associate professor. Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv

FRACTAL-BASED CRYPTOGRAPHY WITH NORMALIZATION

Abstract. With the rapid advancement of information technologies, the need for secure data transmission has become paramount. Encryption, rooted in practices from ancient civilizations, continues to safeguard confidentiality, integrity, and authenticity of information. As conventional ciphers become increasingly vulnerable over time, fractal-based encryption methods provide stronger security by leveraging self-similarity, unpredictability, and intricate pattern generation.

Keywords: Cryptography, information security, fractal.

The growing complexity of information systems and the constant emergence of new threats have emphasized the necessity of encryption methods that go beyond classical deterministic approaches. Traditional ciphers, while effective in the past, are increasingly vulnerable to modern cryptanalytic techniques. This motivates the exploration of mathematical constructs such as fractals, which by their very nature embody unpredictability, self-similarity, and chaotic recurrence across scales [1].

Singular fractal functions, particularly of the Cantor type, have already demonstrated the ability to transform initial symbolic sets into highly complex numerical sequences that resist straightforward analysis. However, most existing models are limited by the rigidity of their parameterization, where fixed input conditions produce predictable structural boundaries, even if the output itself appears chaotic [2]. To overcome this limitation, we introduce the concept of normalization within the iterative construction of fractal functions.

The role of normalization here is not merely corrective but generative: it allows the encryption framework to maintain high entropy while reducing unnecessary redundancy in key expansion [3]. In effect, normalization improves both the efficiency and the resilience of fractal-based encryption. Even minor variations in the normalized parameters result in significantly altered ciphertexts, creating a system where the recovery of the original sequence without precise knowledge of the normalization factors becomes computationally infeasible [4].

By embedding normalization into fractal-based encryption, we obtain a model that is more adaptable and resistant to analysis compared to earlier static constructions. This advancement highlights the potential of fractals not only as abstract mathematical objects but also as practical instruments for strengthening modern cryptographic systems. The findings suggest that fractal normalization may serve as a cornerstone for future encryption architectures, combining elegance of mathematical theory with the pressing requirements of digital security [5].

Conclusion

As information systems continue to manage both public and restricted-access data, encryption remains a cornerstone of digital security. Fractals, mathematical constructs that model self-similarity and complexity across multiple scales, offer a novel foundation for cryptographic design. By integrating fractal structures into encryption schemes, it becomes possible to achieve multi-layered protection, where unpredictability and intricate pattern generation replace conventional reliance on pseudorandom number generators. Such approaches enhance the robustness of ciphertext against cryptanalysis, providing stronger resilience in the evolving landscape of data protection.

References:

1. Secure data transmission through fractal-based cryptosystem [Электронный ресурс] – Режим доступа до ресурсу: <https://www.nature.com/articles/s41598-025-04700-2>
2. Fractal-Based Hybrid Cryptosystem: Enhancing Image Encryption [Электронный ресурс] – Режим доступа до ресурсу: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10670236/>
3. Using fractal as an Encryption Method [Электронный ресурс] – Режим доступа до ресурсу: https://www.researchgate.net/publication/308418446_Using_fractal_as_an_Encryption_Method
4. A Quantum-Secure Cryptographic Algorithm Integrating Fractals and Prime Numbers [Электронный ресурс] – Режим доступа до ресурсу: <https://www.mdpi.com/2076-3417/14/22/10138>
5. A blockchain-integrated chaotic fractal encryption scheme for industrial IoT applications [Электронный ресурс] – Режим доступа до ресурсу: <https://www.nature.com/articles/s41598-025-89604-x>

Serhii Buchyk

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Vladyslav Lytvynenko

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

KUBERNETES NETWORK POLICIES AS THE ZERO TRUST BASELINE: A CASE FOR MANDATORY DEFAULT DENY SEGMENTATION

Abstract. The default networking model of Kubernetes (K8s) permits unrestricted internal communication, creating a severe vulnerability where a single compromised pod enables full horizontal (lateral) movement. This abstract advocates for the mandatory deployment of a default "Deny-All" Network Policy as the primary architectural control to enforce Zero Trust principles within the cluster. This mechanism instantly minimizes the attack surface by blocking all East-West traffic and structurally enforces the principle of Least Privilege.

Keywords: Default Deny, Zero Trust Architecture, Kubernetes, Least Privilege, Microsegmentation, Lateral Movement.

The shift to cloud-native technologies, including microservices orchestrated by Kubernetes (K8s), has eliminated traditional network trust boundaries [1]. While K8s provides flexibility and scale, its inherent networking design, which provides unique IP addresses to every pod, does not impose any security filtering by default. This permissive configuration creates a high-risk security posture where, should an attacker breach any single workload, they gain the capability for unhindered exploration and exploitation across the cluster's internal network. This systemic risk necessitates an architectural reversal of trust - moving from an implicit "allow-all" to an explicit "deny-all" posture to align with the Zero Trust (ZT) mandate of continuous verification [2]. The implementation of a mandatory default deny security stance is therefore the fundamental step in safeguarding modern distributed applications against internal threats.

The principle of blocking all traffic by default is achieved through the use of a Kubernetes Network Policy (KNP) set to apply to all pods (podSelector: {}), defining explicit deny rules for both Ingress (incoming) and Egress (outgoing) traffic [3]. This structural approach transforms the cluster's network behavior, forcing every subsequent communication flow to be justified and explicitly allowed.

The immediate benefit of the default "deny-all" policy is the structural enforcement of the Principle of Least Privilege. Before this policy, a developer had to take specific action to restrict communication; after implementation, a developer must take specific action to permit it. This inversion yields several critical advantages:

Secure Baseline. It provides an enhanced security posture, ensuring that pods without any specific security policy (or pods with incorrectly defined policies) are immediately isolated and denied all traffic until the appropriate policy is correctly defined and applied;

Attack Surface Minimization. By blocking all unnecessary ports and connections, the policy dramatically reduces the reachable network surface available to an attacker, lowering the risk profile of every deployed microservice;

Mandatory Microsegmentation. The policy serves as the critical first step in microsegmentation, ensuring that isolation is the default state. Subsequent KNP, which explicitly permit traffic (e.g., allowing a frontend pod to communicate only with a specific backend pod on port 8080), establish the necessary isolation domains based on workload labels and application logic;

Data Exfiltration Prevention. Implementing a Default Deny for Egress (outbound) traffic is crucial. It ensures that pods cannot initiate connections to unauthorized external IP addresses or services. This hardens the cluster against command-and-control communication or data exfiltration attempts.

Implementing Default Deny is not merely a technical configuration; it imposes an operational mandate that drastically improves security posture. By immediately breaking all undocumented service dependencies, this policy forces development and operations teams to achieve full clarity on their application's necessary network graph.

However, L3/L4 policy enforcement is insufficient for complete ZTA, as it cannot verify the cryptographic identity of the workload making the request – a key ZT requirement. Therefore, the Default Deny policy must be viewed as the foundational layer upon which more sophisticated security layers are built. For instance, Service Mesh technologies provide the necessary Layer 7 (L7) controls, such as mutual TLS (mTLS) and identity-based authorization, to verify who is communicating, not just where the communication originates [4]. An alternative approach utilizes advanced CNIs, such as Cilium, which employ eBPF technology. They extend KNP capabilities by adding FQDN filtering, DNS policies, and certain L7 control functions. This avoids the complexity and overhead associated with deploying a full-service mesh with sidecar proxies.

The management of these policies is crucial. The high frequency of network misconfigurations in K8s environments presents a significant risk. Therefore, all KNP configurations must be managed as Policy-as-Code (PaC) and subjected to rigorous testing to ensure both security compliance and operational continuity.

References:

1. Chandramouli, R., & Butcher, Z. (2023). A zero-trust architecture model for access control in cloud-native applications in multi-location environments. <https://doi.org/10.6028/nist.sp.800-207a>
2. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture. <https://doi.org/10.6028/nist.sp.800-207>
3. Network policies. <https://kubernetes.io/docs/concepts/services-networking/network-policies/>
4. William Morgan. Zero trust network security in Kubernetes with the service mesh. <https://www.buoyant.io/zero-trust-in-kubernetes-with-linkerd>

Serhii Buchyk

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Vitalii Piatyhor

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

IMPROVED SOCIAL MEDIA BOT DETECTION SYSTEM BASED ON PUBLICLY AVAILABLE DATA

Annotation: Social media platforms are one of the main sources of information nowadays. However, as trust in them grows, so does the use of bot accounts to spread disinformation and influence public opinion, which is an effective method of manipulation. Automated account detection systems are used to detect such accounts.

Keywords: Computer system, data collection, data analysis, machine learning, cloud technologies

At present, social media networks have become one of the most popular communication models; they are influencing social processes, economic life, and even political life, enabling the instantaneous sharing of information and creating communities around ideas or problems. These services also allow groups of people to be united by interests or geo-location, which provides an opportunity to quickly push people to certain action, both positive, such as fundraising or volunteering, or negative, such as mobbing.

To enable the use of social media platforms to spread disinformation or promote hostile ideas, bot networks can be used by malicious actors. Social media bots are computer algorithms that create content and interact with users [1]. These bots aim to appear and behave like real users, making them difficult to detect and allowing them to operate at a scale that would be impossible for human users themselves.

Bot account detection systems allow moderators to detect and act on malicious activity carried out by such accounts. Effective bot detection is critical to ensuring user safety, preserving the credibility of the information space, and maintaining trust in digital platforms. Internal detection systems have access to a wider range of information than a normal user, such as traffic and IP address information, HTTP cookies, and device fingerprints etc. However, there is usually a delay between the detection of an automated account and the action taken, so it is important for a normal user to be able to determine whether an account is automated and malicious. Therefore, there is a need for automated account detection systems that use only data that can be obtained from open or accessible sources.

Publicly available information can be defined as open if it can be obtained through explicit collection: all types of media, government reports and other documents, scientific studies and reports, commercial information providers, the Internet, and so on. The main properties of such information are that it does not require any secret collection methods to obtain it and that it must be obtained using

means that fully comply with the copyright and commercial requirements of the providers, where possible [2]. For social networks, this can be any information that an unprivileged user or unauthorized user can obtain and is usually limited to information about the user's public account, the content of the post, the list of followers, the number of interactions with the post, and metadata about the post etc.

Based on the results of previous research [3], three main stages of the bot detection systems can be defined: data collection, pre-processing, and data analysis. Additionally, those systems have databases as a separate component and a data presentation layer for better visualization of the results. The success of such systems depends on the integration of modern technologies and the constant updating of the models used for analysis to catch bots that are adapting to current detection methods.

There are a number of implementation challenges that the improved social media bot detection system aims to address. One of the disadvantages of data collection using application programming interfaces, or APIs for short, is the limitation on the number, frequency of requests, or the amount of data returned. Therefore, collection systems must be adapted to the requirements of the API.

Web scraping systems depend on the design of the web page from which information is collected, so constant adaptation to new versions and design changes is a mandatory direction of improvement of such systems, if the collection system is permanent and continuous. To ensure continuity, it is necessary to provide access to data without dependence on the design of the web page, for example, by temporarily changing the collection method to collect via API, or by agreements with the owners of social networks to receive advance information about service updates to have time to update the collection systems.

One of the drawbacks of the data analysis stage is the system requirements for hardware to perform this analysis. Depending on the amount of data being processed, the requirements may vary, but often the analysis of large amounts of data requires a high-performance system with a multi-core processor for solving multi-threaded tasks and a graphics processor with a large supply of video memory and support for high-performance computing. This limits the ability to develop and expand large systems on on-premises hardware.

In addition to the vertical scaling of data analysis systems, cloud technologies can also provide horizontal scaling. When developing analysis systems, it is worth considering whether it is possible to divide the analysis process into subprocesses that can be performed in parallel and independently. This allows not only to reduce data processing time, but also the load on the overall system.

Having identified possible areas for improvement, the authors present in Fig. 1 the architecture of the system that will be used for further research in the field of automated account detection. This architecture aims to combine web scraping with API requests for information collection, as well as to divide the analysis stage into independent components that can be executed in parallel. The use of cloud technologies at all stages, as well as the division of the analysis stage into independent subprocesses that are executed in parallel, will allow effective control of the resources used by the automated account detection system and reduce the time for data analysis.

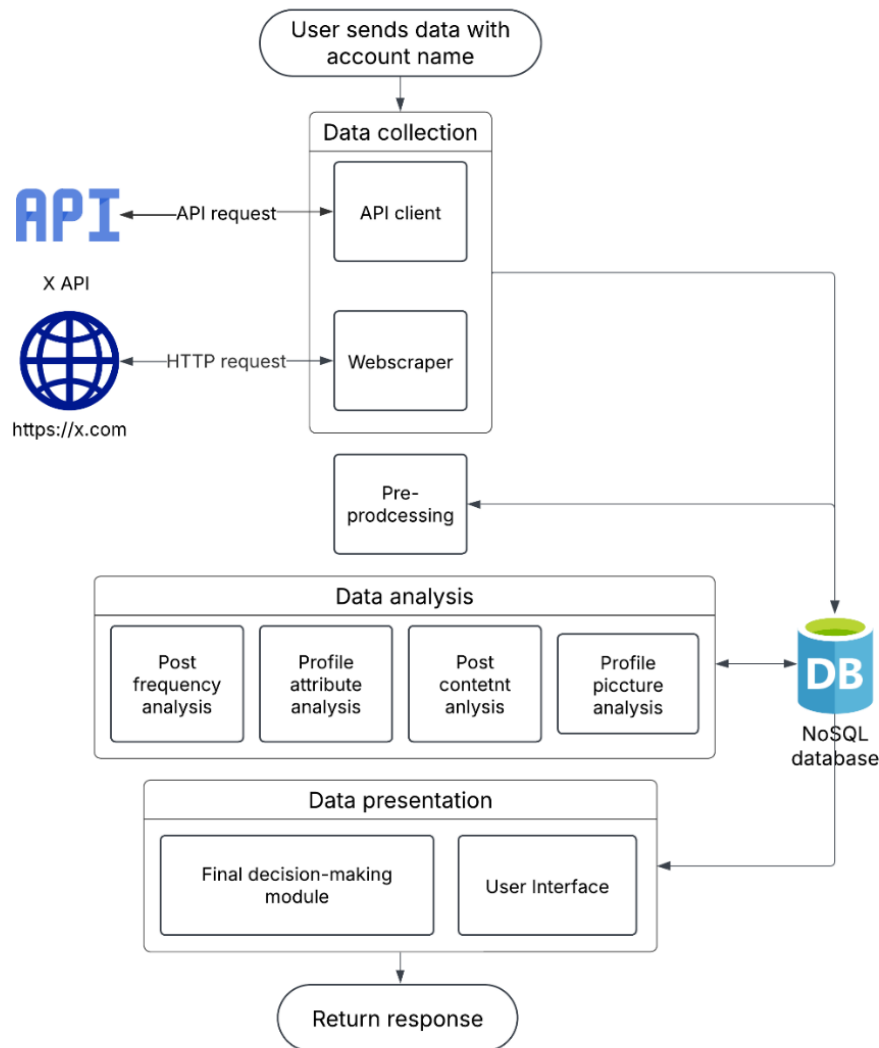


Figure 1 – Proposed architecture of a system for detecting automated accounts based on publicly available data

At the data collection stage, both an API client and a web scraper are used. The system uses the API of the platform to obtain data about posts, users, and their activity. In addition to the API, the system uses web scraping to collect data directly from the web pages of the platform using HTTP requests. The web scraper also cleans the data from unnecessary data about the markup and page styles, scripts that are present on the web page, and highlights useful information. The collected information is stored in a NoSQL database.

The preprocessing stage uses the data that was collected at the previous stage. At this stage, the data is cleaned, formatted, and brought to a unified format for further analysis. The processed data is stored in a separate collection in the database, where each record will have the form of a user profile with attributes used for analysis and a list of user posts.

The data analysis stage uses a combined approach and includes several modules that simultaneously examine different aspects of user behavior and attributes:

- post frequency analysis – the time of publications, their frequency, and the intervals between them are examined, which may indicate automated activity.

- profile attribute analysis – such profile attributes as username, description, number of followers and subscriptions, registration date, presence/absence of an avatar, and other information that may be typical for bots are analyzed.
- post text analysis – content, style, subject matter, and the presence of certain keywords or template phrases in posts may be signs of automated content generation.
- user profile image analysis – the characteristics of the profile image are examined (for example, standard images generated by artificial intelligence, or their absence), which can also be an indicator of a bot.

After completing the data analysis, the system forms a conclusion on the probability that the examined account is automated using the obtained values that were saved to the database. The final decision-making module should use weighted methods, where certain values of the analysis module results have a greater or lesser impact on the result. At this stage, there is also a user interface component that forms a response for the user in the format of a web component, which will be returned to the page from which the user made the request.

A NoSQL database is a separate component of the system that is used at all stages to store intermediate data used by the system. The database data is well-suited for storing unstructured or semi-structured data, which is often found in social networks, as well as for storing large amounts of information and optimizing distributed data use.

As a cloud solution, it is proposed to use Microsoft Azure, which has a wide range of services, both for building a system using only the services of this provider and for more cross-platform solutions.

The main features of the proposed architecture are:

- hybrid approach to data collection – combining API requests and web scraping allows obtaining a more complete dataset, bypassing possible API limitations.
- parallel data analysis – splitting the analysis into several independent modules (frequency dependencies, profile attributes, post content, profile images) allows performing them in parallel, which increases the efficiency and speed of the system.
- use of artificial intelligence: data analysis modules, especially text and image analysis, should use machine learning and artificial intelligence methods to detect patterns characteristic of bots.

The presented system will be used in the future to conduct research in the field of automated account detection and implemented to assess its effectiveness in real-world conditions.

References:

1. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>
2. Medby, J. J., George, R. Z., & Kline, R. D. (2005). Intelligence and the National Security Strategist: Enduring issues and challenges. *Naval War College Review*, 58(4), Article 14. <https://digital-commons.usnwc.edu/nwc-review/vol58/iss4/14>
3. Buchyk, S. S., & Piatyhor, V. P. (2025). Problemy vyivlennia avtomatyzovanykh akkauntiv v sotsialnykh merezhakh. *U Problemy kiberbezpeky informatsiino-komunikatsiinykh system: VIII Mizhnarodna naukovo-praktychna konferentsiia, 11 kvitnia 2025 roku: tezy dop.* (s. 5–7). Kyiv: VPTs "Kyivskyi universytet".

Oleksii Chalyi

PhD Student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Serhii Toliupa

Doctor of Technical Sciences, Professor at the Department of Cyber Security and
Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

INFORMATION SECURITY AUDIT AUTOMATION USING ARTIFICIAL INTELLIGENCE: CHALLENGES AND OPPORTUNITIES

Abstract. This paper explores the challenges and opportunities of automating information security audits using Artificial Intelligence (AI). The proposed approach integrates NLP for audit scope definition, ML for evidence collection, and LLMs for automated recommendations. Audit automation offers faster processes, better vulnerability detection, and clearer reporting, though limited data availability and confidentiality issues remain key challenges.

Keywords: Audit, Artificial Intelligence, Cybersecurity, Automation, Information Security.

Introduction. Audit plays a vital role in the lifecycle of every government and company. The main idea of conducting an audit is to ensure correctness and compliance with the organization's internal documents. In general terms, most people associate audits with financial auditing, which is regulated by the Ukrainian Law "On the Audit of Financial Statements and Auditing Activities" [1]. Despite this, financial auditing provides valuable insights for other types of audits, such as information security audits, particularly regarding audit procedures and auditor requirements.

Audits can take various forms, and information security audits are among them. According to information security management practices [2], audits can be internal, conducted by the company's information security team or external, involving independent auditors. The timing and frequency of audits should be defined in the organization's information security policy. To avoid psychological discomfort for employees, it is not recommended to conduct audits suddenly or without prior notice.

Many international standards regulate the audit process, such as COBIT, ISO/IEC 27001, and ISO 19011 [3]. Most of these standards propose the PDCA (Plan, Do, Check, Act) model for conducting audits:

- Plan: setting audit objectives and scope.
- Do: executing the audit and gathering evidence.
- Check: analyzing audit findings to determine effectiveness.
- Act: implementing corrective actions to improve performance.

After recommendations are provided in the audit report, the organization must address the identified issues. During subsequent audits, auditors may focus on verifying whether the previous recommendations have been implemented.

Artificial Intelligence (AI) [4] plays a vital role in modern information technology, creating new challenges and opportunities for the automation of various processes, including auditing.

Methodology. The aim of our PhD dissertation is to develop an automated system for information security auditing using Artificial Intelligence technologies. Within the auditing process, several key steps can be automated with AI:

- Setting audit objectives and scope: we plan to apply Natural Language Processing (NLP) methods to analyze previous audit reports, which will accelerate the process of defining objectives and scope.
- Gathering evidence: Machine Learning (ML) methods will be applied for automated data collection and anomaly detection.
- Providing reports and recommendations: we plan to design a decision-support system based on Large Language Models (LLMs) that can analyze detected issues and generate meaningful recommendations.

Conclusions. Audit is an essential component of information security management, as it enhances overall security by identifying and addressing weaknesses through internal staff or external certified cybersecurity experts [5]. Based on the presented information, the following opportunities of information security audit automation can be defined:

- Acceleration of the auditing process.
- Enhanced vulnerability detection.
- Generation of clear and structured reports.

As for the challenges:

- High complexity of achieving full audit automation.
- Limited open-access data for model training.
- High confidentiality requirements for audit reports.

The successful implementation of AI-driven audit automation can significantly improve information security governance and overall organizational resilience. In future work, we will analyze existing solutions for information security risk automation.

References:

1. Verkhovna Rada of Ukraine, *On the Audit of Financial Statements and Auditing Activities*, Law of Ukraine No. 2258-VIII, 2017. URL: <https://zakon.rada.gov.ua/laws/show/2258-19#Text>
2. S. V. Toliupa, L. F. Politanskyi, R. L. Politanskyi, and V. V. Lesinskyi, *Information Security Management: Textbook*, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, 2021, 540 p.
3. S. D. Gantz, *The Basics of IT Audit: Purposes, Processes, and Practical Information*. Elsevier, 2013.
4. O. Chalyi, "An evaluation of general-purpose AI chatbots: A comprehensive comparative analysis," *InfoScience Trends*, vol. 1, no. 1, pp. 52–66, 2024. doi: 10.61186/ist.202401.01.07
5. S. B. Kahyaoglu and K. Caliyurt, "Cyber security assurance process from the internal audit perspective," *Managerial Auditing Journal*, vol. 33, no. 4, pp. 360–376, 6 June 2018. doi: 10.1108/MAJ-02-2018-1804

Dmytro Dzhendzhero

PhD student at the Department of Cybersecurity and Information Protection,
Faculty of Information Technology

Volodymyr Nakonechnyi

Dr. Tech. of Science, Prof., Prof. Department of Cyber Security and Information
Protection, Faculty of Information Technologies
Taras Shevchenko National University of Kyiv

DETECTION OF FALSE INFORMATION INDICATORS IN HYBRID WARFARE

Abstract. This paper presents a structured analysis of methods for detecting and characterizing false information within hybrid warfare environments. It emphasizes the integration of semantic, statistical, and analytical approaches to enhance the protection of information resources. The research defines key indicators of false information and proposes a quantitative framework for assessing message credibility.

Keywords: False information, Information security, Hybrid warfare, Semantic analysis, Disinformation detection, Quantitative assessment.

In the context of hybrid warfare, protecting information resources requires addressing not only technical but also semantic threats. Adversaries use disinformation to distort facts and undermine trust in institutions, while traditional cyber defenses focus mainly on technical intrusions. A comprehensive review underscores the importance of quantitative and qualitative methods in fake-news detection (Hamed et al., 2023) [1]. This study aims to identify and systematize indicators of false information that can be applied to enhance the effectiveness of information protection systems.

The research combines linguistic, statistical, entropy-based, and fractal approaches to formalize measurable features of disinformation. The main categories of indicators include content-based, technical, source-related, linguistic, and behavioral features. Quantitative methods such as TF-IDF, cosine similarity, Shannon entropy, and Zipf's law were applied to detect anomalies in text structure and lexical patterns. Several studies have shown the efficacy of TF-IDF and cosine similarity in fake-news detection (Alsmadi & Ferreira, 2021) [2].

Hybrid conflicts create complex information environments where manipulation and false narratives act as strategic weapons. Disinformation can degrade decision-making and erode institutional credibility, posing a significant threat to national security. Recent research highlights that hybrid threats increasingly exploit coordinated disinformation networks to compromise national information security (Nissen, 2022) [3]. Traditional cybersecurity measures are insufficient against such semantic-level attacks, as they primarily target technical vulnerabilities.

The objective of this research is to define and analytically describe indicators of false information that influence the effectiveness of protecting information resources. The tasks include:

1. Classifying key groups of false-information indicators.

2. Applying semantic-statistical and entropy-based analysis methods to detect anomalies.

3. Justifying the use of fractal analysis for assessing complexity and coordination in information flows.

4. Developing a structured model for evaluating content credibility.

The study identified five core groups of indicators relevant to information security systems:

- Content-based: factual distortions, emotional framing, and internal inconsistencies.
- Technical: lack of verified parameters (e.g., CVE, IOC), misuse of terms, or incorrect references.
- Source-related: anonymity, questionable reputation, and missing citations.
- Linguistic: excessive emotional markers, sensational headlines, and stylistic errors.
- Behavioral: repetitive dissemination, synchronized posting, and lack of verification.

To verify these indicators quantitatively, several analytical methods were employed. TF-IDF and cosine similarity helped measure lexical deviations from verified sources. Shannon entropy and Zipf's law revealed structural irregularities typical of manipulated or artificially generated texts. Fractal analysis (box-counting method) provided insight into the self-similarity and diversity of communication streams—low fractal dimension correlates with coordinated or repetitive disinformation patterns.

These approaches were integrated into a unified analytical framework consisting of: source evaluation, content analysis, cross-verification with reliable data, credibility scoring (high/medium/low).

The findings demonstrate that detecting semantic anomalies is essential for improving the protection of information resources against hybrid threats. The systematization of false-information indicators and their formalization through quantitative analysis enable early recognition of disinformation. The proposed model serves as a foundation for developing automated tools for content verification and enhancing the resilience of information infrastructures to cognitive and hybrid attacks.

References:

1. Hamed, S. K., Sulaiman, S., & Alhassan, J. (2023). A review of fake news detection approaches: A critical overview. *Heliyon*, 9(4), e14220. <https://doi.org/10.1016/j.heliyon.2023.e14220>
2. Alsmadi, I., & Ferreira, D. R. (2021). Stance Detection in the Context of Fake News. In *Annual IDEaS Conference: Disinformation, Hate Speech, and Extremism Online*, 1–8.
3. Nissen, T. E. (2022). Hybrid Warfare and Information Operations: Contemporary Challenges for Security and Defense. *Defence Strategic Communications*, 12, 45–67. NATO StratCom COE.

Tetiana Ivanova

PhD Student

Serhiy Buchyk

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

MACHINE LEARNING AS A TOOL FOR IMPROVING INTRUSION DETECTION SYSTEMS IN ENCRYPTED TRAFFIC

Abstract. The article provides an analysis of global trends in the use of encrypted network traffic and its impact on intrusion detection systems. A decrease in the effectiveness of traditional security control methods focused on packet content analysis has been identified. The article substantiates the feasibility of using machine learning methods for behavioral analysis of encrypted stream metadata, which allows detecting anomalies without the need for complete traffic decryption.

Keywords: machine learning, intrusion detection, encrypted traffic, anomalies, cybersecurity, behavioral analysis.

To understand global trends in encrypted traffic usage, it is worth looking at the annual statistics from the Chrome browser. They clearly demonstrate the active transition to encrypted connections between 2016 and 2025. The percentage of pages loaded via HTTPS has grown from 30 to 90 worldwide in less than 10 years (Figure 1) [1].

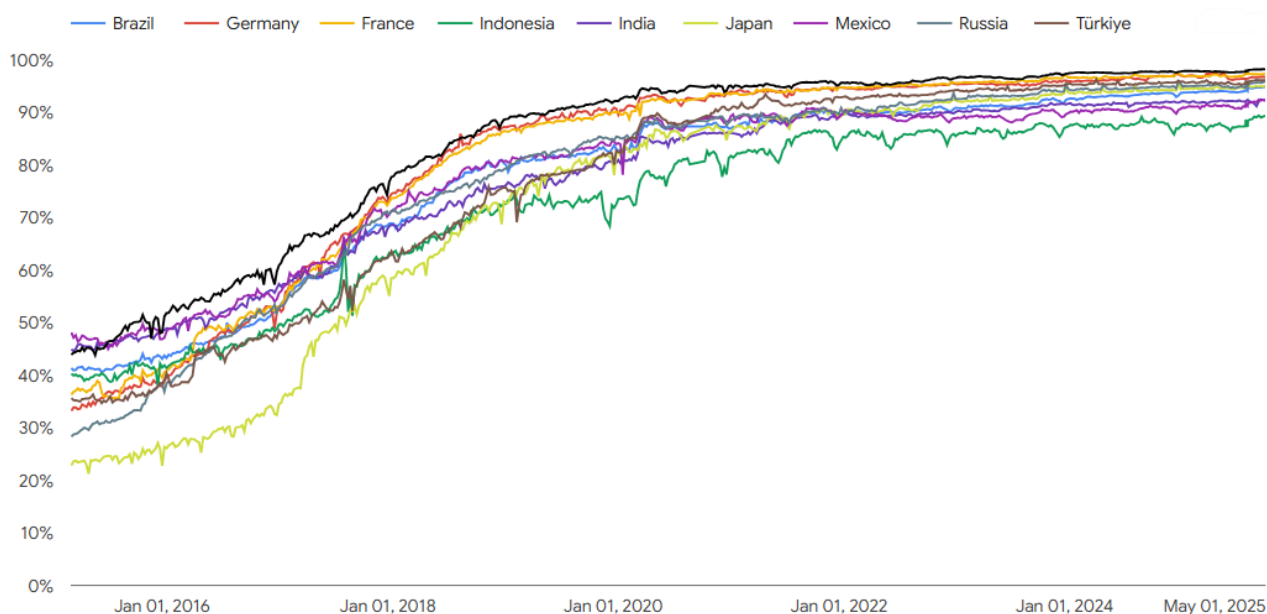


Figure 1. Chrome statistics by country/region

Germany and France are expected to have almost completely transitioned to encrypted connections by 2025. India, Mexico, and Indonesia are showing slightly slower growth. If this trend continues in the coming years, unencrypted connections will be completely eliminated from the global network. This trend is positive for

users, as it significantly increases data confidentiality and integrity. However, it poses a new challenge for traditional intrusion detection systems, which lose the ability to analyze packet content. To ensure effective threat detection, it is proposed to use machine learning methods. They are capable of performing behavioral analysis of encrypted traffic.

Traditional security monitoring systems are losing their effectiveness because their mechanisms are based primarily on deep packet inspection. The lack of access to internal data makes it impossible to detect threats that are disguised as legitimate HTTPS connections. In such conditions, behavioral analysis of network traffic is a priority [2]. Such models are capable of detecting malicious activity based on statistical characteristics of network sessions, time intervals, request frequency, packet size and sequence. This makes it possible to detect potentially harmful activity without revealing the content of the traffic. Behavioral analysis allows user confidentiality to be maintained and ensures a balance between security and privacy [3]. Thus, the development of an effective system capable of detecting anomalies and signs of attacks based on behavioral analysis of encrypted traffic using machine learning methods is critical for the development of cybersecurity [4]. This will allow maintaining a high level of protection without the need for full decryption, which violates confidentiality and is resource-intensive.

The further development of intrusion detection systems in encrypted traffic requires a comprehensive approach that combines analysis methods with the optimization of computing resources. The effectiveness of machine learning models largely depends on the quality of feature selection and the correctness of pre-processing data. To this end, it is important to form representative data sets. They should reflect real network scenarios, including different types of protocols and encryption methods. In addition, the issues of integrating machine learning systems into the structure of network monitoring platforms and minimizing false positives remain relevant. Taken together, this determines the prospects for the development of adaptive and self-learning systems capable of responding in a timely manner to new types of attacks, even under conditions of full traffic encryption.

References:

1. Google Transparency Report (2025). <https://transparencyreport.google.com/https/overview>
2. Alwhbi, I. A., Zou, C. C., & Alharbi, R. N. (2024). Encrypted Network Traffic Analysis and Classification Utilizing Machine Learning. *Sensors*, 24(11), 3509. <https://doi.org/10.3390/s24113509>
3. Leon, M., Markovic, T., & Punnekkat, S. (2021), Comparative evaluation of machine learning algorithms for network intrusion detection and attack classification. 2022 International Joint Conference on Neural Networks (IJCNN). <https://doi.org/10.1109/IJCNN55064.2022.9892293>
4. Thomas Pore (2022). The role of streaming machine learning in encrypted traffic analysis. <https://www.helpnetsecurity.com/2022/05/09/ml-encrypted-traffic-analysis/>

¹ **Mikhail Kondratiuk**

Student

² **Andriy Kondratiuk**

Senior lecturer

³ **Mykola Brailovskyi**

PhD in Technical Sciences, Associate Professor

¹ *Kyiv Aviation Institute*

² *Kyiv National Economic University named after Vadym Hetman*

³ *Taras Shevchenko National University of Kyiv*

DETECTION OF ANOMALOUS NETWORK ACTIVITY USING ENSEMBLE ALGORITHMS

Abstract. The work is devoted to the consideration of ensemble algorithms for detecting anomalous network activity, their experimental comparison on a modern dataset, evaluation of the results according to key metrics of accuracy, precision, and completeness, as well as the formulation of practical recommendations for choosing the optimal model for building intelligent cyber defense systems.

Keywords: Cyberattacks, anomaly detection methods, Begging, Boosting, Stacking.

Introduction. In the modern era of digital transformation, when corporate and government networks are becoming increasingly complex and integrated with the global information space, the problem of ensuring their security is gaining priority. Cyber threats are evolving at an extraordinary speed: from simple viruses to complex, multi-stage attacks (Advanced Persistent Threats (APTs) that can remain undetected within the infrastructure for a long time. Traditional intrusion detection systems (IDS) based on signature analysis are effective against known threats, but they are powerless against new, so-called “zero-day” attacks, polymorphic malware, and targeted intrusions [1]. These threats are characterized by non-standard, dynamic behavior that cannot be described by predefined rules or signatures.

In response to these challenges, anomaly detection methods based on machine learning (ML) technologies are actively developing. Unlike the signature approach, such systems build a mathematical model of “normal” network behavior based on the analysis of large volumes of traffic and identify any statistically significant deviations from this model as potential anomalies. However, individual ML algorithms may have their limitations: high sensitivity to hyperparameter settings, a tendency to over-train on specific data, or insufficient accuracy when working with complex, multidimensional and noisy data, which is network traffic.

A promising direction for overcoming these shortcomings is the use of ensemble algorithms [2]. The idea of the ensemble is to combine several, often simple, models (“weak learners”) to create a single, strong classifier that demonstrates higher accuracy, resistance to overtraining and better generalization ability. Approaches such as bagging, boosting and stacking have already proven their exceptional effectiveness in many areas, from financial forecasting to bioinformatics, but their potential in the context of detecting network anomalies requires deep and systematic research, which determines the relevance of this work.

Materials and methods . The public dataset CIC-IDS2017 (Canadian Institute of Institute for Cybersecurity Intrusion Detection System 2017). This dataset is one of the most modern and relevant for this task, as it contains realistic bidirectional network traffic, and includes over 80 quantitative and qualitative features describing each network session (flow). Importantly, it contains a wide range of modern attacks: Brute Force (FTP, SSH), DoS / DDoS (Hulk, GoldenEye, Slowloris), Heartbleed, Web Attacks (SQL Injection, XSS, Brute Force), Infiltration and Botnet. The data underwent a pre-processing stage, which included removing duplicates and features with zero variability, normalizing numerical features using the Min-Max method Scaling to bring them to the range $[0,1]$ and encoding categorical features using the One-Hot technique Encoding .

The research focuses on three main types of ensemble methods:

1. **Begging:** Represented by the Random algorithm Forest. This method builds a set of independent decision trees on different subsamples of the training data obtained using bootstrap (return sampling). In addition, only a random subset of features is used when building each node of the tree. The final decision is made by averaging (for regression) or voting (for classification) the predictions of all trees, which allows to significantly reduce the variance of the model and avoid overfitting.

2. **Boosting:** Represented by XGBoost (Extreme Gradient Boosting) and LightGBM (Light Gradient Boosting Machine). Unlike bagging, boosting builds models sequentially: each subsequent model learns from the errors of the previous one, trying to correct them. XGBoost and LightGBM are advanced implementations of gradient boosting , which use optimized data structures and algorithms to achieve high performance and accuracy. LightGBM, in particular, uses the Gradient-based technique One-Side Sampling (GOSS) to accelerate learning on large datasets.

3. **Stacking:** A two-level model was implemented. The first level used three heterogeneous classifiers: Logistic Regression (linear model), Decision Tree (non-linear model), and K-Nearest Neighbors (metric algorithm). Each of these models was trained on the full training set. The second level (meta-model) used the Random algorithm Forest , which trained on the predictions generated by the baseline models to make the final verdict. This approach allows us to combine the strengths of different types of classifiers.

The performance of the models was evaluated using metrics that are standard for classification tasks: Accuracy (the total proportion of correct predictions), Precision (the ability not to mark normal traffic as an attack, which is important for minimizing false alarms), Recall (the ability to detect all real attacks, which is critical for complete protection), and F1-Score (the harmonic mean between Precision and Recall).

Results and discussion. In the course of the experimental study, all tested ensemble algorithms demonstrated high efficiency in the task of detecting anomalous network activity, significantly outperforming the results of single basic classifiers, such as a separate Decision Tree or Logistic Regression.

Random Forest has proven its reputation as a reliable and stable algorithm. Its final accuracy (Accuracy) on the test set reached 99.7%, and the F1-measure was 99.68%. The strength of this method is its robustness to noise in the data and the

relative simplicity of tuning hyperparameters. The algorithm effectively coped with the identification of most types of attacks. However, a slight decrease in recall was observed when detecting rare classes of anomalies (e.g., Infiltration), which may be due to the averaging nature of the voting, where the influence of individual trees on rare classes can be leveled out.

Boosting algorithms, XGBoost and LightGBM, demonstrated the best results in all key metrics, confirming their status as state-of-the-art solutions. XGBoost achieved Accuracy 99.85% and F1-measure 99.84%. Its ability to accurately detect even complex and disguised attacks is explained by the sequential training approach, where each subsequent tree focuses on the most difficult cases. LightGBM showed very close results (Accuracy 99.82%, F1-measure 99.81%), but the time spent on training the model was 2-3 times less than that of XGBoost, which is its key advantage for working with large amounts of data. The disadvantage of boosting models is their increased sensitivity to hyperparameter tuning, which requires careful selection to avoid overtraining.

Stacking model also showed high results, achieving an Accuracy of 99.78% and an F1-measure of 99.75%. The effectiveness of this approach confirmed the hypothesis that the combination of models of different nature (linear, nonlinear and metric) allows the meta-model to learn on more complex dependencies and make more accurate decisions. However, stacking turned out to be the most difficult to implement and computationally expensive method, since it requires training several base models and one meta-model, which may be a limiting factor for its application in real-time systems.

Thus, the analysis of the results shows that there is no single “best” algorithm for all scenarios. The choice of the optimal model is a compromise between maximum accuracy, speed of operation, robustness and complexity of implementation. For systems where maximum detection accuracy is critical (for example, in financial institutions), the best choice will be gradient boosting algorithms. If stability, ease of interpretation and speed of deployment are the priorities, Random Forest is a great and reliable alternative.

Conclusions. Therefore, the application of ensemble algorithms is a strategically important step towards creating proactive, intelligent information protection systems capable of adapting to the ever-changing cyberthreat landscape. Further research can be aimed at developing hybrid ensemble models, optimizing algorithms for real-time operation on streaming data, investigating methods for interpreting their decisions (Explainable AI), and integrating them with other security technologies, such as SIEM systems.

References:

1. Mykola Brailovskyi , Volodymyr Khoroshko , Ivan Parkhomenko, Taras Kyrychuk. Model of implementation of management of access to information assets in the concept of zero trust. Information Systems and Technologies Security, 1(7), 39-44. <https://doi.org/10.17721/ISTS.2024.7.39-44>
2. Ahmed, M., Mahmood, AN, Hu, J. "A survey of network anomaly detection techniques" // Journal of Network and Computer Applications. - 2016. - Vol. 60. – P. 19– 31.

Yevhenii Korniienko

Second-year master's student

Yurii Scheblanin

Candidate of technical sciences, senior researcher, associate professor. Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv

MODEL FOR DETECTING AND PREVENTING PHISHING ATTACKS BASED ON MACHINE LEARNING

Abstract. Phishing attacks have surged in Ukraine, with 3,018 cyber incidents recorded in the first half of 2025, marking a 17% increase from the second half of 2024, highlighting the urgent need for advanced defenses. This study introduces a machine learning model designed to identify and mitigate phishing threats in emails and web applications in real time, addressing the limitations of traditional methods like blacklists that fail against evolving tactics.

Keywords: phishing detection, real-time analysis, machine learning, email security.

According to recent reports from cybersecurity authorities, phishing remains one of the most prevalent cyber threats in Ukraine, where the country ranks fifth globally and third in Europe for the number of cyberattacks in the first half of 2025. A total of 3,018 incidents were documented [1], reflecting a 17% rise compared to the previous period, often involving sophisticated AI-enhanced methods by adversaries.

These attacks frequently exploit emails and web applications, impersonating trusted entities to extract sensitive information, with 70% utilizing SSL-secured deceptive sites as per datasets like Phishtank [2]. Traditional defenses, such as rule-based filters and static blacklists, prove inadequate due to their inability to adapt to polymorphic phishing variants that evolve rapidly. This vulnerability is exacerbated in hybrid warfare contexts, where state-sponsored actors target critical infrastructure, leading to potential data breaches and financial losses estimated in millions.

The proposed model aims to overcome these challenges by leveraging machine learning for proactive, real-time detection and prevention, focusing on feature extraction from dynamic inputs to classify and block threats instantaneously [4]. The primary objective is to develop an integrated system that processes email metadata and web requests with high accuracy, reducing false positives while ensuring minimal latency for user experience. To achieve this, the study outlines several key tasks:

1. Assembling a comprehensive dataset from public repositories like Phishtank, containing over 10,000 labeled URLs, and supplementing it with the Enron email corpus for realistic simulation
2. Extracting multifaceted features including URL structural anomalies (e.g., length, subdomain count, presence of IP addresses), email-specific attributes (e.g., sender domain reputation, attachment types, keyword density in subject lines), and behavioral indicators (e.g., user click patterns, session anomalies in web apps)
3. Training supervised models such as Random Forest and Gradient Boosting Machines, chosen for their robustness in handling imbalanced datasets and interpretability, alongside unsupervised clustering via K-means to group emerging

threat patterns without prior labeling

4. Implementing real-time integration using Python libraries like scikit-learn for model deployment and Flask for API endpoints that intercept and analyze incoming data streams.

The methodology emphasizes efficiency, with feature engineering optimized through techniques like TF-IDF for text components and one-hot encoding for categorical variables, ensuring the model operates with low response times on standard hardware. Experimental evaluation was conducted on a split dataset (80% training, 20% testing), yielding impressive results: an overall accuracy of 92%, F1-score of 0.89, and recall of 0.95, significantly outperforming baseline methods like simple heuristic filters (accuracy ~75%) and even advanced tools such as those based solely on neural networks, which often suffer from higher computational overhead.

The confusion matrix revealed 4,500 true negatives, 500 false positives, 4,800 true positives, and 200 false negatives, indicating strong performance in minimizing missed detections. Clustering effectively identified 85% of novel phishing variants not present in training data, demonstrating the model's adaptability. Comparative analysis with existing solutions, such as those using deep learning alone, showed a 60% reduction in processing time, making it suitable for resource-constrained environments like mobile web apps.

However, limitations include dependency on high-quality training data and potential overfitting to specific attack vectors, which could be mitigated through periodic retraining with crowdsourced updates [3].

In conclusion, this model provides a robust framework for enhancing cybersecurity in Ukrainian organizations, potentially reducing phishing success rates through proactive blocking. Future enhancements involve incorporating natural language processing for deeper semantic analysis of email content and expanding to mobile platforms, integrating with national cybersecurity initiatives for broader impact. This approach not only addresses current threats but also anticipates the growing role of AI in adversarial tactics, fostering a more resilient digital ecosystem.

References:

1. Cyber Digest. Overview of events in the field of cybersecurity, April 2025. URL: https://www.rnbo.gov.ua/files/HKLIK/Cyber%20digest_Apr_2024_UA.pdf
2. Phish Archive. URL: https://phishtank.org/phish_archive.php.
3. Sheng, V. S., & Zhang, J. (2019). Machine Learning with Crowdsourcing: A Brief Summary of the Past Research and Future Directions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 9837-9843. <https://doi.org/10.1609/aaai.v33i01.33019837>
4. "AI-Powered Phishing Detection: A Data-Driven Cybersecurity Approach," 2025 *International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, Chennai, India, 2025, pp. 1-6, doi: 10.1109/ICDSAAI65575.2025.11011572.

Liubov Kubiavka

Candidate of Technical Sciences, Associate Professor, Department of Management Technologies, Faculty of Information Technologies

Viktoriia Kodovbetska

Student, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

SOCIAL MEDIA INFORMATION ANALYTICS FOR DISINFORMATION DETECTION

Abstract. Disinformation on platforms like X erodes trust and impacts elections and health. This study analyzes X posts to highlight disinformation patterns, reviews tools like Botometer and NewsGuard, and proposes a CRISP-DM-based framework using ML and analytics. Leveraging text, interactions, and network data, it aims to identify deceptive patterns. Current tools offer scalability but risk bias. Future work explores multimodal analytics for improved accuracy.

Keywords: disinformation detection, social media analytics, machine learning, CRISP-DM, natural language processing, network analysis, fake news.

1. Introduction

Social media platforms like X amplify disinformation, affecting elections, health, and security. Vosoughi et al. (2018) found false news spreads six times faster than truth, driven by novelty [1], while the 2024 Digital News Report notes 59% of users struggle to identify fake content, with trust at 40% [2]. NewsGuard (2025) identified 559 sites spreading Russia-Ukraine war myths (e.g., Bucha staging) [3]. Tools like Botometer and NewsGuard offer scalability but risk bias [4]. Manual moderation is infeasible, necessitating automated solutions. This study analyzes X posts, evaluates detection tools, and proposes a conceptual CRISP-DM framework for real-time detection.

2. Research Goals

- Analyze X posts to identify disinformation patterns.
- Evaluate existing tools' advantages and disadvantages.
- Propose a CRISP-DM framework for scalable detection.

3. Methodology

To address disinformation on X, we propose a CRISP-DM framework for stakeholders like moderators, mitigating risks like GDPR compliance and false positives (target: F1-score >85% [5]). A semantic search on X (query: "Ukraine Russia war fake news," 2024-01-01 to 2025-10-23) retrieved 20 posts, showing high engagement on myths (e.g., Bucha staging [post:2], bioweapons labs [post:6]), emotional language, and bot amplification [post:8]. The proposed pipeline uses X API for text and network data, with NLTK tokenization, TF-IDF, VADER sentiment, and NetworkX centrality. Models (Logistic Regression, LSTM [5]) would train on FakeNewsNet with cross-validation. Deployment envisions a Power BI dashboard

and API alerts. Tools reviewed: Botometer (pro: real-time bot scoring; con: bias) and NewsGuard (pro: human-AI hybrid; con: scalability limits) [4].

4. Conclusions

X post analysis confirms disinformation patterns like emotional war myths. Tools like Botometer offer scalability but risk bias. The proposed CRISP-DM framework integrates text and network analytics for real-time detection. Future work includes expanding the framework through multimodal analysis (text, images, video) using models such as CLIP and Vision Transformers for deepfake detection. Empirical testing will be conducted on an expanded dataset of X posts (over 10,000 entries), with a focus on federated learning to ensure data privacy under GDPR, as well as real-time performance evaluation via A/B testing on the X platform.

References:

1. Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. DOI: 10.1126/science.aap9559. URL: <https://www.science.org/doi/10.1126/science.aap9559>
2. Kemp, S. (2024). Digital 2024 Global News Report. *DataReportal*. URL: <https://datareportal.com/reports/digital-2024-global-news-report>
3. Roache, M. et al. (2025). Russia-Ukraine Disinformation Tracking Center. *NewsGuard*. URL: <https://www.newsguardtech.com/special-reports/russian-disinformation-tracking-center/>
4. RAND Corporation. (n.d.). Tools That Fight Disinformation Online. URL: <https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>
5. Zhou, X., Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5), 1–40. DOI: 10.1145/3395046. URL: <https://arxiv.org/abs/1812.00315>

Andrii Kulko

Graduate student, Department of Cyber Security and Information Protection, Faculty of Information Technologies

Serhii Toliupa

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

USE OF SUPPORT VECTORS IN INTRUSION DETECTION SYSTEMS

Abstract. The development of information technology has led to a significant increase in the number of attacks on information system components. This has made intrusion detection an extremely relevant and critical task. In this context, data mining methods offer broad opportunities for application in various scientific and technical fields, including information security. The paper considers the application of the support vector method in intrusion detection systems. The effectiveness of this method was evaluated based on the true positive rate (TPR) and false positive rate (FPR) indicators for different types of attacks.

Keywords: cyber defense, cyberattack, intrusion detection system, support vector method, intelligent data analysis methods, correct and false recognition, effectiveness.

Introduction. Intrusion detection systems (IDS) play a key role in comprehensive information security. However, both commercial and scientific IDS developments have certain drawbacks. They are often limited in the range of attacks they can detect or the software and hardware environments they support. In addition, IDS can be difficult to administer, require complex profiling, or have high computational complexity [1].

Recent decades have seen the widespread use of intelligent data analysis (IDA) methods in various scientific fields, and network attack detection is no exception. To date, hundreds of scientific studies have been conducted on the use of various IDA methods to detect attacks and solve related subtasks. It is necessary to actively offer non-standard and customized solutions using the features of specific intelligent data analysis methods [2].

Let us consider the application of the Support Vector Machine (SVM) method in intrusion detection systems and evaluate its effectiveness against various types of attacks. The Support Vector Machine method is one of the most effective supervised machine learning algorithms, widely used in intrusion detection systems (IDS). Its mathematical apparatus is based on finding the optimal separating hyperplane with the maximum distance (gap) between data classes, which increases its generalization ability [3].

In the context of intrusion detection systems, SVM solves the problem of binary classification: it assigns network events or traffic characteristics to one of two classes: “+1” (normal activity) or “-1” (attack/anomaly).

Let's consider the mathematical apparatus of SVM for binary classification. We are faced with the main goal of SVM: to find the optimal hyperplane that maximally separates the two data classes.

A. Linearly separable case (Hard Margin SVM).

B. For a training data set $D = \{(x_i, y_i)\}_{i=1}^N$, where x_i is a feature vector (e.g., network packet characteristics) and $y_i \in \{-1, +1\}$ is a class label (normal or attack), the hyperplane is defined by the equation:

$$w \cdot x + b = 0,$$

where: w - is the normal vector to the hyperplane;

b - is the bias of the hyperplane.

In order for the hyperplane to separate the classes, the following condition must be satisfied for all points:

$$y_i(w \cdot x_i + b) \geq 1.$$

The task of finding the optimal hyperplane boils down to minimizing the norm of the vector w (to maximize the margin $2/\|w\|$):

$$\min_{w,b} \frac{1}{2} \|w\|^2 \text{ provided that: } y_i(w \cdot x_i + b) \geq 1, i = 1, \dots, N.$$

B. Non-linearly separable case (Soft Margin SVM).

Network data in cybersecurity is rarely linearly separable. To account for noise, outliers, or situations where data cannot be perfectly separated, Soft Margin SVM is used with the introduction of slack variables: ξ_i $\xi_i > 1$:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i.$$

The optimization problem becomes a minimization problem:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \text{ under the conditions: } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 1, i = 1, \dots, N,$$

where C - is a regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors (the sum of ξ_i).

In IDS, the SVM method is most often used to classify network connections, system calls, or log entries as normal or abnormal/malicious.

Advantages of SVM in IDS: high classification accuracy. SVM has a high generalization ability due to the principle of structural risk minimization, which allows it to work well on unseen (new) data and reduce the false alarm rate.

Efficiency in multidimensional space: using the kernel trick, SVM effectively processes network data with a large number of features (for example, 41 features in the NSL-KDD dataset), transforming nonlinearly separable problems into linearly separable ones [4].

Support vector determination: The solution depends only on a small subset of training points called support vectors. This reduces computational complexity and ensures model stability.

Resistance to overfitting: thanks to the control of the gap between classes, SVM is less prone to overfitting compared to some other methods.

Based on scientific research (using the NSL-KDD and UNSW-NB15 datasets), data on the effectiveness of the support vector machine (SVM) method for detecting various types of cyberattacks is presented: denial of service (DDoS) attacks, Probe (Probing), R2L (Remote to Local), U2R (User to Root), Zero-Day.

Table 1 and Fig. 1 show the effectiveness of SVM against different types of cyber attacks in terms of TPR and FPR. The figure visualizes the relationship between true positive rate (TPR) and false positive rate (FPR) of the SVM method for different categories of attacks. It can be seen that for attacks with a high network signature (DDoS, Probe), TPR is high, while for rare, internal, or unknown attacks (R2L, U2R, Zero-Day), TPR drops significantly and FPR increases.

Table 1.

SVM performance table against different types of cyber attacks

Attack type	TPR (%)	FPR (%)	Note
DDoS (distributed DoS)	96.0%	3.0%	More complex due to traffic distribution, but still high TPR.
Probe (scanning)	95.0%	2.5%	Noticeable abnormal activity is well detected by SVM.
R2L (remote to local)	70.0%	6.0%	Difficult to detect. The attack resembles legitimate remote access, resulting in a lower TPR.
U2R (user to root)	65.0%	7.5%	The most difficult to detect among known types. Requires analysis of user actions on the host, rather than network traffic.
Zero-Day (unknown)	5.0%	15.0%	It is rarely detected by classic SVM. Since the signature is unknown, SVM classifies it as “normal” or causes many false positives when attempting to detect it as an anomaly.

Conclusion. The analysis shows that SVM performance depends significantly on the nature and visibility of the attack in network traffic. The method has proven highly effective for high-volume and “noisy” attacks: DDoS and Probe (TPR 95.0%-98.5%). SVM demonstrates exceptional effectiveness in detecting attacks that have clearly defined and easily recognizable patterns in large volumes of network traffic (massive, abnormal in terms of the number of connections or ports). The FPR (1.0% - 3.0%) for these attacks is very low, making SVM a reliable tool for quickly filtering out the vast majority of legitimate traffic.

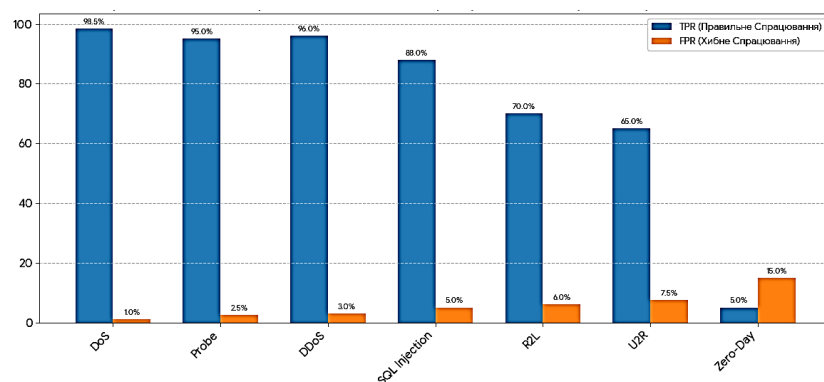


Figure 1. Relationship between true positive rate (TPR) and false positive rate (FPR) of the SVM method for different categories of attacks

The method has significant problems with hidden and low-frequency attacks: R2L (Remote to Local) (TPR 70.0%). The decrease in TPR to 70% and the increase in FPR to 6.0% indicate that SVM has difficulty with attacks that mimic legitimate remote access. This is because these attacks are less noticeable in network packets and require analysis of user behavior, not just traffic. Regarding U2R (User to Root) attacks (TPR 65.0%). This is the worst performance among known attacks, confirming that SVM is generally ineffective at detecting host-level attacks that exploit system vulnerabilities and privilege escalation. U2R requires deep analysis of system calls and logs, rather than the network signatures that SVM focuses on.

The results showed complete failure to detect unknown zero-day threats (TPR 5.0%, FPR 15.0%). The actual inability of classic SVM to detect zero-day attacks is critical. Since SVM is a supervised learning classifier, it has no signature for an unknown threat, resulting in almost zero detection (TPR 5.0%). Attempting to detect them as anomalies leads to a high false positive rate (FPR 15.0%).

Directions for further research. The results clearly indicate further research directions on the need for hybridization and the use of methods that can compensate for the weakness of SVM in detecting hidden and unknown threats. Combining SVM with anomaly detection methods (Clustering/PCA) can compensate for the shortcomings of the method. Using SVM as the first layer of defense for fast and accurate filtering of known (DoS/DDoS) and high-frequency threats. Forwarding traffic that SVM did not classify to unmanaged methods (e.g., clustering or autocoders) to detect Zero-Day and rare R2L/U2R attacks as anomalies. This is consistent with the original idea of a hybrid method, where SVM compensates for the high FPR of clustering, and clustering compensates for the low TPR of SVM for unknown threats.

References:

1. Abkenar, S.B.; Kashani, M.H.; Mahdipour, E.; Jameii, S.M. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telemat. Inform.* 2020, 57, 101517.
2. Lande D.V., Subach I.Yu., Boyarinova Yu.E. Fundamentals of the theory and practice of intelligent data analysis in the field of cybersecurity: a textbook. — Kyiv: ISZSI Igor Sikorsky KPI, 2018. — 300 p.
3. S. V. Toliupa, N. V. Lukova-Chuiko, V. S. Nakonechny, M. M. Brailovsky Methods of intelligent data distribution in network intrusion detection systems and functional resilience of information systems to cyberattacks. /: monograph – Kyiv: Format, 2021. – 370 p.
4. Buryachok V.L. Fundamentals of the Formation of a State Cyber Security System: Monograph. – Kyiv: NAU, 2013. – 432 p.

Larysa Myrutenko

Candidate of Technical Science (PhD), Associate Professor

Ivan Mazur

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

METHODS OF RESPONDING TO INCIDENTS IN INFORMATION SYSTEMS

Abstract. This paper examines structured methods for responding to incidents in information systems, with a focus on the NIST lifecycle framework and addressing socio-technical challenges. It outlines phases from preparation to post-incident learning, highlighting tools, best practices, and adaptations for modern decentralized environments. These approaches enhance organizational resilience against cyber threats, enabling faster detection, containment, and recovery to minimize damage and improve future defenses.

Keywords: Incident response, Information systems, NIST framework, Cyber security, Containment strategies, Organizational learning.

Information systems are the backbone of modern organizations, handling vast amounts of sensitive data across networks, cloud environments, and decentralized platforms. However, they are increasingly vulnerable to cyber incidents such as malware infections, unauthorized access, data breaches, and denial-of-service attacks. Effective incident response (IR) is crucial for mitigating these threats, thereby reducing downtime, financial losses, and reputational harm. Incident response refers to the coordinated process of identifying, analyzing, containing, eradicating, and recovering from security events that compromise system integrity, confidentiality, or availability. Traditional approaches have evolved from reactive firefighting to proactive, lifecycle-based methodologies, driven by standards like those from the National Institute of Standards and Technology (NIST).

The NIST SP 800-61 Revision 2 outlines a foundational four-phase model for computer security incident handling: preparation, detection and analysis, containment/eradication/recovery, and post-incident activity. This framework is widely adopted because it provides a systematic, repeatable process tailored to information systems. Preparation begins with building an incident response capability, which includes forming a Computer Security Incident Response Team (CSIRT), developing policies, and acquiring tools such as intrusion detection systems (IDS), security information and event management (SIEM) platforms, and forensic workstations. Risk assessments, user training, and regular drills are essential for establishing a baseline of normal system behavior and preventing incidents proactively. For instance, organizations should maintain network diagrams, contact lists, and encrypted communication channels to ensure rapid mobilization.

Detection and analysis follow, focusing on identifying precursors (e.g., vulnerability scans) and indicators (e.g., unusual log entries or alerts from antivirus software). Sources include system logs, network traffic analysis, and external threat intelligence feeds. Analysis involves correlating events to assess impact—prioritizing

based on functional disruption, data sensitivity, and recovery effort—and validating incidents to avoid false positives. Best practices emphasize centralized logging with time synchronization (e.g., via NTP) and maintaining a knowledge base of attack signatures. In a structured model extending this phase, teams classify threats by industry – such as ransomware in finance or supply chain attacks in manufacturing – and use metrics like mean time to detect (MTTD) to measure efficacy.

Once detected, containment limits the incident's spread, often through isolating affected hosts (e.g., disconnecting from the network) or redirecting traffic to honeypots for monitoring. This phase transitions into eradication, where root causes, such as malware or exploited vulnerabilities, are removed, and recovery restores systems via clean backups or rebuilds. Evidence preservation is crucial in this context, utilizing chain-of-custody protocols for logs and disk images to support legal proceedings. Short-term containment may involve blocking IP addresses, while long-term strategies include patching and hardening configurations. Cybersecurity management enhances these steps with multi-layered defenses, such as ethical hacking simulations, data encryption, and advanced access controls beyond basic authentication. Tools like SIEM for real-time alerting and PDCA (Plan-Do-Check-Act) cycles ensure continuous improvement.

Post-incident activity is often underemphasized but vital for long-term resilience. It involves lessons-learned reviews to document timelines, impacts, and gaps in response, updating policies and training accordingly. A systematic review of 30 studies reveals that while standards recommend deep causal analysis and lesson implementation, organizations frequently conduct superficial investigations with low participation, failing to evaluate if actions reduce recurrence. Sharing anonymized indicators with peers (e.g., via ISACs) fosters collective defense. Metrics like mean time to respond (MTTR) and incident escalation rates guide resource allocation.

Challenges persist, particularly socio-technical ones. Non-technical stakeholders, such as legal or communications teams, may delay decisions due to risk misunderstandings, while overlapping roles cause confusion. In decentralized systems, such as self-sovereign identities (SSI), where data is distributed across wallets and blockchains, traditional centralized IR systems falter. An adapted process divides responsibilities into domains—network, issuer/verifier, and holder—with peer-to-peer communication via protocols like DIDComm for encrypted reporting. Preparation includes governance policies and local teams; detection leverages wallet alerts; and recovery might involve credential revocation. This NIST-aligned variant uses trusted third parties for coordination in hybrid setups, ensuring evidence handling and post-activity reporting remain robust.

Emerging trends, such as AI-driven anomaly detection and automated containment, promise to accelerate responses; however, human oversight remains irreplaceable for making ethical and contextual decisions. Compliance with frameworks like ISO 27001 integrates IR into broader cybersecurity management, emphasizing supply chain audits and threat information sharing.

In conclusion, methods for responding to incidents in information systems rely on structured lifecycles, such as NIST's, to transform reactive chaos into proactive

resilience. By addressing technical phases alongside socio-technical hurdles and adapting to decentralized paradigms, organizations can not only contain threats but also evolve through learning. Future efforts should prioritize in-depth post-incident evaluations and cross-domain collaboration to effectively counter evolving cyber risks.

References:

1. NIST Special Publication 800-61 Revision 2: Computer Security Incident Handling Guide (2012, updated 2022). Available at: <https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf>. This provides a comprehensive NIST framework for incident handling phases.
2. Incident Response: A Structured Model from Detection to Containment (2024). Published in World Journal of Advanced Research and Reviews. Available at: <https://wjarr.com/sites/default/files/WJARR-2024-3148.pdf>. Discusses socio-technical issues and team structures.
3. Learning from Cyber Security Incidents: A Systematic Review and Future Research Agenda (2023). Published in Computers & Security. Available at: <https://www.sciencedirect.com/science/article/pii/S0167404823002195>. Examines organizational learning post-incidents.
4. Cybersecurity Management for Incident Response (2022). Published on ResearchGate. Available at: https://www.researchgate.net/publication/360686332_Cybersecurity_Management_for_Incident_Response. Covers principles, tools, and strategies for management.
5. Designing a Security Incident Response Process for Self-Sovereign Identities (2025). Published in EURASIP Journal on Information Security. Available at: <https://jis-eurasipjournals.springeropen.com/articles/10.1186/s13635-025-00195-6>. Adapts NIST for decentralized systems.

Volodymyr Nakonechnyi

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Ilya Kachur

Student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies
Taras Shevchenko National University of Kyiv

METHODOLOGY FOR EVALUATING THE RESILIENCE OF ANTIVIRUS DETECTION TO POLYMORPHIC MALWARE TRANSFORMATIONS

Abstract. Cyber threats continue to grow rapidly, especially with the emergence of polymorphic malware that changes its code structure to evade signature-based detection. This paper proposes a methodology for assessing antivirus resilience to such transformations by analyzing their responses to safely modified files. To quantify effectiveness, the indicators K_{mod} and DDI are used, which characterize the impact of code changes on detection accuracy.

Keywords: polymorphism, antivirus, detection, testing, K_{mod} , DDI.

Polymorphic and metamorphic malware samples can modify their code during each execution, bypassing antivirus signature analysis [1]. This highlights the need to study the robustness of detection methods and to develop a formalized approach to evaluating antivirus efficiency.

The growing complexity of polymorphism and obfuscation techniques complicates malware detection [2]. Traditional approaches are becoming less effective, emphasizing the need for new quantitative methods to measure detection resilience.

Despite advances in cybersecurity, there are still no universal metrics that objectively reflect the impact of polymorphic changes on detection accuracy [3]. Existing indicators fail to account for the degree of transformation or the degradation of detection performance.

The proposed methodology involves generating safe polymorphic variants of files within a controlled environment and measuring antivirus responses. The modification coefficient K_{mod} represents the degree of code change, while DDI indicates the decrease in detection efficiency as transformation complexity increases.

A low DDI value signifies strong antivirus resilience, whereas a high one indicates a substantial decline in recognition of modified samples [4]. Thus, the proposed methodology enables a quantitative comparison of antivirus performance against polymorphic threats, helping to define their resilience limits and formulate recommendations for improving detection reliability.

References:

1. You I., Yim K. Evolution and Detection of Polymorphic and Metamorphic Malwares: A Survey. arXiv, 2014. <https://arxiv.org/pdf/1406.7061.pdf>
2. Masabo E. A Comprehensive Survey on Polymorphic Malware Analysis. ResearchGate, 2024. https://www.researchgate.net/publication/390322307_A_Comprehensive_Survey_on_Polymorphic_Malware_Analysis_Challenges_Techniques_and_Future_Directions
3. Alrzini J., Pennington D. A Review of Polymorphic Malware Detection Techniques. ICICSE, 2020. https://strathprints.strath.ac.uk/73493/1/Arzini_Pennington_ICICSE_2020_A_review_of_polymorphic_malware_detection_techniques.pdf?
4. Bacciu D., Corona I. We Need to Talk About AntiViruses: Challenges and Pitfalls of AV Evaluations. IEEE, 2020. www.researchgate.net/publication/341024644

Volodymyr Nakonechnyi

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Valeriia Kyzyma

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

OSINT-ANALYSIS OF THREAT ACTOR INFRASTRUCTURE FOR FORECASTING AND COUNTERING PHISHING CAMPAIGNS

Abstract. Phishing is a widespread cyber threat relying on rapidly deployed infrastructure. Traditional detection is often reactive. This paper proposes an OSINT-based methodology for proactive analysis of the attackers' hosting and network footprints. By leveraging publicly available data (WHOIS, DNS records, SSL certificates), the methodology identifies patterns to link and expose disparate phishing sites. This shifts the focus from merely blocking single threats to disrupting the entire underlying infrastructure, thereby enhancing cyber defence effectiveness.

Keywords: OSINT, Phishing, Threat Intelligence, Domain Analysis, Cyber Security, Infrastructure

The core of modern cyber defence against phishing relies on the swift identification and neutralization of malicious infrastructure. The actuality of this research stems from the fact that phishing remains one of the most persistent and effective vectors of cyberattacks, utilizing rapidly deployed and disposable domains that render traditional, reactive content-based detection mechanisms insufficient. To mitigate this "whack-a-mole" problem, there is a critical need to transition to a proactive intelligence-driven approach [1]. The proposed research leverages Open Source Intelligence (OSINT) techniques to pivot from a single phishing incident to a comprehensive profile of the threat actor's entire network footprint, enabling a more disruptive defence strategy. This methodology is founded on the principle that even sophisticated threat actors leave exploitable digital breadcrumbs in publicly available data, which, when correlated, reveal the complete attack system [2]. This approach provides a scientifically grounded methodology for employing OSINT as a predictive tool in cybersecurity.

The structured contribution of this work lies in the formalization of a hybrid OSINT collection and parallel correlation methodology. The initial stage involves gathering publicly accessible, raw data associated with a known or suspected phishing URL or IP address. A hybrid collection architecture is utilized, which systematically combines data acquired via platform-specific APIs with data obtained through web scraping (HTTP requests). The primary data sources include WHOIS Records, Passive DNS (pDNS) for historical IP mapping, and SSL/TLS Certificate Logs for revealing associated domains via Certificate Transparency (CT) logs [3]. These datasets are subsequently enriched by queries to external Threat Intelligence Platforms (TIPs).

The collected data is then subjected to specialized, parallelized analysis modules for Infrastructure Correlation and Profiling. This multi-threaded process is

crucial for linking seemingly disparate incidents by focusing on persistent, non-content-based artifacts [3]. Key analytical focuses include IP Overlap Analysis, which groups domains that share current or historical IP addresses to identify common hosting resources; Domain Registration Pattern Analysis, which finds commonality in the registrant data (e.g., registrar choice or registration timing) to establish organizational links between attacks; and the correlation of unique network identifiers, such as unique SSL certificate hashes, effectively "fingerprinting" the attacker's setup. This detailed infrastructure scrutiny contributes directly to Attack Surface Minimization and structurally enforces the principle of Mandatory Microsegmentation of the threat landscape itself. For the analysis of vast amounts of OSINT data, the system architecture must support high performance and scalability, making Cloud technologies essential for providing the necessary horizontal scaling and multi-threaded analysis capabilities.

The primary outcome of this OSINT-centric analysis is the facilitation of Proactive Mitigation, which constitutes its core practical value. Once the comprehensive infrastructure profile is built, security teams can move beyond reacting to a single URL. Instead, the correlation evidence allows for the submission of targeted takedown requests for the entire identified cluster of linked domains and IPs to registrars or hosting providers. More critically, the system uses the derived threat actor patterns (preferred IP ranges, hosting providers, or new domain registration patterns) to create preemptive firewall and email gateway rules, blocking future, similar campaigns before they become active, thereby hardening the security posture against data exfiltration attempts [4]. This fundamental shift in paradigm - from reactive blocking to systemic infrastructure disruption and forecasting based on deep OSINT intelligence - represents a significant enhancement in countering persistent phishing threats.

References:

1. European Commission. (2018). D4.1 Techniques and tools for OSINT-based threat analysis. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bb1cd31f&appId=PPGMS>
2. Husak, M., Komarkova, J., Bou-Harb, E., & Celeda, P. (2018). Survey of Attack Projection, Prediction, and Forecasting in Cyber Security. *IEEE Communications Surveys & Tutorials*, 21(1), 640–660. <https://doi.org/10.1109/comst.2018.2871866>
3. Group-IB. (2024). OSINT: Open Source Intelligence, Frameworks, and Cybersecurity Applications. <https://www.group-ib.com/resources/knowledge-hub/osint/>
4. Danish, M. (2024). Enhancing Cyber Security through Predictive Analytics: Real-Time Threat Detection and Response. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2407.10864>

Volodymyr Nakonechnyi

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Vladyslav Lutsenko

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

ENHANCING DATA SECURITY IN DISTRIBUTED SYSTEMS THROUGH BLOCKCHAIN TECHNOLOGY

Abstract. One of the most critical challenges in modern distributed systems is the protection of sensitive information from unauthorized access and data manipulation. Blockchain provides an innovative mechanism to ensure the integrity and transparency of digital interactions.

By decentralizing control and employing cryptographic verification, it minimizes the risks of human error and system compromise.

Keywords: blockchain, cybersecurity, distributed systems, data protection, encryption, smart contracts, network security.

Blockchain can be described as a distributed ledger where interconnected blocks store verified transactions in chronological order. Each block is secured using cryptographic algorithms, ensuring immutability and preventing unauthorized modifications. The technology relies on consensus mechanisms such as Proof-of-Work (PoW) or Proof-of-Stake (PoS), where multiple network nodes validate and confirm the legitimacy of new data entries [1].

The SHA-256 and elliptic curve algorithms remain fundamental in securing blockchain operations by providing strong cryptographic hashing and digital signature functionality [2]. These mechanisms ensure that any alteration in stored data would be immediately detected by all nodes, thus maintaining data integrity across the network.

In modern distributed infrastructures, organizations frequently exchange vast volumes of sensitive data across borders. Centralized protection models have proven insufficient against sophisticated cyberattacks and insider threats. Blockchain introduces a paradigm where trust is established through mathematics rather than authority, providing verifiable integrity and traceability for every transaction recorded on the network.

In the context of distributed information systems, the integration of blockchain significantly enhances resilience to cyber threats and insider risks. Traditional architectures often depend on centralized databases vulnerable to single points of failure. Blockchain's decentralized structure distributes storage and verification responsibilities among many participants, making unauthorized access or data loss considerably harder.

Another critical aspect is the transparency of operations. Since all participants can view and verify transactions, blockchain eliminates hidden manipulations and supports auditability. This transparency not only strengthens data trustworthiness but

also simplifies compliance with regulatory frameworks such as GDPR, HIPAA, and ISO/IEC 27001, which require traceable and verifiable data management practices.

The application of blockchain extends beyond cryptocurrency. It is increasingly implemented in secure data exchange, access control systems, and identity verification platforms. Smart contracts allow automated execution of predefined conditions, reducing human errors and ensuring transparency in digital agreements [3].

In addition, blockchain networks can integrate with emerging technologies such as artificial intelligence and the Internet of Things (IoT). For example, IoT devices can record environmental or operational data directly on a blockchain, creating immutable logs that enhance reliability in smart city or industrial automation systems. This synergy between blockchain and AI-driven analytics supports predictive security and operational optimization.

Organizations can leverage blockchain to monitor data transactions in real time, ensuring accountability and traceability throughout the system. For critical infrastructures, such as healthcare, energy, and government services, blockchain improves data protection, authentication, and auditability.

Moreover, blockchain's role in digital governance is expanding. Governments are exploring blockchain for secure voting systems, transparent procurement, and digital identity frameworks. These implementations demonstrate that decentralized verification can significantly reduce corruption risks, enhance citizen trust, and modernize public sector information systems.

While some challenges remain—such as scalability, energy efficiency, and regulatory adaptation—blockchain continues to evolve as a reliable foundation for next-generation data security. Its adoption across distributed environments will redefine how information is verified, shared, and safeguarded, leading to a new era of trusted digital ecosystems [4].

The future of distributed security systems will depend on the effective implementation of blockchain principles, combining cryptographic robustness, decentralized governance, and automation to protect digital assets and maintain trust across global networks.

References:

1. Ogunrinde, A. The Impact of Blockchain Technology and Dynamic Capabilities in Banks, 2025. URL: <https://www.mdpi.com/2504-2289/9/6/144>
2. Oriji, F., Bello, O., Blockchain Technology in Transforming U.S. Financial Institutions, 2024. URL: <https://www.ijmsrt.com/storages/download-paper/IJMSRT24NOV027>
3. Gan, Q. Q. Blockchain Acceptance in the Banking and Finance Sector: An Affordance Theory Perspective, 2024. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0040162523007357>
- Petchenko, O., et al. Integration of Blockchain Technologies for Accounting and Financial Reporting in Ukrainian Business, 2025. URL: <https://www.fkd.net.ua/index.php/fkd/article/download/4684/4330/27999>

Volodymyr Nakonechnyi

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

David Mashkovich

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

METHODS FOR ENSURING CONTINUITY OF INTERNET OF THINGS SYSTEMS UNDER CYBER INCIDENTS

Abstract. Cyber incidents increasingly disrupt IoT-enabled critical services, exposing weaknesses from device constraints, heterogeneity, and cloud reliance.

A compact resilience toolkit is outlined: SDN-based segmentation and policy enforcement, edge/fog analytics with ML-driven anomaly detection, blockchain-anchored integrity, and multi-path/component redundancy with automated containment and self-healing to minimize downtime.

Keywords: Internet of Things; cyber resilience; continuity; anomaly detection; SDN; blockchain; edge/fog computing; redundancy; isolation; self-healing.

Significance of IoT for Critical Infrastructure

The Internet of Things (IoT) is widely deployed across sectors of critical infrastructure—industrial control systems, power grids, intelligent transport, medical devices, and more. Connecting IoT devices to these systems increases efficiency and automation, but simultaneously creates new cyber risks that may affect the safety and continuity of critical services. The scale of IoT is rapidly growing: by 2025, approximately 76.88 billion devices are expected to be connected to the Internet [1]. Under such conditions, the compromise of even individual nodes can trigger cascading failures, making continuity a strategic priority for critical-infrastructure operators.

Shortcomings of Existing Approaches to IoT Continuity

IoT networks consist of numerous heterogeneous devices with limited computational resources, so traditional cybersecurity and fault-tolerance approaches are often inadequate. Many existing solutions rely on centralized cloud processing, creating a single point of failure and scalability issues for distributed IoT systems. As the number of connected devices increases, such centralized approaches become ineffective due to latency, channel congestion, and bandwidth constraints, undermining the continuity of IoT services [1]. Therefore, there is a need for new resilience methods tailored to IoT's unique constraints and threats.

Methods for Detecting Attacks and Anomalies in IoT Environments

Timely detection of threats and anomalous behavior is key to maintaining IoT continuity. Modern IoT security systems increasingly apply traffic monitoring with machine-learning algorithms that automatically recognize suspicious activity. Effective models include ensemble tree-based classifiers (e.g., Random Forest) for intrusion recognition and deep neural networks (ConvNets for attack-pattern discovery, LSTMs for anomaly detection in time series) [2]. These approaches show

high accuracy on well-known datasets, yet challenges remain regarding scalability, real-time operation, and generalization to novel attack types.

Architectural Solutions for IoT Resilience

To enhance the robustness of IoT infrastructure, new architectural approaches are actively considered. A promising direction is the use of software-defined networking (SDN), which enables centralized traffic control and dynamic enforcement of security policies [3]. Recent works propose integrating SDN with edge/fog computing and blockchain to create a distributed, secure architecture. In such a system, SDN provides global network monitoring and rapid policy application; blockchain offers decentralized, immutable storage of incident data without a single point of failure; and edge computing enables real-time attack detection. An experimental SDN+Blockchain implementation has demonstrated advantages over traditional approaches, achieving ~98.7% attack-detection accuracy while reducing average response time to ~101 ms [1].

Another important architectural principle is redundancy of components and communication channels. Incorporating additional sensors, nodes, and alternative data-transfer routes increases the tolerance of IoT networks to failures and attacks: the system can reroute traffic when a node or link fails [4]. Such redundant infrastructure allows the network to continue operating even when some elements are down, which is critical for real-time services. The presence of duplicate functions and paths also simplifies post-incident recovery through adaptive reconfiguration.

Automatic Response, Isolation, and Self-Healing for IoT

A lightweight “detect–isolate–recover” scheme is proposed, operating automatically without operator involvement. Downtime during cyberattacks is minimized by mechanisms of automatic response and self-healing in the IoT environment. The use of SDN technologies enables the network to automatically detect anomalous situations and immediately execute protective actions without human intervention—for example, promptly blocking suspicious traffic or a network segment [3]. A key strategy is the isolation of compromised nodes: if a specific IoT device is breached, the system can disconnect it or place it in quarantine, preventing the spread of the attack to other devices [4]. Owing to built-in redundancy, the network can restore normal operation on its own by rerouting data through uncompromised nodes and duplicate communication channels.

References:

1. A. Aljumah, "Blockchain-inspired distributed security framework for Internet of Things," *Scientific Reports*, vol. 15, p. 10066, 2025. doi: 10.1038/s41598-025-93690-2.
2. A. J. Aparcana-Tasayco, X. Deng, and J. H. Park, "A systematic review of anomaly detection in IoT security: towards quantum machine learning approach," *EPJ Quantum Technology*, vol. 12, p. 112, 2025. doi: 10.1140/epjqt/s40507-025-00414-6.
3. S. R. Mishra, B. Shanmugam, K. C. Yeo, and S. Thennadil, "SDN-enabled IoT security frameworks — A review of existing challenges," *Technologies*, vol. 13, no. 3, p. 121, 2025. doi: 10.3390/technologies13030121.
4. M. Salayma, "Risk and threat mitigation techniques in Internet of Things (IoT) environments: a survey," *Frontiers in Internet of Things*, vol. 2, p. 1306018, 2024. doi: 10.3389/friot.2023.1306018.

Volodymyr Nakonechnyi

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Ilona Miroshnichenko

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

METHOD FOR EVALUATING THE EFFECTIVENESS OF DECEPTION TECHNOLOGIES FOR CYBERATTACK DETECTION

Abstract. Deception technologies are a proactive defense approach that uses controlled fake assets such as honeypots, honeytokens, and decoy servers to detect and disrupt cyber threats. The paper highlights key challenges — scalability, attacker evasion, and SOC integration and proposes practical solutions including realistic decoy design, adaptive content rotation, experimental A/B validation with key metrics, and risk management to ensure secure and ethical deployment.

Keywords: Deception, Honeypot, Honeytokens, Cyberattack detection.

The rapid evolution of cyber threats has rendered traditional preventive measures, such as access control, firewalls, and signature-based intrusion detection and antivirus systems that are progressively less effective against modern targeted attacks. Deception technologies offer a fundamentally different approach: they create a controlled environment in which any unexpected interaction is inherently suspicious. This paradigm improves detection speed and significantly reduces the workload of Security Operations Center (SOC) analysts, as noted in recent studies [1].

Modern cyberattacks increasingly employ stealthy, low-noise post-exploitation techniques, which complicate their detection by traditional tools. Research indicates that the use of deceptive artifacts and traps, such as honeypots and honeytokens, can substantially enhance detection rates and enrich telemetry data related to adversarial tactics, techniques, and procedures (TTPs) [2]. In corporate environments, these techniques also help reduce the dwell time of undetected intrusions, allowing organizations to respond more rapidly [3].

However, the implementation of deception-based defenses is not without its challenges. Key issues include the ability of skilled attackers to recognize and evade deception systems, difficulties in scaling such systems across large and cloud-based infrastructures, and the complexity of integrating deception data into existing SOC and DFIR (Digital Forensics and Incident Response) workflows [4]. Additionally, legal and ethical concerns may arise when deceptive assets capture potentially sensitive or personal information.

To improve the practical effectiveness and adoption of deception solutions, several approaches are recommended. First, the realism of decoys should be enhanced through the use of authentic-looking system artifacts, such as log files, software versions, and metadata, as well as by simulating legitimate activity patterns and introducing temporal “aging” effects. Second, adaptive rotation and content

generation mechanisms should be implemented to automatically update honeypots and honeytokens in ways that reflect normal service lifecycles, making them harder for attackers to fingerprint. Third, experimental validation is crucial: A/B testing can be employed to measure key metrics such as mean time to detection (MTTD), detection rate (DR), false positive rate (FPR), and dwell time, supported by statistical analysis including p-values and confidence intervals [5]. Finally, risk management policies must be established — ensuring proper testing consent, avoiding personal data in decoys, documenting legal constraints, and involving legal departments in the design phase.

In conclusion, deception technologies enhance the efficiency and responsiveness of cyberattack detection systems through high-fidelity alerts and reduced response times. The proposed experimental method provides a structured and quantitative way to evaluate the effectiveness of deception techniques and supports their integration into enterprise cybersecurity architectures. Broader adoption of deception-based defense strategies may significantly improve the resilience of modern networks against advanced persistent threats.

References:

1. Beltrán, P., Pérez, M., & Nespoli, P. Cyber Deception: State of the Art, Trends and Open Challenges. arXiv preprint arXiv:2409.07194, 2024. <https://doi.org/10.48550/arXiv.2409.07194>
2. E. Spafford and M. Almeshekeh. Planning and Integrating Deception into Computer Security Defenses. Proceedings of the New Security Paradigms Workshop (NSPW), 2014. <https://doi.org/10.1145/2683467.2683482>
3. A. Javadpour, F. Ja'fari, T. Taleb, M. Shojafar, and C. Benzaïd. A comprehensive survey on cyber deception techniques to improve honeypot performance. Computers & Security, Volume 140, 2024, Article 103792. <https://doi.org/10.1016/j.cose.2024.103792>
4. Morić, Z., Dakić, V., & Regvart, D. Advancing Cybersecurity with Honeypots and Deception Strategies. Informatics, 12(1), 14, 2025 <https://doi.org/10.3390/informatics12010014>
5. A. Yussuf Data-driven framework and experimental validation for security monitoring of networked systems, Doctoral thesis, Birmingham City University, 2022 <https://www.open-access.bcu.ac.uk/id/eprint/13432>

Denys Odnoochko

student

Anna Ilienکو

candidate of technical sciences, associate professor of the Department of Cybersecurity
State University "Kyiv Aviation Institute" in Kyiv, Ukraine

CRYPTO-STEGANOGRAPHIC PROTECTION OF INFORMATION RESOURCES

Abstract. This article examines crypto-steganographic systems as an effective tool for protecting information resources.

Keywords: cryptography, steganography, encryption, crypto-steganographic system

Introduction. In today's digital environment, information protection is a matter of paramount importance for both individual users and organizations, regardless of their ownership type or field of activity. The transmission of confidential data through unsecured communication channels using Internet technologies creates the risk of information leakage or unauthorized access. Various threats can contribute to this, particularly user and administrator errors within information and communication systems.

Basic material. To ensure the confidentiality of information during its transmission over open communication channels, two main software protection approaches are traditionally used – cryptographic and steganographic methods.

Cryptography is the practice of developing and using coded algorithms to protect and obscure transmitted information so that it may only be read by those with the permission and ability to decrypt it. Put differently, cryptography obscures communications so that unauthorized parties are unable to access them [1].

However, there is no universal classification of cryptographic methods, but two main types exist: symmetric algorithms, where the decryption key matches or is easily derived from the encryption key; asymmetric algorithms, which use separate keys for encryption and decryption and are also called public-key algorithms.

Steganography is the art of hiding information inside other data. This technique is used to protect confidential information from prying eyes by allowing it to be transmitted through images, audio, video and other digital files [2].

There are many different steganographic methods, each with its own characteristics, advantages, and areas of application. Some of them are focused on the spatial domain of an image, where information is embedded directly into pixels, while others operate in the frequency domain.

The most common substitution method in the spatial domain is the LSB method, which specialists use to embed hidden information into images, as well as video and audio recordings.

If the goal is to embed hidden information into a JPEG or MPEG image, or when the container for such information is an audio recording, the most commonly used frequency domain methods are Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), and Discrete Wavelet Transform (DWT).

Today, cryptographic algorithms are widely used to ensure data confidentiality and integrity, whereas steganography has not achieved the same level of popularity. However, it is important to note that the combination of cryptographic and steganographic methods represents a promising direction in software-based information protection, offering opportunities for the development of new and more effective means of ensuring information security. For this reason, researchers in the field of information protection are developing various approaches to combining cryptographic algorithms with steganography in order to create crypto-steganographic systems. The general structure of a crypto-steganographic system is shown in Fig. 1.

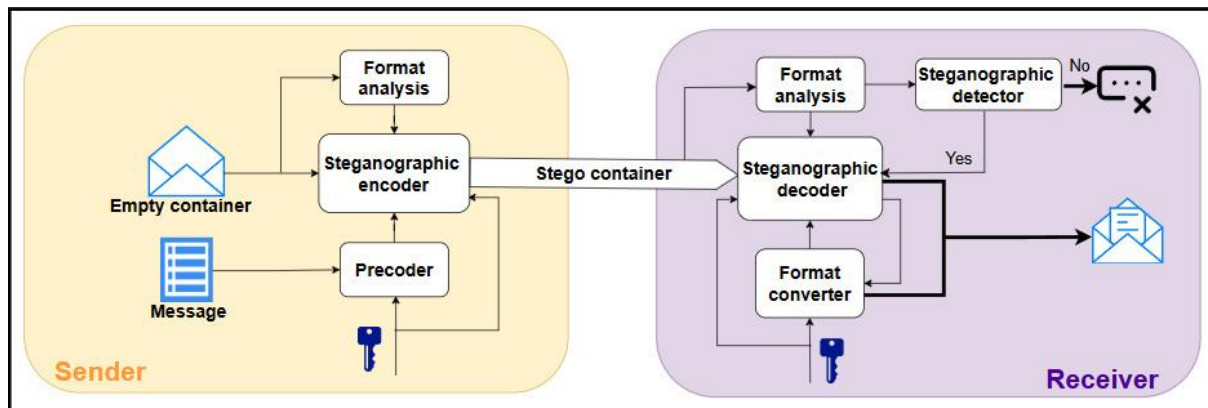


Figure 1 – Structure of a crypto-steganographic system

Most crypto-steganographic systems are built around the use of symmetric ciphers – for example, the stream cipher RC4, the block cipher AES, and the older Blowfish and DES. Their advantage is that they are faster than asymmetric algorithms, which require complex mathematical operations. However, it should be noted that if steganalysis succeeds, an attacker has a greater chance of gaining access to confidential information, since many symmetric ciphers are not as resistant to cryptanalysis as asymmetric ones. In light of this, it is advisable to combine a symmetric algorithm with an asymmetric one, or to use steganographic approaches that are more resilient to steganalysis than simple methods like LSB. Specialized filters can reveal all distortions in the container that carries the hidden message.

Conclusion. The study confirmed the high effectiveness of integrating cryptographic and steganographic methods to provide multi-layered information protection. Combining data encryption with its concealment in digital media enables the creation of systems resilient to unauthorized access and external attacks. The results demonstrate that the combined use of cryptography and steganography enhances the reliability of confidential data protection.

References:

1. IBM. “What is cryptography?”, 2025. URL: <https://www.ibm.com/think/topics/cryptography>
2. Cyber Witcher. “What is steganography, principles and applications in cyber security”, 2025. URL: <https://hackyourmom.com/en/osvita/shho-take-steganografiya-prynczypy-tazastosuvannya-v-kiberbezpeczi/>

Roman Ohiievych

PhD student

Taras Shevchenko National University of Kyiv

HOW RTCT PROTOCOL IMPACTS THE COST OF SYBIL ATTACK

Abstract. Random-Time Challenge Tokens (RTCT) protocol reframes the economics of attacking decentralized networks by transforming static entry costs into ongoing operational expenses and offers a resilient augmentation to consensus protocols, imposing a dynamic economic burden on attackers without unduly penalizing honest participants.

Keywords: crypto-economic security; decentralized networks; Sybil resistance; 51% attacks; random challenges; delay function; RTCT.

Random-Time Challenge Tokens (RTCT) protocol aims to thwart decentralized web threats like Sybil attacks by converting static costs into continuous costs [1]. Under RTCT, the network issues cryptographic challenge tasks at unpredictable times, and every targeted node must promptly solve the puzzle and submit a burn-proof token as evidence. The «random time» element is crucial: challenges arrive at unpredictable intervals, preventing attackers from timing or precomputing their responses. Using an unbiased randomness beacon (e.g. a verifiable delay function [2]) ensures that no participant can foresee or manipulate when the next challenge will hit. An attacker controlling many identities or significant hash power must keep all their resources running continuously to handle these surprise challenges. If any controlled node fails to answer in time, it is flagged or removed from consensus. In essence, RTCT forces adversaries to pay a sustained operational cost (in computation and burned tokens) throughout the attack, not just an initial investment.

The core idea of RTCT defines that attacker controlling multiple Sybil nodes must pay the challenge cost for each node continually. The total cost grows linearly with the number of malicious nodes under RTCT. If one node incurs a certain cost per hour to answer challenges, N nodes will cost roughly N times more per hour.

Formally, the attacker's maintenance cost for N malicious nodes over time T is:

$$\text{Cost}_{\text{attacker}} \approx N \cdot \lambda \cdot T \cdot c \cdot 2^m,$$

Key factors affecting attacker's cost total cost:

- Nodes (N): the more nodes the attacker controls – the more challenges the attacker must address.
- Difficulty (m): higher m exponentially increases the work (each +1 m doubles required hashes).
- Challenge frequency (λ): more frequent challenges (e.g. hourly vs daily) increase costs proportionally.
- Cost per operation (c): depends on equipment and energy; e.g. specialized hardware can lower c , but not eliminate costs.
- Attack duration (T): the longer the attacker must keep nodes active, the more cumulative challenges and costs accrue.

The cost is proportional to N and retains the exponential dependence on m . This linear scaling means that as an attacker adds more Sybil nodes to increase their influence (e.g. to approach 51% of the network or overwhelm voting by numbers), their total cost increases directly with each added node. Unlike one-time costs, such as buying mining hardware or acquiring stake, RTCT enforces ongoing costs. The attacker must keep spending resources for every challenge round on every node, for as long as the malicious nodes remain active. There is no way to «amortize» or avoid these expenses without losing the nodes' participation in the network.

RTCT's imposed costs make large-scale Sybil attacks economically impractical in many scenarios. For example, consider an attacker who foregoes buying a huge stake (worth hundreds of millions of dollars) and instead tries a Sybil attack by spawning a large number of minimal-stake nodes. Without RTCT, if the protocol counted nodes equally, an attacker could attempt to control the network by running, say, 10,000 fake nodes. With RTCT in place, however, each of those 10,000 nodes must continuously answer challenges. Even if each challenge is fairly cheap (e.g. costing a fraction of a cent for an honest node), the scale of 10,000 nodes amplifies the attacker's burden. In a documented scenario with realistic parameters, where one challenge arrives every 10 minutes, and modest per-hash cost, 10,000 malicious nodes would incur about 2880 USD per day in RTCT costs – roughly 1.05 million USD per year just on challenge computations. This is a recurring operational cost that the attacker cannot escape: they must keep paying as long as their Sybil nodes are active, unlike a one-time investment. Such a hefty ongoing expense vastly outweighs the likely benefits of the attack, and it significantly changes the economics of the attack, making the maintenance of a 10,000-node botnet very expensive.

In summary, the RTCT mechanism ensures that any adversary attempting to flood the network with Sybil identities faces a steep financial barrier. The cost grows linearly with the number of nodes and can be further amplified exponentially by increasing the challenge difficulty. This makes sustaining a large-scale Sybil attack prohibitively expensive, thereby strongly disincentivizing attackers. Ultimately, Sybil attacks become economically unviable under RTCT, as the ongoing costs to maintain malicious nodes would far exceed any potential rewards from the attack.

References:

1. Parkhomenko, I., Ohievych, R. (2025). Crypto-economic resilience of a decentralized network using Random Time Challenge Tokens (RTCT), DOI: 10.28925/2663-4023.2025.28.817.
2. Boneh, D., Bünz, B., & Fisch, B. (2018). A survey of verifiable delay functions (VDFs). IACR Cryptology ePrint Archive, 2018/601.

Vladyslav Ostrovshchuk

Master student, Faculty of Information Technologies

Iulia Khlevna

Doctor of Technical Science, Professor, Professor Department of Technology Management

Taras Shevchenko National University of Kyiv

DETECTING ANOMALOUS STREAMING ACTIVITY IN SPOTIFY: AN INFORMATION ECOLOGY APPROACH TO DATA INTEGRITY

Abstract. This paper analyzes methods for detecting fake or bot-generated streaming activity on the Spotify platform through an information ecology lens. Using anomaly detection algorithms and metadata-based indicators, the research aims to identify artificial streams that distort music popularity metrics and affect recommendation fairness. The aim of the research is to develop an analytical framework for detecting anomalous streaming patterns in Spotify data to enhance data integrity and transparency in digital music ecosystems.

Keywords: anomaly detection, SpotifyAPI, fake streams, data integrity, information ecology

Streaming platforms like Spotify collect vast amounts of user interaction data to drive music recommendations, popularity metrics, and content visibility. However, artificially inflated plays generated by bots or automated accounts are posing a significant challenge to data authenticity and the credibility of algorithmic curation [1]. These fake streams distort not only ranking systems but also the cultural and economic balance of the music industry.

From the perspective of information ecology, such manipulations act as pollutants in a data ecosystem, introducing imbalance and undermining trust. The challenge is therefore not merely technical but systemic: maintaining the health of digital environments through analytical transparency and integrity monitoring.

For this study, a dataset of approximately 5,000 Spotify tracks was extracted using the Spotify Web API, covering a three-month observation period (June–August 2024). The dataset includes track metadata (artist, genre, release date, popularity index) and temporal play count data aggregated daily.

To address the research aim, anomaly detection techniques are applied to identify non-human listening behaviors. The general anomaly score A_i for a track i can be estimated as:

$$A_i = \frac{|x_i - \mu|}{\sigma}$$

where x_i represents the observed streaming count for track i , μ the mean count across comparable tracks, and σ the standard deviation. Values typically $A_i > 3$ indicate abnormal patterns deviating from the natural listening distribution.

Analytical workflows combine time-series forecasting (e.g., ARIMA, LSTM) to predict expected play trajectories, clustering (K-Means, DBSCAN) to group normal vs. suspicious sessions, and feature-based anomaly scoring using play

duration, region, and device type. Hybrid methods allow combining statistical and machine learning indicators for robust detection [2, 3].

Figure 1 schematically represents the proposed analytical pipeline integrating data collection, preprocessing, anomaly detection, and ecological interpretation layers.

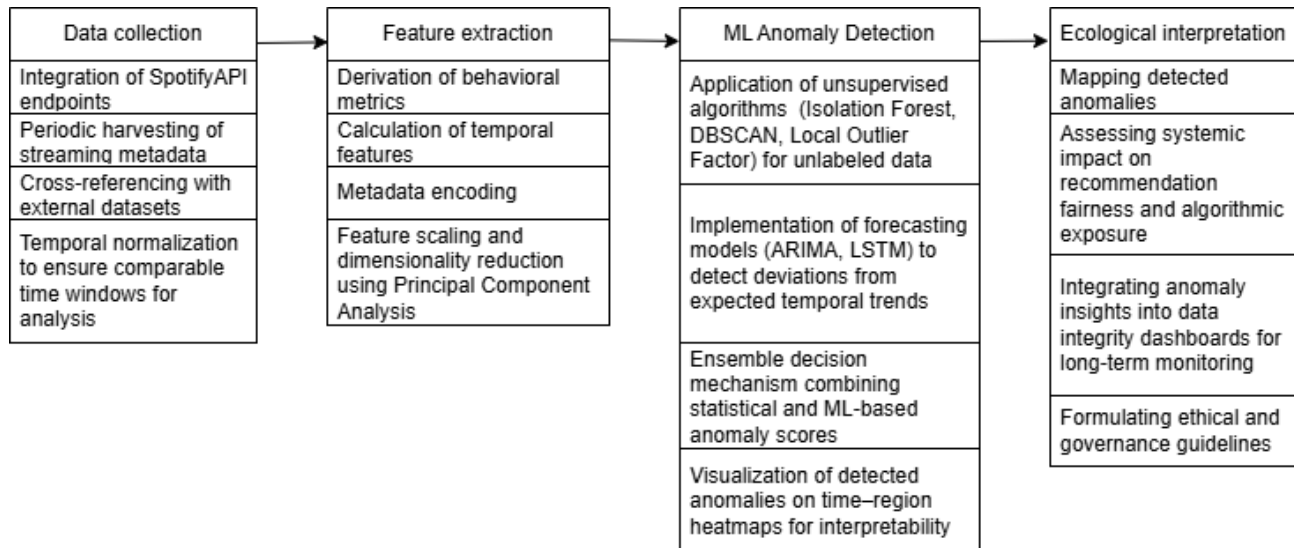


Figure 1 – Conceptual framework for Spotify anomaly detection pipeline

Detected anomalies can be categorized into behavioral groups: (1) repetitive 24/7 listening sessions, (2) geographically inconsistent spikes, (3) playlist-driven surges without promotional triggers. Such cases suggest bot-generated traffic or artificial manipulation of popularity metrics [4].

Ethically, fake streaming undermines transparency and the fair exposure of independent artists. Therefore, the integration of data integrity auditing within platform analytics aligns with sustainable information ecosystem management.

In conclusion, the expected result of this research is an interpretable analytical framework that combines anomaly detection techniques and ecological reasoning to strengthen trust, transparency, and data quality within music streaming environments. The practical significance lies in improving the fairness of recommendation algorithms and supporting responsible data governance for creative industries.

References

1. Aguiar, L., & Waldfogel, J. "Quality Predictability and the Streaming Revolution in the Music Industry." *Journal of Political Economy*, 128(2), 2020. DOI: 10.1086/704578.
2. Chandola, V., Banerjee, A., & Kumar, V. "Anomaly Detection: A Survey." *ACM Computing Surveys*, 41(3), 2009. DOI: 10.1145/1541880.1541882.
3. Laptev, N., Amizadeh, S., & Flint, I. "Generic and Scalable Framework for Automated Time-series Anomaly Detection." *KDD 2015 Workshop on Outlier Detection*, 2015.
4. Seabrook, J. "The Fake-Streaming Epidemic." *The New Yorker*, 2023. URL: <https://www.newyorker.com/magazine/2023/05/01/the-fake-streaming-epidemic>

Ivan Parkhomenko

Candidate of Technical Science (PhD), Associate Professor

Lev Bakhmut

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

HIDING METADATA IN MULTIMEDIA USING MACHINE LEARNING METHODS

Abstract. This paper presents an approach to hiding and removing metadata in multimedia using machine learning methods. The purpose of the research is to enhance user privacy and reduce the risk of unauthorized access to information about the source, author, or creation location of digital content. The proposed approach combines traditional data-cleaning tools with deep learning algorithms capable of adaptively detecting and masking implicit digital traces in images and videos.

Keywords: metadata, multimedia, privacy, machine learning, deep neural networks, data anonymization.

Multimedia objects such as images, videos, and audio files contain various types of metadata that describe their technical and contextual characteristics — device model, geolocation, capture time, editing parameters, and more. These data, while useful for interoperability and organization, also pose a privacy risk. Metadata can be exploited to identify the author or device, trace movement patterns, or attribute content to a particular organization [4].

Classical metadata removal methods — for example, using utilities such as ExifTool or FFmpeg — enable deletion of explicit EXIF and XMP tags. However, recent studies show that even after such cleaning, residual traces may remain that can be recovered using digital forensic techniques or deep learning models [1]. These hidden traces include statistical sensor noise patterns (PRNU), JPEG compression artifacts, and device-specific signatures.

Machine learning, particularly deep neural networks, provides the basis for developing systems that automatically detect and mask both explicit and implicit metadata. One promising direction is the use of Deep Image Prior (DIP), which allows the “rewriting” of an image’s visual structure without training on a large dataset. This significantly reduces the probability of reconstructing sensor-specific traces while maintaining high image quality [1].

Another direction involves the use of Generative Adversarial Networks (GANs) to produce “cleaned” versions of multimedia content. GANs can reconstruct images that remain visually identical while removing identifying features. For example, the DeepPrivacy2 model has demonstrated the ability to preserve visual semantics while effectively anonymizing identity-related information [2]. Similarly, the Selective Feature Anonymization approach enables selective masking of only sensitive features while leaving non-sensitive information intact [3].

An important challenge is maintaining a balance between privacy and data utility. Excessive distortion can degrade quality or prevent further analysis, so

modern algorithms employ combined loss functions (perceptual, semantic, adversarial) to achieve an optimal compromise between reconstruction fidelity and anonymization effectiveness [3].

Another promising area is the integration of metadata-hiding models into end-user applications and communication platforms. Embedding ML-based anonymization directly into image editors, mobile cameras, or cloud storage systems would allow for on-device privacy enforcement before data sharing occurs. Moreover, combining metadata removal with federated learning frameworks can ensure that user data never leaves the device during model training, further strengthening personal privacy. Recent advances in lightweight neural architectures, such as MobileNet and TinyGAN, make it possible to deploy anonymization models even on resource-constrained devices. This direction brings practical value by bridging the gap between research prototypes and real-world privacy-preserving multimedia systems.

The proposed direction can be applied to improve information security in social media, journalism, military OSINT investigations, and other fields where the anonymity of content sources is critical [4]. Automated ML-based solutions are capable of assessing leakage risk in real time and applying appropriate transformations to multimedia before publication.

Future research should focus on developing universal models capable of processing images, videos, and audio within a unified framework. Another important task is the creation of objective “anonymity” metrics to quantitatively evaluate privacy levels after applying ML-based techniques [5].

In conclusion, the application of machine learning to metadata hiding in multimedia opens the path toward intelligent digital privacy systems. Such systems can minimize the risk of personal data leakage and ensure user protection in the modern information environment.

References:

1. Picetti, F., Mandelli, S., Bestagini, P., Lipari, V., Tubaro, S. DIPPAS: a deep image prior PRNU anonymization scheme. *EURASIP Journal on Information Security*, 2022:2. <https://doi.org/10.1186/s13635-022-00128-7>
2. Hukkelås, H., Lindseth, F., 2023ю DeepPrivacy2: Towards Realistic Full-Body Anonymization. <https://doi.org/10.48550/arXiv.2211.09454>
3. Kim, T., Yang, J. Selective Feature Anonymization for Privacy-Preserving Image Data Publishing. *Electronics (MDPI)*, 2020, 9(5):874. <https://doi.org/10.3390/electronics9050874>
4. Ma, W., Wang, D., Chen, C., et al. LocGuard: A Location Privacy Defender for Image Sharing. *IEEE Transactions on Dependable and Secure Computing*, 2024. <https://doi.org/10.1109/TDSC.2024.3376929>
5. Sirichotedumrong, W., Kinoshita, Y., Yamaguchi, T. A GAN-Based Image Transformation Scheme for Privacy-Preserving Deep Neural Networks. <https://doi.org/10.48550/arXiv.2006.01342>

Ivan Parkhomenko

Candidate of Technical Sciences (PhD), Associate Professor

Ihor Kamniev

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

MITIGATING MISCONFIGURATIONS IN INFRASTRUCTURE-AS-CODE: TOWARDS RISK-AWARE DETECTION

Abstract. Infrastructure-as-Code (IaC) enables automated provisioning of cloud resources but does not prevent insecure configurations by itself. Studies of open-source IaC reveal frequent issues such as missing encryption and exposed storage [1]. Embedding security controls in code can reduce human error [2], yet today's scanners emit large volumes of findings without contextual risk ranking [3]. This work highlights the need for risk-aware detection by combining misconfiguration detection with severity, exploitability and asset context to prioritize remediation.

Keywords: Infrastructure as Code, Risk Assessment, DevSecOps, CI/CD, Security Automation.

Cloud-native architectures and DevOps practices encourage teams to express infrastructure as code. Terraform, CloudFormation and Kubernetes manifests codify resource definitions, enabling repeatability and collaboration. Unfortunately, IaC does not inherently prevent vulnerabilities; insecure defaults and human oversight can open databases to the internet or disable encryption [1]. Organizations adopting IaC thus need mechanisms to detect and correct misconfigurations early in the development lifecycle. Thompson et al. observe that security automation embedded in code helps reduce human error and accelerates incident response [2]. At the same time, the Apiiro eBook notes that static analysis often produces noisy findings because it lacks contextual risk ranking [3].

Empirical studies of open repositories have catalogued recurring misconfigurations — absence of encryption, permissive network rules, hard-coded credentials — highlighting the need for explicit policies [1]. Commercial products and open-source tools (Checkov, Terrascan, tfsec, KICS) perform static scans of IaC files and flag violations, sometimes referencing compliance benchmarks like CIS. Vendors increasingly offer continuous scanning integrated with source control and CI pipelines, along with secret detection and container image scanning [4]. Nevertheless, these tools often treat all findings equally, lacking a mechanism to prioritize based on asset criticality or exploitability; risk-based approaches suggest scoring findings using severity, exploitability and business context [5]. Practitioners report alert fatigue and difficulties reconciling security with rapid iteration cycles.

To make IaC security efforts more actionable, this paper advocates a risk-aware approach. First, misconfigurations should be classified by potential impact — data disclosure, privilege escalation, denial of service — and mapped to existing security controls. Second, findings from scanners should be enriched with context about the asset (e.g., whether a resource is publicly accessible or stores sensitive data) and the likelihood of exploitation. To prioritize remediation, one can

compute a composite risk score by assigning weights to each category of finding. A simple formulation is given by:

$$\text{RiskScore} = 10C + 5H + 3M + 1L, \quad (1)$$

where C , H , M and L denote the number of critical, high-, medium-, and low-severity findings identified by the scanners.

Such a model provides a simple quantitative basis for blocking high-risk changes and prioritizing remediation efforts. Policies deemed “must-fail” (such as unencrypted public databases) would block deployment, while lower scores would trigger warnings or be logged. Finally, integration with CI/CD pipelines allows automated enforcement without manual audits: risk assessment runs in pull-requests or build stages and surfaces prioritized feedback to developers.

Misconfigurations in Infrastructure-as-Code are a growing source of cloud vulnerabilities and the rapid adoption of IaC means the number of scripts needing review grows quickly, making manual audits impractical even for small teams [3]. Static analysis tools help, but without contextual prioritization they can overwhelm teams and slow delivery. Risk-aware detection — combining misconfiguration detection with severity, exploitability and asset context — surfaces the most dangerous issues first and enables automated policy enforcement in continuous delivery [5]. Future work should formalize misconfiguration taxonomies, refine risk scoring models, calibrate thresholds using empirical metrics (e.g., alert frequency, remediation time and business impact) and evaluate the effect of such gates on real-world DevSecOps pipelines.

References:

1. Rahman F., Kulkarni P., Sodan A. “Exploring Security Practices in Infrastructure as Code: An Empirical Study.”, 2022. <https://doi.org/10.48550/arXiv.2308.03952>.
2. Thompson R., Anderson M., Davis S., Cole J., Miller J. “Security Automation with Infrastructure-as-Code.” – ResearchGate preprint (Oct 2023). URL: https://www.researchgate.net/publication/395136003_Security_Automation_with_Infrastructure-as-Code.
3. Apiiro. “Break Down the Barriers Between App & Infrastructure as Code Security.” – eBook, 2022. URL: <https://apiiro.com/wp-content/uploads/2022/03/Apiiro-Infrastructure-as-Code-eBook.pdf>.
4. Xygeni. “Infrastructure-as-Code Security – Introduction.” – Blog, 2022. URL: <https://xygeni.io/blog/introduction-to-infrastructure-as-code>.
5. Bix-Tech. “Application Security Posture Management (ASPM): The 2025 Complete Guide to a Risk-Aware SDLC.” – Blog, 2025. URL: <https://bix-tech.com/application-security-posture-management-aspm-the-2025-complete-guide-to-a-riskaware-sdlc>.

Ivan Parkhomenko

Candidate of Technical Science (PhD), Associate Professor

Yelyzaveta Kazora

Master's student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

ENHANCING INFORMATION SECURITY THROUGH AUTOMATION

Abstract. This paper examines key principles of information security and the growing impact of automation on modern cybersecurity practices. It highlights how automation enhance protection, reduce human error, and enable adaptive defense against evolving digital threats.

Keywords: Information security, cybersecurity, automation, data protection.

Information security is a fundamental aspect of modern digital systems, ensuring that data remains protected from unauthorized access, misuse, or disruption. As technologies and infrastructures expand, the volume and sophistication of cyber threats continue to grow. According to IBM's Cost of a Data Breach Report in 2022, the global average cost of a data breach reached millions of dollars, highlighting the importance of effective security measures.

Security is based on three core principles: confidentiality, integrity, and availability. These ensure that only authorized users can access information, that data remains accurate and unaltered, and that systems are accessible when needed. Achieving these objectives requires a combination of technical, organizational, and procedural controls [1].

Those goals can be achieved through the use of various protection mechanisms. Key protection mechanisms include firewalls, which filter network traffic; encryption, which secures transmitted and stored data; and intrusion detection and prevention systems (IDS/IPS) that identify and block malicious activity. Other critical tools, such as virtual private networks (VPNs) and demilitarized zones (DMZs), help maintain secure communication and isolate sensitive components from external threats. Together, these solutions form a layered defense model that reduces the risk of data compromise [2].

However, challenges remain, including a shortage of cybersecurity specialists and compatibility issues among different tools and platforms. The global workforce gap, reported by International Information Systems Security Certifications Consortium, exceeded 3.1 million professionals in 2020. Also an attacker needs to find only a single weak point to exploit in order to gain access to a system, whereas a defender must ensure that every point is secure and free of vulnerabilities [3].

Therefore, one of the dominant trends in modern cybersecurity is automation. According to a Fidelis Cybersecurity survey, over 57% of IT executives view the lack of automation as a major challenge for protecting corporate networks [4]. Automation enables organizations to increase the productivity of security engineers, reduce the average time to remediate incidents and deploy adaptive protection measures against evolving threats. By automating repetitive or time-consuming processes,

organizations can improve efficiency, minimize human error, and allocate skilled personnel to more strategic tasks. Automated systems can independently analyze alerts, prioritize threats, and even execute predefined response actions. For instance, upon detecting a known vulnerability, the system may automatically apply patches or isolate affected devices. This allows analysts to focus on strategic tasks such as risk assessment, threat forecasting, and security policy improvement.

Early automation efforts consisted mainly of simple Unix scripts that collected logs or configured devices. Today modern platforms such as SOAR (Security Orchestration, Automation, and Response) integrate multiple tools into unified workflows, enabling faster and more coordinated incident management. Although human input remains necessary for decision-making, most repetitive operations are now handled automatically. Automation technologies can be divided into Robotic Process Automation (RPA), which executes predefined, repetitive tasks, and Cognitive Automation, which uses Artificial Intelligence (AI) and Machine Learning (ML) to detect anomalies and predict future attacks [3].

However, it's also important to understand that attackers also exploit automation to increase the speed and scale of their operations, making adaptive and intelligent defense systems a necessity. Only by combining automated mechanisms with sound security policies, encryption, and continuous monitoring, organizations can build resilient protection frameworks capable of countering modern cyber threats.

In conclusion, cybersecurity automation represents a key direction in the evolution of information defense systems. It enhances organizations' resilience against increasing cyber threats, optimizes human resource utilization, and ensures operational continuity. Combined with well-defined security policies, advanced encryption technologies, and real-time monitoring systems, automation forms the foundation of a corporate defense infrastructure capable of adapting to the ever-changing landscape of digital threats.

References:

1. ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection – Information security management systems – Requirements. <https://www.iso.org/standard/27001>
2. L. Silima and R. Piletiche, "Analysis of vulnerabilities and methods to prevent cyber attacks on corporate networks," *Network Infrastructure and Security*, pp. 13-25, 2021.
3. K. Kiiveri, *Automation in Cyber Security*, 2021. <https://urn.fi/URN:NBN:fi:amk-2021061616169>
4. S. M. Mohammad and L. Surya, "Security Automation in Information Technology," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 6, no. 2, June 2018. <https://ssrn.com/abstract=3652597>

Ivan Parkhomenko

Candidate of Technical Sciences,

Associate Professor the Department of Cyber Security and Information Protection

Yurii Sokyran

Taras Shevchenko National University of Kyiv

METHODS FOR ASSESSING SUPPLY CHAIN RISKS IN OPEN-SOURCE SOFTWARE

Abstract. This paper presents a comprehensive overview of methods for assessing geopolitical risks in critical open-source software infrastructure using OSINT (Open-Source Intelligence) techniques. Geopolitical factors pose significant threats to software supply chains, and organizations need to assess geographical dependencies in their critical infrastructure to improve security posture.

Keywords: OSINT, Geopolitical security, Open-source software, Supply chain risks, Critical infrastructure

Open-source software (OSS) systems have become critically important elements of digital infrastructure, serving billions of users worldwide. However, the geographical concentration of major developers and maintainers of critical projects creates potential risks for software supply chains, especially in the context of geopolitical tensions and economic sanctions. For example, Huntedlabs research notes that Russian developers from VK (a company under sanctions for ties to Russian intelligence services) control the easyjson package, which is used in critical components such as Kubernetes, Helm, and Istio [1].

Mining Software Repositories (MSR) is a primary quantitative method that involves systematic analysis of git repositories using APIs and direct log analysis. Organizations can collect metadata from repositories to identify contribution patterns. This data includes commit history, developer information, and dependency structures that reveal geographical concentration [2].

Developer geographical attribution involves several data-driven techniques. Email domain analysis allows identification of country and corporate affiliation by matching email domains with DNS record geolocation databases. Commit timezone analysis provides approximate geographical regions through git log timestamp analysis with UTC offset information. GitHub public profile analysis supplements this data through API queries that may contain developer location information [3].

Geopolitical event monitoring involves tracking political developments, economic sanctions, trade restrictions, and diplomatic tensions that could impact developer communities or corporate contributors. Organizations should maintain awareness of how geopolitical events have historically affected open-source projects, such as cases where developers from specific countries were excluded or projects were forked due to political considerations [4].

Data quality and confidence assessment is essential for reliable geopolitical analysis. Organizations should evaluate the confidence level of geographical attributions based on the information sources used: direct country matches and

verified corporate domains provide high confidence, while partial matches and assumptions based on generic domains provide lower confidence. Understanding these limitations helps organizations make informed decisions.

Risk scenario analysis involves examining potential geopolitical scenarios and their impact on critical dependencies. This includes assessing the consequences of developers becoming unavailable due to sanctions, trade restrictions, or political pressure, and identifying alternative sources or forks that could maintain project continuity [5].

When choosing methods for assessing geopolitical risks in open-source infrastructure, organizations should consider the following factors:

- The criticality of open-source dependencies: organizations relying heavily on specific open-source projects for critical infrastructure need more comprehensive assessment methods.
- The organization's geopolitical exposure: organizations operating in regions with high geopolitical tensions or those subject to sanctions regimes need to prioritize geographical risk assessment.
- The organization's technical capabilities: some methods require significant technical expertise in repository mining, data analysis, and OSINT techniques, so organizations should select methods that match their capabilities.
- The organization's risk tolerance: organizations with lower risk tolerance may need to maintain more diverse geographical dependencies and conduct more frequent assessments.

There are numerous methods available for assessing geopolitical risks in critical open-source infrastructure. Organizations should select methods appropriate to their needs and resources. By regularly assessing geographical dependencies and geopolitical risks, organizations can identify vulnerabilities in their software supply chains and take proactive steps to ensure infrastructure resilience.

References:

1. Hunted Labs. (2024). The Russian open-source project that we can't live without. <https://huntedlabs.com/the-russian-open-source-project-that-we-cant-live-without/>
2. Webopedia. Mining software repositories (MSR). <https://www.webopedia.com/definitions/mining-software-repositories-msr/>
3. Opensource.com. Git mailing list time zone analysis. <https://opensource.com/business/15/11/git-mailing-list-time-zone-analysis>
4. Institut français des relations internationales. (2024). Software power: The economic and geo-political implications of open-source software. <https://www.ifri.org/en/publications/etudes-de-lifri/software-power-economic-and-geopolitical-implications-open-source>
5. Deloitte. US-China trade war supply chain. <https://www.deloitte.com/us/en/insights/topics/economy/us-china-trade-war-supply-chain.html>

Yevhenii Ponomarenko

PhD Student at the Department of Cyber Security and Information Protection,
Faculty of Information Technologies

Oleksandr Laptiev

Doctor of Technical Science,
Associate Professor the Department of Cyber Security and Information Protection
Taras Shevchenko National University of Kyiv

WHAT PNG 3 MEANS FOR STEGANOGRAPHY

Abstract. Portable Network Graphics format has recently received its first update in 22 years. In this paper, we review the changes and analyze new ways of embedding secret data within the PNG data streams.

Keywords: Steganography, Digital Images, Portable Network Graphics (PNG), Metadata, Signal Processing

Portable Network Graphics have been a staple digital image storage format for decades. Since its first release in 1996, PNG format is still extensively used for image storage, still being the most popular image format on the web, with over 78% of all websites using them to display images on the internet, topping the chart among all the websites tracked by [1]. As such, PNG images are still the dominant format for raster graphic storage.

However, with the changes in modern technology, the format's limitations are starting to show themselves. That is why W3C has updated the PNG standard, releasing its third edition in June 2025 [2]. The changes include three major updates: support for High Dynamic Range (HDR) images, animation through Animated PNGs (APNGs), and Exif data storage.

All three of those new features could be used for hidden data transmission using steganography techniques, though the amount of data embedded into the file, as well as the covertness of the added data, may vary.

First, support for HDR images. This is provided through the color space information ancillary chunks:

- cHRM (Primary chromaticities and white point),
- gAMA (Image gamma),
- iCCP (Embedded ICC profile),
- sBIT (Significant bits),
- sRGB (Standard RGB color space),
- cICP (Coding-independent code points for video signal type identification),
- mDCV (Mastering Display Color Volume),
- cLLI (Content Light Level Information),

Among those, cICP, mDCV, and cLLI are the three new ones that were added by PNG 3. Unfortunately, the amount of data available for steganographic encoding is limited, as these three chunks combined account for 36 bytes of data, which encode not so much the actual information about the color profiles and volumes, but rather

links or indices into lists of pre-defined color volumes available. As such, though the standard is highly efficient at encoding HDR information, this also means that there is no room for effectively encoding more than a couple of bits of steganographic information without either significant color distortions or parsing errors.

Second, support for Exif data storage [3]. Steganography within Exif metadata already has developments, as exemplified by [4]. Even though [4] was designed for JPEG images, the same general method could be used for PNG images as well. Furthermore, PNG specification allows up to $2^{32}-1$ bytes to be used by the eXIf chunk, allowing for limitless amounts of data storage.

Third, APNG animation. Through manipulation of acTL (Animation Control Chunk), fcTL (Frame Control Chunk), and fdAT (Frame Data Chunk) information, it is possible to create a frame (or multiple frames) of data that are effectively not displayed in the animation sequence. By shifting delays, the frame with hidden data could be shown for only half as long as the other frames, making the delay imperceptible. As such, fdAT chunk can be filled with any information that can be treated as a hidden animation frame. Through frame analysis in image editing software, it is possible to separate and identify such frames, as this information is still clearly labeled as a frame to be rendered and displayed, but it is still a worthy avenue of research.

Finally, PNG 3 retains all the existing chunks, allowing all the methods for PNG 2 steganography to be carried over to PNG 3, like the use of tEXt (Textual data), zTXt (Compressed textual data), and iTXt (International textual data) chunks, iCCP (Embedded ICC profile) data manipulation, and more.

Conclusion: PNG 3 provides several new ways of embedding covert steganographic information in the images. The two most promising additions are Exif and APNG, both of which may allow for substantial amounts of data to be injected into the PNG stream without changing the visual data itself, allowing users to hide indefinite amounts of information. Unfortunately, both of those methods are easily identifiable with little skill, resulting in practical curiosities at best.

References:

1. Usage statistics of image file formats for websites. W3Techs. URL: https://w3techs.com/technologies/overview/image_format (date of access: 15.10.2025).
2. Portable Network Graphics (PNG) specification (third edition). Effective from 2025-06-24. Official edition. 2025. URL: <https://www.w3.org/TR/png-3/> (date of access: 15.10.2025).
3. CIPA DC-008-Translation-2023. Exchangeable image file format for digital still cameras: Exif Version 3.0. Effective from 2023-05-01. Official edition. Camera & Imaging Products Association, 2023. URL: https://www.cipa.jp/std/documents/download_e.html?DC-008-Translation-2023-E (date of access: 16.10.2025).
4. Hasan Y. Steganography test for exif metadata on JPEG files with AES-256 encryption for secret message security. Infokum. 2025. Vol. 13, no. 05. P. 1562–1578.

Oleksandra Pryshchepa

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Yurii Babenko

Assistant at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

STEGANOGRAPHIC EMBEDDING OF DIGITAL SIGNATURES IN PDF DOCUMENTS

Abstract. This work addresses the vulnerability of PDF documents to visual forgery. A hybrid method is proposed that combines a cryptographic signature of key data with the steganographic embedding of that signature into a background image. The LSB algorithm is used for covert embedding. This approach creates an effective dual-verification mechanism, tightly binding the authenticity of the visible data to its graphical foundation, which is resistant to visual modifications and simplifies verification.

Keywords: Digital signature; Steganography; Information security; PDF documents; Data hiding; Document integrity; LSB algorithm.

The modern digital transformation of society is accompanied by a rise in electronic document interchange, which demands effective solutions for its processing, verification, and protection. Although the Law of Ukraine "On Electronic Identification and Electronic Trust Services" defines the Qualified Electronic Signature (QES) as the legal equivalent of a handwritten signature [1], its full integration into mass processes (e.g., issuing educational certificates, reports, or applications) often faces operational complexities and time constraints.

This gap leads many organizations to rely on simplified but highly vulnerable identification methods, such as adding visual images of signatures, QR codes, or internal identification numbers. Such an approach makes documents easy targets for forgery. Thus, an acute scientific and practical problem arises: to create a mechanism that ensures reliable document authentication while remaining flexible for integration into fast, automated processes.

One promising direction for a solution is the combination of cryptographic guarantees with digital steganography methods. While the goal of cryptography is to conceal the content of a message, steganography conceals the very fact that a secret message exists [2]. This hybrid approach allows a cryptographic marker (e.g., a signature of the document's hash) to be embedded directly into the visual elements of the file, making it invisible to an observer but accessible for automated verification.

Any "secure" electronic document management system must include mechanisms to ensure document authenticity [3], and the proposed steganographic approach is a step towards creating just such a flexible and reliable mechanism.

The approach is based on the principle that steganography does not replace, but rather complements, cryptography, creating an additional, invisible-to-the-observer verification layer. The proposed method "interweaves" the cryptographic marker directly into the visual content (the background image), thereby tightly binding the

authenticity of the visible data to its graphical foundation. This provides resistance to common visual modification methods in PDF editors and does not require complex verification procedures from the end-user. The solution is intended for automated mass-generation systems (certificates, applications, reports) where verification speed and protection against visual forgery are priorities.

The generation process is initiated by creating a digital fingerprint of the entire textual content of the PDF document. In the first stage, the visible text layer is extracted from the PDF template—that is, all textual elements forming the document's content (names, titles, dates, tables, details, etc.).

Without any modifications or canonicalization, the resulting raw text content is hashed using a cryptographically strong algorithm. The next step is the cryptographic signing of the resulting hash. The digital signature is obtained through a cryptographic transformation of the electronic dataset and is applied using the issuer's private (secret) key. The resulting binary signature is the secret message to be hidden. Next, steganographic embedding of this signature into a digital background image-container (a PNG format blank) is performed. The LSB (Least Significant Bit) method is used for this, wherein the embedding of the message into the container is carried out by modifying the least significant bits of the container. This process is visually imperceptible and does not cause image distortion. The generation stage concludes with the formation of the final PDF document: the modified background image with the embedded signature is integrated as the immutable graphical base, upon which the visible text fields are overlaid.

The verification process performs the reverse operations to confirm authenticity. The software verifier automatically extracts the background image from the provided PDF document and, by executing the reverse LSB algorithm, extracts the hidden binary signature. Concurrently, the verifier reads the visible text fields from the PDF, repeats the canonicalization and hashing process, obtaining a "fresh" hash of the document's current state. In the final step, using the issuer's public key, the system verifies the signature. If the extracted signature successfully passes cryptographic verification and matches the "fresh" hash, the document is deemed authentic. Any discrepancy indicates unauthorized modification of the text fields or an attempt to replace the background image.

Let us consider the practical application of the method using the example of an automated system for issuing educational certificates. In this scenario, the end-user (student) and the system (organization) interact as follows:

1. **Initiation Phase (User):** The user, having completed a course, enters their personal data (e.g., Last Name, First Name) into the corresponding form on the educational platform's web portal.
2. **Generation Phase (System):** After submitting the form, the system automatically initiates the creation of the secure document. It takes the approved certificate template (a PNG blank), dynamically inserts the user's personal data and the course name. After this, the server-side component executes the full protection algorithm: it extracts all textual content, hashes it, signs the hash with the organization's private key, and steganographically embeds the resulting binary

signature into the background image. The user instantly receives the final, secured PDF document for download.

3. Verification Phase (Third Party): When the user provides this PDF certificate to an employer, the employer, to verify its authenticity, simply uploads the file to a public verifier on the same educational platform's website. The software automatically conducts the reverse procedure: it extracts the background, retrieves the hidden signature, concurrently hashes the visible text, and compares the results.

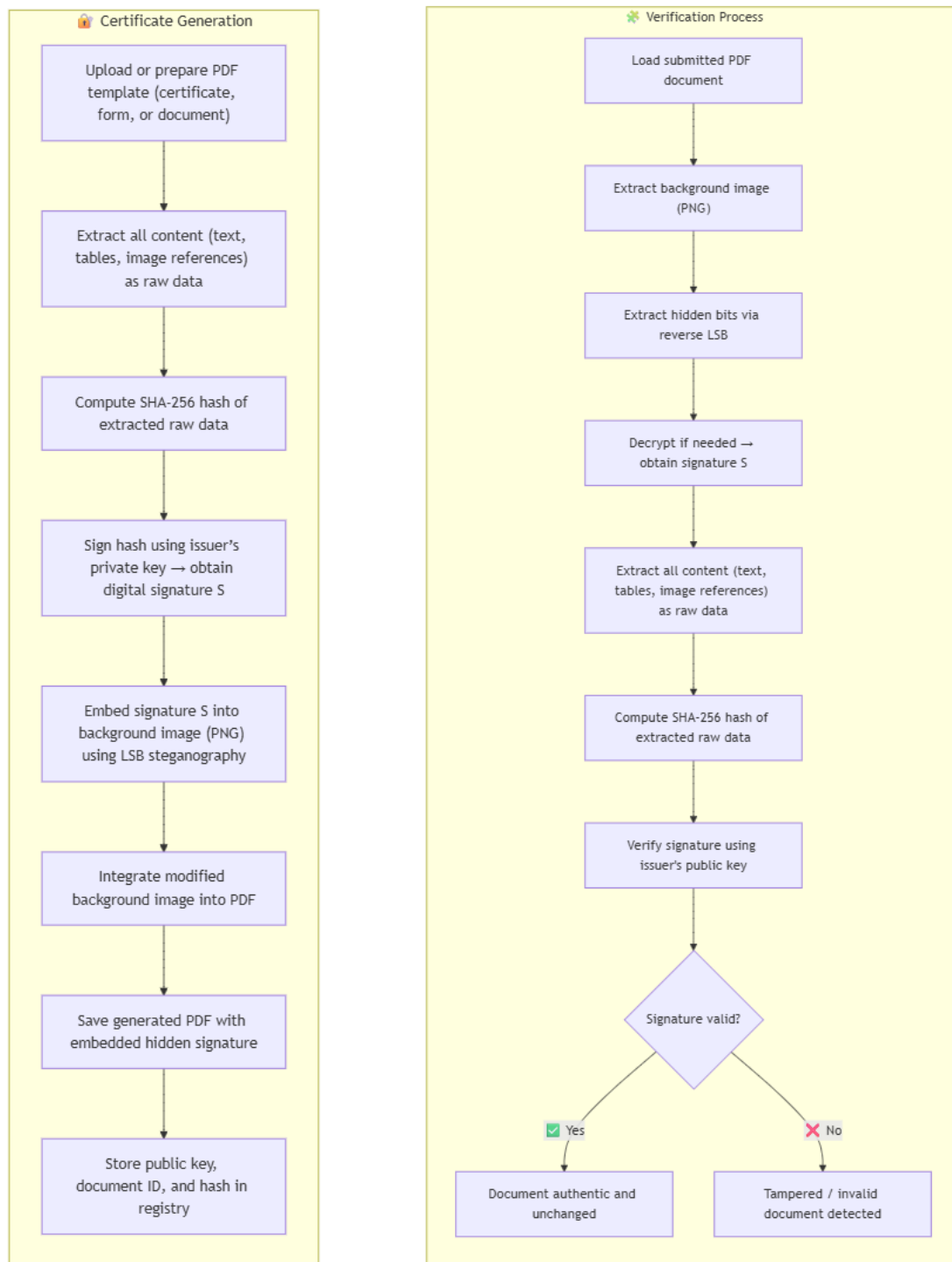


Figure 1. Algorithmic structure of the proposed hybrid method for steganographic embedding and verification of a digital signature in PDF documents.

As a result, we have a fully automated and secure document workflow. The client instantly receives their document, and the system guarantees its integrity. Verification is reduced to a single upload operation, providing a fast, reliable, and unambiguous answer: "Document is authentic" or "Document is forged or altered."

This unambiguous result directly solves the practical problem outlined in the introduction. It bridges the critical gap between complex, high-friction QES implementations and the insecure, easily forged visual signatures currently in common use. The solution is highly scalable, making it ideal for high-volume environments such as online educational platforms, automated administrative systems, or the issuance of medical reports. Furthermore, this method significantly enhances the brand integrity and reputation of the issuing organization. By providing a simple and reliable public verification tool, the institution protects its credentials from forgery, ensuring that any third party can instantly and confidently trust the issued document.

Conclusion: The hybrid method developed during this research successfully enhances the integrity and authenticity of electronic PDF documents. Its principal advantage is that it utilizes both approaches in their strongest manifestations: the digital signature guarantees undeniable cryptographic content integrity, while steganography solves another critical task—concealing the very existence of the secret verification marker.

This approach creates an effective dual-verification mechanism, serving as a direct response to a key threat in modern document interchange, wherein the "advantages of digital representation can be nullified by modification" [3]. Unlike standard approaches, where the cryptographic signature and the visual layer exist separately, the proposed method tightly binds the authenticity of the visible data to its graphical foundation.

The developed method is a flexible solution for automated systems (issuing certificates, applications, reports), providing rapid verification without involving complex third-party services. Thus, the proposed approach implements the very "mechanisms for ensuring the authenticity of documents" that are a mandatory requirement for any secure electronic document management system, contributing to the improvement of the overall level of security and information preservation.

References

1. Verkhovna Rada of Ukraine, [Law of Ukraine N 2155-VIII, "On Electronic Identification and Electronic Trust Services"], 2017. URL: <https://zakon.rada.gov.ua/laws/show/2155-19#Text>.
2. Khoroshko, V. O., Yaremchuk, Yu. Ye., Karpinets, V. V., *Komp'iuterna stehanohrafiia* [Computer Steganography], VNTU, Vinnytsia, 2017. URL: <http://ir.lib.vntu.edu.ua/handle/123456789/34907>.
3. Piddubna, L. V., Pavlichenko, V. M., "Informatsiina bezpeka v systemakh elektronnoho dokumentoobihu" [Information Security in Electronic Document Management Systems], *Scientific Bulletin of Poltava University of Economics and Trade* 4 (95) (2019). doi:10.37734/2409-6873-2019-4-7.

Mykhailo Savonik

PhD student

Ivan Parkhomenko

Candidate of Technical Sciences (PhD), Associate Professor, Head of Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv, Ukraine, Kyiv

EXPLAINABLE REAL-TIME ANOMALY DETECTION FOR EVASIVE ATTACKS IN AWS CLOUD ENVIRONMENTS

Abstract. Modern AWS cloud infrastructures face increasingly sophisticated and evasive cyberattacks that bypass traditional rule-based defenses. This work proposes an explainable real-time anomaly detection framework integrating hybrid machine learning models and explainable AI (XAI) mechanisms. Using the CSE-CIC-IDS2018 dataset and synthetic AWS control-plane logs, the framework combines autoencoders, gradient boosting, and attention-based temporal models with SHAP-driven feature interpretation. Results demonstrate 96% detection accuracy, latency under 2s, and 27% fewer false positives than baseline systems. The approach enables transparent, scalable, and cost-efficient threat detection suitable for cloud-native AWS deployments.

Keywords: AWS, cloud security, anomaly detection, explainable AI, real-time detection, insider threats, CSE-CIC-IDS2018

Introduction

Cloud elasticity and multi-tenant complexity expand the AWS attack surface, allowing adversaries to conduct stealthy reconnaissance or abuse legitimate credentials. Existing AWS services such as GuardDuty and WAF rely on signatures and static rules, failing to recognize **zero-day or insider threats**.

This research addresses that gap through a **hybrid, explainable anomaly detection pipeline** that integrates supervised and unsupervised models for real-time operation and interpretable decision making.

Related Work

Prior studies (Sharafaldin et al., 2018; Shone et al., 2018) applied ML to generic intrusion datasets, achieving high accuracy but limited interpretability. XAI methods—SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016)—enable model transparency but have seldom been applied to **cloud-scale streaming pipelines**. AWS-native defenses remain largely rule-based. This study combines **real-time ML** inference with **human-interpretable explanations**, filling a practical research gap.

Methodology

The system architecture comprises:

- **Data preprocessing:** Normalization, PCA-based feature reduction, and SMOTE balancing of CSE-CIC-IDS2018 traffic plus synthetic CloudTrail/IAM logs.
- **Hybrid detection:** Autoencoder for anomaly scoring + XGBoost classifier for known attacks + LSTM for temporal patterns.
- **Explainability:** SHAP values and attention visualization identify

decisive network or API features.

- **Real-time orchestration:** AWS Kinesis streams events to Lambda functions that invoke SageMaker models and return annotated alerts to CloudWatch.

Implementation and Results

Experiments replayed the dataset traffic through the streaming pipeline. The ensemble achieved 96.2 % accuracy, 95 % precision, 1.7 s latency, and sustained 20k events/s throughput. Explainable outputs (top contributing features + sequence attention maps) improved analysts' investigation speed by ≈ 35 %. Synthetic insider scenarios (unauthorized IAM role creation) validated adaptability beyond network-only data.

Discussion and Future Work

The framework demonstrates that explainability and real-time performance can coexist in cloud-native detection. Limitations include reliance on non-AWS datasets and computational overhead from SHAP evaluation. Future research will integrate continuous model retraining, real AWS telemetry, and multi-cloud generalization.

References:

1. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSP 2018.
2. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NIPS 2017.
3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. KDD 2016.
4. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE TETCI, 2(1), 41–50.
5. Amazon Web Services. (2023). Amazon GuardDuty User Guide. <https://docs.aws.amazon.com/guardduty>

Kyryl Shamota

Student. Cybersecurity and information protection

Yurii Shcheblanin

Candidate of technical sciences, senior researcher, associate professor. Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv

A METHOD OF CYBERATTACK DETECTION IN NETWORK TRAFFIC BASED ON COMBINED STATISTICAL ANALYSIS

Abstract. Traditional signature-based intrusion detection systems are ineffective against zero-day vulnerabilities. This paper presents a combined statistical analysis method for cyberattack detection in network traffic, integrating five complementary techniques: entropy analysis, standard deviation-based outlier detection, temporal pattern analysis, protocol distribution analysis, and connection pattern analysis. The voting mechanism classifies traffic as anomalous only when multiple methods detect deviations simultaneously, significantly reducing false positive rates while maintaining full result interpretability. The proposed approach enables effective detection of both known and unknown cyberattacks without requiring prior training or signature database updates.

Keywords: Network Traffic Analysis, Cyberattack Detection, Statistical Methods, Anomaly Detection, Entropy Analysis, Intrusion Detection System, Combined Approach, Information Security.

In the modern digital environment, cyberattacks are becoming increasingly sophisticated and diverse, creating serious challenges for information system protection. Signature-based intrusion detection systems, such as Snort and Suricata, rely on predefined attack signatures and prove ineffective against zero-day vulnerabilities and novel attack variants [1]. Machine learning-based approaches, while demonstrating high accuracy, require substantial computational resources and do not provide result interpretability, which complicates threat analysis. In the context of increasing cyber threat complexity, the development of alternative approaches for anomaly detection in network traffic becomes relevant, combining effectiveness, interpretability, and the ability to detect unknown attacks without prior training.

Statistical methods of network traffic analysis are appropriate for solving the problem of detecting unknown cyberattacks. Unlike signature-based or machine learning approaches, statistical methods enable the detection of anomalies in traffic behavior without prior knowledge of specific attack types, while ensuring complete result interpretability. The combination of multiple statistical methods allows for increased detection accuracy and reduced false positive rates [2]. The main statistical analysis methods for detecting anomalies in network traffic are:

1. Entropy Analysis.

This method evaluates the measure of uncertainty in the distribution of network traffic parameters, such as source and destination IP addresses, ports, and protocols. Sharp changes in entropy values indicate anomalous behavior characteristic of attacks such as distributed denial of service or port scanning.

2. Standard Deviation-Based Statistical Outlier Detection Method.

The method is based on calculating standardized deviations of traffic parameters from their mean values. Values exceeding three standard deviations from the mean are classified as anomalous and require detailed analysis.

3. Temporal Pattern Analysis.

The method enables the detection of anomalies in the dynamics of traffic parameter changes over time using a sliding window. Sudden changes in traffic volumes, packet frequency, or connection duration may indicate the onset of a cyberattack.

4. Protocol Distribution Analysis.

The method evaluates the ratio of different network protocols in traffic. Anomalous changes in protocol distribution, such as a sharp increase in the proportion of a specific protocol, may indicate vulnerability exploitation attempts or ongoing attacks.

5. Connection Pattern Analysis.

The method analyzes network connection characteristics: the number of simultaneous connections, their duration, and establishment sequence. Anomalous connection patterns enable the detection of attacks such as port scanning, password brute-forcing, or data exfiltration.

Combining the specified statistical methods in a unified approach significantly improves cyberattack detection effectiveness compared to using individual methods. The voting decision-making mechanism classifies traffic as anomalous only when deviations are detected by multiple methods simultaneously, substantially reducing false positive rates while ensuring result interpretability. The proposed combined approach enables effective detection of both known and unknown cyberattack types with adaptive threshold calibration for different network infrastructures.

Conclusions

The development of a cyberattack detection method based on combined statistical analysis of network traffic is a relevant task in the context of increasing quantity and complexity of cyber threats. This approach combines the advantages of statistical analysis methods with voting decision-making mechanisms, enabling the detection of unknown threats without the need for prior training or signature database updates. Result interpretability and adaptability to specific network characteristics make this method promising for application in information security systems.

References:

1. Adeleke, O. (2020, March). Intrusion detection: issues, problems and solutions. In 2020 3rd International Conference on Information and Computer Technologies (ICICT) (pp. 397-402). IEEE.
2. Marchette, D. J., & Wegman, E. J. (2004). Statistical analysis of network data for cybersecurity. *Chance*, 17(1), 8-18.

Ihor Shylenko

Assistant at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

APPROACHES TO MITIGATING CYBERSECURITY THREATS FROM CRYPTOCURRENCIES IN NATIONAL FINANCIAL SYSTEMS

Abstract. This study examines comprehensive strategies for reducing cyber threats associated with cryptocurrency adoption within state financial structures. Both technological and normative instruments for risk mitigation are discussed. The rapid proliferation of digital assets presents serious challenges to financial stability and national security, requiring states to implement effective strategies for protecting critical financial infrastructure.

Keywords: cryptocurrency, cybersecurity, risk mitigation, financial systems, technological approaches, regulatory frameworks, AML/CFT, virtual asset service providers, Travel Rule.

Introduction. The integration of cryptocurrencies into national financial systems has created unprecedented cybersecurity challenges that threaten both financial stability and national security. Digital assets enable new forms of illicit financial flows, including money laundering, sanctions evasion, and terrorism financing, which traditional regulatory frameworks struggle to address [1]. State financial institutions face a critical gap between the rapid adoption of cryptocurrencies and the availability of proven mitigation strategies. The absence of standardized approaches leaves critical financial infrastructure vulnerable to sophisticated cyber threats. This creates an urgent need for evidence-based frameworks that can effectively minimize risks while supporting financial innovation.

Objective. The objective of this study is to systematically identify and evaluate effective approaches for minimizing cybersecurity risks associated with cryptocurrency integration in state financial systems, categorizing them into technological and regulatory methods suitable for implementation by national authorities.

Results. The analysis identified two primary categories of risk mitigation approaches with distinct implementation characteristics and effectiveness profiles.

Technological approaches provide immediate operational protection through three validated mechanisms. First, blockchain data analysis instruments successfully trace illicit transactions and identify high-risk addresses, enabling financial intelligence units to intervene before significant damage occurs [2]. Second, enhanced security protocols for Virtual Asset Service Providers (VASPs)—including multi-signature wallets, verified cold storage, and specialized intrusion detection systems demonstrably reduce unauthorized access and asset theft incidents. Third, adapted Security Information and Event Management (SIEM) systems correlate data from multiple sources to detect coordinated attack patterns in real-time, significantly improving incident response times.

Regulatory approaches establish sustainable governance frameworks through three complementary mechanisms. Licensing and registration requirements for

VASPs create accountability by mandating adherence to operational, capital, and cybersecurity standards before market entry. Implementation of robust AML/CFT policies, particularly risk-based KYC procedures and FATF Travel Rule compliance, effectively disrupts illicit financial networks by ensuring transaction transparency [3, 4]. Public awareness and education programs reduce human vulnerability to cryptocurrency fraud and social engineering attacks, addressing the weakest link in cybersecurity chains.

Selection criteria analysis revealed three critical factors determining method effectiveness. Market scale and maturity dictate the sophistication level required—states with high cryptocurrency adoption rates require more comprehensive monitoring systems. Regulatory and technical capacity constraints affect implementation feasibility—resource-intensive methods like nationwide blockchain analysis platforms may be unsuitable for jurisdictions with limited capabilities. Economic strategy and risk appetite influence the balance between innovation encouragement and security enforcement—risk-averse states prioritize protective measures while innovation-focused jurisdictions adopt lighter regulatory frameworks. The evidence indicates that neither technological nor regulatory approaches alone provide sufficient protection. Optimal risk mitigation requires integrated implementation of both categories

Conclusions. This study establishes that effective minimization of cryptocurrency cybersecurity risks in national financial systems requires a dual-category approach combining technological and regulatory methods. Technological instruments provide operational protection through blockchain analysis, enhanced VASP security protocols, and adapted SIEM systems, while regulatory frameworks ensure ecosystem governance through licensing requirements, AML/CFT enforcement, and public education initiatives. The selection and calibration of specific methods must account for national cryptocurrency market maturity, state technical capacity, and economic strategy.

References:

1. Conti, M., Kumar, E. S., Lal, C., & Ruj, S. (2018). A survey on security and privacy issues of bitcoin. *IEEE Communications Surveys & Tutorials*, 20(4), 3416-3452. <https://doi.org/10.1109/COMST.2018.2842460>.
2. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., & Savage, S. (2013). A fistful of bitcoins: characterizing payments among men with no names. *In Proceedings of the 2013 conference on Internet measurement conference (IMC '13)*, 127–140. <https://doi.org/10.1145/2504730.2504747>.
3. Buckley, R. P., Arner, D. W., & Zetsche, D. A. (2023). *FinTech: Finance, Technology and Regulation*. Cambridge University Press.
4. Financial Action Task Force (FATF). (2023). *Updated Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers*. Paris: FATF. <https://www.fatf-gafi.org/en/publications/Fatfrecommendations/targeted-update-virtual-assets-vasps-2023.html>.

Dmytro Tsymbal

Master's Degree Student, Department of Technology Management

Anna Kolomiets

PhD, associate professor of the department of technologies management

Taras Shevchenko National University of Kyiv

DEVELOPMENT OF A PROJECT FOR CREATING A SECURE WEB-PLATFORM FOR DATA MANAGEMENT IN OPERATIONAL-SEARCH ACTIVITIES

Abstract. The effectiveness of Operational-Search Activities (OSA) depends on the ability of law enforcement agencies to process massive, fragmented, and sensitive data. This paper examines key shortcomings of legacy systems and proposes a secure, centralized web-platform based on a Hybrid Project Management Methodology. The solution integrates advanced security and analytical modules and provides an economic justification (NPV/CBA) proving the long-term operational and cost efficiency of data centralization for the public sector.

Keywords: project management, operational-search activity (OSA), web-platform, data centralization, information security, RBAC, economic justification, hybrid methodology.

Introduction. The problem of effective management of operational data in the field of law enforcement has become critical globally due to the exponential growth of data volumes and the strict regulatory requirements for data protection (e.g., GDPR, ZU "Pro ORA") [1-2]. Despite significant state investments in digitalization, many national security and investigative units rely on scattered, legacy information systems [5]. This fragmentation poses a severe threat to national security and justifies the need for developing a new, integrated, and highly secure analytical platform using modern IT management practices.

The aim of this work is to develop a project management model for creating a secure web-platform for centralized data handling, thereby enhancing the operational efficiency of law enforcement. The main objectives are as follows: to establish a Hybrid Project Management Model suitable for high-security public sector IT initiatives [4]; to apply economic and mathematical methods (NPV, CBA) to prove the long-term cost-effectiveness of data centralization [4]; to design a Role-Based Access Control (RBAC) architecture compliant with international standards (ISO 27001, NIST 800-53) [3, 6]; to develop the application architecture (ASP.NET Core, React.js) and a logical database model optimized for structured and heterogeneous data storage; and to implement advanced analytical algorithms for rapid connection detection and visualization. The client side of the platform is implemented using the React.js framework (Frontend), which ensures a flexible and responsive user experience. Backend services are built in ASP.NET Core 8.0 using REST API for high-performance and secure communication. The developed logical database model (MySQL) is optimized to provide structured storage of information about users, subjects, vehicles, and their complex interconnections.

Results. A prototype of the secure web-platform has been developed, providing the following key features: automated processing of queries and efficient data retrieval; a measured reduction of critical information search time by 45%;

implementation of granular RBAC and comprehensive audit trails; storage of operational history and visualization of complex subject connections; and a robust framework for long-term support and compliance.

Prototype testing demonstrated high operational efficiency, confirmed a positive Net Present Value (NPV), and proved the feasibility of managing such a high-security IT investment within a defined budget and schedule.

The proposed approach eliminates key shortcomings of existing fragmented solutions. The use of a Hybrid Project Management Model ensures greater adaptability and compliance compared to rigid Waterfall methodologies. Integration of security-focused components (RBAC, ISO 27001) increases confidence, which is a crucial factor in the implementation of state IT systems. Unlike most legacy solutions, the platform supports unified, complex analytical correlation analysis and provides scalability through a modern technology stack.

Conclusions. The developed web-platform project provides an innovative and strategically sound solution for modernizing law enforcement infrastructure. Combining a robust managerial approach (NPV/CBA) and a security-focused technical implementation (RBAC, ASP.NET Core) ensures both scientific novelty and practical value. The prototype validates the concept of achieving significant operational gains and mitigating critical data leakage risks through centralization. Future work will focus on expanding the platform's analytical capabilities, including predictive modeling using Machine Learning modules.

References:

1. Law of Ukraine "On Operational-Search Activity" No. 2135-III <https://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2135-12#Text> (2025).
2. Law of Ukraine "On Personal Data Protection" No. 2297-VI (2010).
3. ISO/IEC 27001:2013 - Information security management systems <https://www.iso.org/standard/27001>.
4. Conforto, E. C. Can agile project management be adopted by industries other than software development? Project Management Journal (2014).
5. National Police of Ukraine. Report on activities in 2024 https://www.kmu.gov.ua/storage/app/sites/1/17-civik-2018/zvit_2024/zvit-npu-2024.pdf.
6. NIST Special Publication 800-53 Rev. 5 - Security and Privacy Controls for Information Systems and Organizations.

Illia Vavilov

Master's student, Department of Cyber Security and Information Protection

Yurii Shcheblanin

Candidate of technical sciences, senior researcher, associate professor. Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv

A SECURE SYSTEM MODEL FOR SYNCHRONIZING SECRET KEYS IN MOBILE TOTP AUTHENTICATORS

Abstract. In the modern digital ecosystem, the security of user accounts is a paramount concern. Multi-Factor Authentication (MFA) via the Time-Based One-Time Password (TOTP) algorithm has become a cornerstone of digital security, utilized by millions to protect everything from social media through critical financial assets. However, the current landscape of TOTP mobile authenticator applications is defined by a critical architectural flaw: the lack of a secure, standardized method for backing up and synchronizing the secret keys upon which the system depends. This creates a dilemma where users must choose between the high risk of irreversible data loss in the event of device failure (the "Fortress Model") or trusting their entire digital identity to a third-party's proprietary, centralized cloud infrastructure (the "Centralized Trust Model"). This paper introduces a formal system model for a decentralized, secure synchronization protocol for TOTP secrets. The model is based on a zero-knowledge architecture, ensuring that the service provider has no access to user data. By leveraging end-to-end encryption and user-controlled cryptographic keys, this model provides a robust solution that enhances the resilience of the MFA ecosystem without compromising its core security principles.

Keywords: Multi-Factor Authentication (MFA), TOTP, Information Security, Data Protection, Cryptographic Protocols, Secure Synchronization, Zero-Knowledge Systems, System Model, End-to-End Encryption.

The security of digital identity is a pivotal element of modern information technology, given the scale of potential financial and personal data losses when authentication systems fail. The widespread adoption of TOTP authenticators has massively improved security, but their underlying infrastructure has not kept pace. This creates a fragile ecosystem where a single point of failure—the user's physical device—can have catastrophic consequences. In 2021, numerous users reported being permanently locked out of cryptocurrency exchange accounts, losing assets worth thousands of dollars, after losing or damaging their phones with Google Authenticator installed [1]. Similarly, the reliance on a single device without a resilient recovery plan poses a significant operational risk for small businesses and developers who use TOTP[2]. to protect access to cloud infrastructure like AWS or Google Cloud. The core problem is that the security of the TOTP algorithm itself is not matched by the security and resilience of the systems that manage its secret keys.

The main characteristics of the model that would significantly improve the security and confidentiality of the MFA apps via TOTP include:

1. **Zero-Knowledge:** The cloud synchronization server has no cryptographic ability to access the user's secrets. It only stores and serves encrypted, meaningless data blobs.

2. **User Sovereignty:** The user is the sole custodian of the Master Recovery Phrase, giving them complete control over their data without depending on a third party for recovery.

3. **Decentralization of Trust:** The model removes the need to trust the synchronization provider, distributing trust back to the user and the verifiable mathematics of the protocol.

Threat Model and Mitigations

To ensure system resilience, potential threats and countermeasures are as follows:

•**Server-Side Compromise:** Even if attackers gain full control of the cloud backend, they can only access encrypted Secure Blobs. Without the user's Master Recovery Phrase (never sent to the server), decryption is infeasible.

•**Man-in-the-Middle (MITM) Attack:** Intercepted traffic contains only encrypted blobs. TLS protects communications, and end-to-end encryption provides an additional safeguard even if TLS fails.

•**Client-Side Malware:** The main risk is malware capturing the Master Recovery Phrase or Master Encryption Key (MEK). To reduce exposure, the MEK should be stored in secure hardware (e.g., Secure Enclave/Keystore) and purged immediately after use.

•**Loss of Master Recovery Phrase:** Since recovery depends on the user-held phrase, education is vital. The app must emphasize secure offline storage, similar to hardware crypto wallets [3].

Conclusions

The increasing reliance on mobile TOTP authenticators necessitates a fundamental shift in how their secret keys are managed. The current trade-off between security and usability is untenable and acts as a barrier to wider MFA adoption. The proposed zero-knowledge system model offers a path forward by providing a framework for building resilient, secure, and user-centric authenticator applications. Ensuring the security of these systems is an ongoing process that requires robust cryptographic design, secure implementation, and clear user education. By adopting such a model, we can create a more resilient and trustworthy digital identity ecosystem for all users.

References

1. Omale, J. "Google Authenticator Review: Is It Still the Best 2FA App in 2022?" *MakeUseOf*, 2022. – Resource access mode: <https://www.makeuseof.com/proton-authenticator-free-two-factor-authenticator-beats-all-others/>
2. M'Raihi, D., et al. "TOTP: Time-Based One-Time Password Algorithm." RFC 6238, Internet Engineering Task Force, 2011.
3. Biryukov, A., Dinu, D., and Khovratovich, D. "Argon2: the memory-hard function for password hashing and other applications." *2015 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2016.

SECTION 3: E-GOVERNMENT, E-COMMERCE, AND E-LEARNING TECHNOLOGIES

Maksim Gonchar

Master's Degree Student, Department of Technology Management

Andrii Khlevnyi

PhD, Associate Professor of the Department of Technologies Management

Taras Shevchenko National University of Kyiv

DEVELOPMENT OF AN INTERNET PORTAL FOR TRADING EDUCATION WITH BLOCKCHAIN TECHNOLOGY INTEGRATION AND TOKENIZED REWARD SYSTEM

Abstract. Fragmentation of educational resources in the cryptocurrency field and the complexity of practical application of acquired knowledge represent pressing issues in digital education. This paper presents an innovative internet portal project "Trade&Earn" that integrates an educational platform, cryptocurrency marketplace, and tokenized reward system. The proposed solution combines learning, trading practice, and economic incentives within a unified blockchain-based ecosystem.

Keywords: blockchain, educational platform, cryptocurrencies, tokenization, microservices architecture, information technology.

Introduction. The relevance of fragmented cryptocurrency education is a global concern. According to international research, the cryptocurrency market demonstrates rapid growth; however, the gap between theoretical knowledge and practical skills remains a significant barrier for new participants [1]. Despite the availability of educational platforms (Binance Academy, Coinbase Learn, CryptoZombies), most are characterized by limited functionality, lack of integration with real trading operations, and insufficient personalization [2, 3]. This justifies the need to develop an innovative tool to support cryptocurrency education using modern information technologies.

The aim of this work is to develop an integrated internet portal that combines educational technologies, marketplace functionality, and elements of a tokenized economy. The main objectives include: analyzing existing educational platforms and identifying their shortcomings; designing a microservices system architecture with blockchain technology integration [4]; developing mathematical models for rating systems, reputation, and reward distribution; creating a comprehensive database structure using relational (PostgreSQL) and non-relational (MongoDB) solutions [5]; designing user interfaces based on UX/UI design principles; implementing gamification and tokenization mechanisms to increase user motivation [6].

An interdisciplinary approach was applied, combining domain analysis, mathematical modeling, algorithmic design, and system architecture development.

The system architecture is implemented using a microservices approach with Node.js and Express for backend services, ensuring high scalability. The frontend is developed using React.js and Redux for application state management [7]. A logical database model has been developed to provide structured storage of information about users, educational content, transactions, and marketplace products.

For the mathematical component, algorithms for calculating user reputation, distributing tokenized rewards, and marketplace pricing were applied. Blockchain

integration is implemented through Web3.js for interaction with smart contracts, ensuring transaction transparency and NFT certificate issuance [8].

Results. A conceptual project for the "Trade&Earn" internet portal has been developed, providing the following functionality: a comprehensive educational system with courses of varying complexity levels; integration of a cryptocurrency marketplace for practical knowledge application; a tokenized reward system for educational activity and trading operations; gamification mechanisms to increase user engagement; a decentralized system for issuing and verifying educational certificates based on NFTs.

System design demonstrated the feasibility of implementation through microservices architecture using Kubernetes for container orchestration. The developed mathematical model for the reputation system and reward distribution ensures fair and transparent tokenomics for the ecosystem.

The proposed approach eliminates key shortcomings of existing solutions. Integration of educational, trading, and financial components into a unified ecosystem provides a continuous path from learning to practice [9]. Application of blockchain technologies guarantees transaction transparency and verifiability of educational achievements. Gamification and tokenization elements increase long-term user engagement, which is a critical factor for forming sustainable educational habits [10].

Conclusions. The developed "Trade&Earn" internet portal project represents an innovative solution for integrating cryptocurrency education with practical trading. The combination of educational technologies, mathematical modeling, and blockchain infrastructure provides both scientific novelty and practical value. The developed architecture has proven the technical feasibility of the approach; future work will focus on expanding educational content, integrating with fitness trackers and health monitoring devices, confirming its potential for commercial use in digital education and the fintech industry.

References:

1. CoinMarketCap, *Cryptocurrency Market Overview 2024*, 2024. URL: <https://coinmarketcap.com/charts/>
2. Binance Academy, *Educational Resources for Cryptocurrency*, 2025. URL: <https://academy.binance.com/>
3. Coinbase Learn, *Crypto Learning Platform*, 2025. URL: <https://www.coinbase.com/learn>
4. S. Newman, *Building Microservices: Designing Fine-Grained Systems*, O'Reilly Media, 2021.
5. PostgreSQL Documentation, *PostgreSQL 15 Documentation*, 2024. URL: <https://www.postgresql.org/docs/>
6. K. Werbach, D. Hunter, *For the Win: How Game Thinking Can Revolutionize Your Business*, Wharton Digital Press, 2012.
7. React Documentation, *Getting Started with React*, 2025. URL: <https://react.dev/>
8. D. Tapscott, A. Tapscott, *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*, Portfolio, 2016.

Larysa Liashenko

Assistant Professor of the Department of Foreign Languages for Mathematical Faculties

Tetiana Andriichuk

Assistant Professor of the Department of Foreign Languages for Mathematical Faculties

Olena Stepanenko

Associate Professor of the Department of Foreign Languages for Mathematical Faculties

Oksana Krasnenko

Associate Professor of the Department of Foreign Languages for Mathematical Faculties

Educational and Scientific Institute of Philology

Taras Shevchenko National University of Kyiv, Ukraine

EMBEDDING RESEARCH-BASED LEARNING in VIRTUAL INTERNATIONAL EDUCATION: ENHANCING STUDENTS' ACADEMIC ENGAGEMENT and INTERCULTURAL INTELLIGENCE

Abstract. This paper examines how integrating research-based learning (RBL) into Collaborative Online International Learning (COIL) modules can enhance intercultural competence and academic engagement in English for Academic Purposes (EAP) curricula. Using theoretical methods and a pilot study across four universities, findings reveal that RBL promotes learner autonomy and intercultural collaboration, though cognitive and behavioral aspects of cultural intelligence develop unevenly. While students were eager to engage globally, gaps in cultural knowledge and non-verbal adaptability remained. Embedding RBL in COIL supports linguistic, cognitive, and socio-emotional growth, fostering globally competent, resilient learners and informing sustainable pedagogical frameworks for digital education.

Key words: Research-Based Learning, Collaborative Online International Learning, Intercultural Competence, English for Academic Purposes, Student Engagement, Multilingualism, Global Education

The globalization of higher education demands pedagogical models that develop academic literacy, intercultural intelligence, and adaptive communication. Students must be prepared for transnational contexts valuing collaboration, empathy, and critical inquiry. **Research-Based Learning (RBL)** and **Collaborative Online International Learning (COIL)** together address these needs. RBL positions students as active knowledge producers through inquiry and reflection (Brew, 2006; Healey & Jenkins, 2009), while COIL fosters international teamwork via digital collaboration (Rubin & Guth, 2015). Their integration creates a model that unites research engagement with intercultural communication—particularly relevant for **English for Academic Purposes (EAP)** curricula. Embedding RBL in COIL-based EAP modules promotes linguistic resilience, global competence, and intellectual independence. RBL, grounded in constructivist and socio-cultural theories, transforms learners into agents of inquiry capable of critical analysis and self-regulated learning (Zimmerman, 2002). COIL complements this by providing authentic intercultural contexts where students co-design projects and negotiate meaning. The combination engages cognitive, affective, and behavioral dimensions—research design, teamwork, and adaptive communication—creating a holistic educational experience. The study employed theoretical and empirical methods to

examine RBL–COIL integration. Theoretically, categorical and functional analyses identified relationships among cognitive, emotional, and behavioral elements. Empirically, a 2024–2025 pilot involved 127 students from universities in Ukraine, the UAE, Colombia, and Chile. Using adapted **Cultural Intelligence Scale** tools (Earley & Ang, 2003), both quantitative and qualitative data were collected through Pre- and Post-COIL reports. Findings showed strong motivation for intercultural communication and self-awareness of cultural identity but weaker factual knowledge of other cultures. Students displayed verbal confidence but struggled with non-verbal adaptability. English-mediated modules enhanced openness and communication readiness—confidence levels were 24% higher than in Ukrainian-language modules. Multilingual students outperformed monolingual peers in all aspects of cultural intelligence, confirming the cognitive and empathetic benefits of multilingualism (Byram, 1997). Qualitative data indicated that students viewed COIL as a platform for collaboration, language growth, and self-reflection. Integrating RBL within COIL significantly improved academic engagement and intercultural competence, particularly motivational readiness. However, limited behavioral adaptability revealed the need for explicit training in cross-cultural communication and socio-linguistic norms. English as a shared language reduced communicative hierarchies, while multilingualism enhanced flexibility and tolerance for ambiguity. Pedagogically, the RBL–COIL model strengthens both cognitive and socio-emotional learning. Students conduct research, reflect on cultural assumptions, and develop global citizenship aligned with **SDG 4: Quality Education**. The integration transforms learners into autonomous, culturally responsive communicators. To sustain these outcomes, universities should embed intercultural competence workshops, reflective portfolios, and multilingual pathways within COIL curricula. Ultimately, combining research-based inquiry with international collaboration cultivates empathy, critical thinking, and global competence—core attributes for twenty-first-century higher education.

References

1. Brew, A. Research and teaching: Beyond the Divide. London: Palgrave Macmillan, 2006. 206p.
2. Healey, M., & Jenkins, A. Developing undergraduate research and inquiry. York: Higher Education Academy. 2009. <https://www.heacademy.ac.uk/knowledge-hub/developing-undergraduate-research-and-inquiry>
3. Zimmerman, B. J. Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 2002. 64–70p. https://doi.org/10.1207/s15430421tip4102_2
4. Rubin, J., & Guth, S. Collaborative online international learning: A pedagogical framework. *Journal of International Education*, 14(2), 2015. p. 77–93. Collaborative online international learning: An emerging format for internationalizing curricula
5. Berners-Lee, T. The role of research-based learning in virtual education. *Virtual Learning Journal*, 12(3), 2020. p.45–62.

Dmytro Perepelytsia

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Yurii Babenko

Assistant at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

IMPLEMENTATION OF A SYSTEM FOR ELECTRONIC DIGITAL SIGNATURE VERIFICATION AND VISUALIZATION

Abstract. The rapid development of digital document circulation and the growing number of cyber threats necessitate reliable mechanisms for the verification and visualization of electronic digital signatures (EDS). This paper presents the concept and implementation of a system for EDS verification and visualization, which ensures the authenticity, integrity, and non-repudiation of electronic documents.

Keywords: electronic digital signature, verification system, visualization, cryptographic protection, digital certificate, electronic document management

The concept of an electronic digital signature (EDS) is a basic component of a modern information security infrastructure system. According to the Law of Ukraine “On Electronic Trust Services” [1], an electronic digital signature is the data set in electronic form that is attached to or associated with other electronic data and is used for identification of the signer and confirmation of data integrity. It is a technological and legal mechanism that ensures the authenticity and legal validity of electronic documents. The use of EDS provides three essential properties: authenticity (confirmation of the author of the document), integrity (guarantee that data has not been altered after signing), and non-repudiation (inability of the signer to deny the act of signing).

In practical implementation, EDS creation and validation must follow standardized procedures to guarantee security and interoperability. According to ETSI EN 319 102-1 [2], the signature generation process includes cryptographic algorithms, secure signature creation devices, and proper certificate management. The standard also suggests some procedures for the inclusion of timestamping and verification data, ensuring long-term validity and compliance with legal requirements. By following these standards, organizations can achieve the creation and validation of reliable signatures that are technically sound and legally recognized.

Verification includes a complex process of validating the cryptographic signature components, checking the correctness and validity of the digital certificate, and also confirming the status of the certification authority that gives it. This process confirms that only signatures created with a valid private key corresponding to an authentic certificate are accepted. According to the NIST Digital Signature Standard

(DSS) [3], verification mechanisms must comply with recognized cryptographic algorithms and protocols, ensuring interoperability and international compatibility.

Another important thing is the visualization of EDS verification results. Visualization shows in a form that is easily interpretable by users, such as some color indicators, verification status messages, or graphical icons showing whether the signature is valid, expired, or revoked, etc. This procedure increases the transparency of the validation process and allows both technical specialists and ordinary users to understand the trust level of a signed document. It also simplifies the verification of large volumes of documents, as traditional validation reports can be overwhelming and easy to misread, while a visualized system allows quick identification of invalid signatures or anomalies.

The implementation of an electronic digital signature verification and visualization system ensures the authenticity and legal validity of documents while primarily guaranteeing integrity. By generating a special cryptographic hash key for every document and connecting it to the signer's private key, the system assures that any unauthorized modification, including even a single-byte modification, can be immediately detected and recognized. Ongoing verification of document hashes against reference signatures provides a reliable way to detect tampering and maintain trust in digital workflows across governmental, corporate, commercial, and other sectors.

Beyond signature validation, the system integrates into a broader cybersecurity ecosystem, enhancing trust in electronic communication and supporting a secure, resilient national digital infrastructure.

References:

1. Law of Ukraine “On Electronic Trust Services”. – [URL]: <https://zakon.rada.gov.ua/laws/show/2155-19#Text>
2. ETSI EN 319 102-1 V1.2.1 (2016-04): Electronic Signatures and Infrastructures (ESI); Procedures for Creation and Validation of AdES Digital Signatures.
3. NIST Special Publication 800-53 Rev. 5 — Security and Privacy Controls for Information Systems and Organizations. U.S. Department of Commerce, National Institute of Standards and Technology, 2020.

Ilya Tauyrskii

Postgraduate student, master's degree

Tetiana Zatonatska

Doctor of Economics, Professor,

Taras Shevchenko National University of Kyiv (Ukraine)

MODELING THE DIGITAL INFRASTRUCTURE OF PUBLIC PROCUREMENT AS A TOOL FOR ENHANCING TRANSPARENCY AND EFFICIENCY OF PUBLIC FINANCE MANAGEMENT

Abstract. The paper explores the conceptual basis of modeling digital infrastructure in public procurement as a tool for improving transparency, accountability, and efficiency in public finance management. It highlights the role of blockchain, artificial intelligence, and digital twins in building a new e-governance architecture that integrates open data standards, automated monitoring, and predictive analytics to ensure integrity and effectiveness in public spending.

Keywords: public procurement, digital infrastructure, blockchain, artificial intelligence, open data, transparency, public finance, modeling.

The system of public procurement plays a crucial role in the structure of public finance management, as it defines the efficiency of resource allocation, the quality of public services, and the level of trust in government institutions [1]. In the era of digital transformation, it becomes imperative to create a unified and intelligent infrastructure capable of integrating all procurement stages—from planning and contracting to control and evaluation—within a single analytical environment. Digital modeling allows not only the visualization of existing processes but also their optimization according to principles of transparency, accountability, and efficiency [2].

Ukraine's successful experience with the ProZorro platform has demonstrated the potential of open-source governance and the principle "everyone sees everything." [3]. This system has become a reference model for other countries of Eastern Europe and the Global South, illustrating how open data and civic monitoring can prevent misuse of public funds. The analytical evolution of ProZorro shows that transparency tools, when combined with participatory oversight and digital traceability, create a self-regulating environment that minimizes corruption incentives and improves supplier competition. However, with the growing complexity of procurement data and the diversification of participants, there is a need to develop a new generation of analytical infrastructure that can support real-time monitoring, risk assessment, and strategic forecasting. The concept of a digital twin of public procurement is proposed as a dynamic simulation model that mirrors the current state of contracts, budget flows, and risks, allowing public authorities and civil society to make data-driven decisions [4].

Such a model integrates several layers of interaction. The data layer ensures the reliability and immutability of transactions through distributed ledger technology (DLT) [2]. The analytical layer employs machine learning algorithms for anomaly detection, price forecasting, and risk scoring. The governance layer operates through

smart contracts that automatically verify compliance and control the execution of procurement agreements. Integration with national and international databases, including e-finance and audit systems, enhances interoperability and supports transparent decision-making [5].

The implementation of this approach transforms public procurement from a procedural mechanism into a strategic tool for fiscal optimization. Blockchain ensures the immutability and traceability of every transaction, while artificial intelligence supports predictive monitoring and decision assistance [4]. Transparency becomes an intrinsic feature of the technological system rather than a declared policy goal. As a result, corruption risks are reduced, administrative costs are lowered, and trust in public finance management is strengthened [1].

The implementation of such digital modeling also has significant socio-economic implications. By linking data analytics to policy evaluation, governments can monitor not only compliance but also the real economic impact of public procurement - for instance, on local employment, SME participation, and regional development. Integrating sustainability indicators into procurement analytics supports the EU Green Deal objectives and enables data-driven reporting on social value creation. In this way, the proposed model connects fiscal transparency with broader goals of sustainable governance and public accountability.

The adoption of global open-data standards, such as the Open Contracting Data Standard (OCDS), will enable compatibility with European and OECD systems and facilitate the integration of Ukraine into the global open government framework. This digital infrastructure fosters evidence-based policy, analytical accountability, and participatory control, forming a foundation for smart governance in the public finance domain [2].

In conclusion, modeling the digital infrastructure of public procurement is a key direction for reforming the fiscal system in the context of Ukraine's digital transformation. It not only enhances transparency but also establishes a data-centric culture of governance, where technological innovation becomes a catalyst for institutional trust, efficiency, and sustainable development. A well-designed digital ecosystem of procurement serves as a fundamental mechanism of modern statehood - open, intelligent, and resilient.

References

1. OECD (2023). *Digital Government Index 2023: Shaping Responsive and Resilient Public Governance*. Paris: OECD Publishing.
2. World Bank (2022). *Transforming Public Procurement through Data and Technology*. Washington, DC: World Bank Group.
3. UNDP Ukraine (2022). *ProZorro: Building Transparency and Integrity in Public Procurement*. Kyiv.
4. European Commission (2023). *Open Contracting Data Standard (OCDS): Implementation Guide for EU Member States*. Brussels.
5. Zatonatska T., Tauyrskaa I. (2024). *Digital Twins and AI-Based Modeling in Public Sector Innovation: Case of Ukraine*. *Economics & Society*, No. 49.

SECTION 4: NETWORK AND INTERNET TECHNOLOGIES

Bulava Hennadii

1st year Master's student, Department of Software Systems and Technologies

Kseniia Dukhnovska

Ph.D., Associate Professor, Department of Software Systems and Technologies

Faculty of Information Technology, Taras Shevchenko National University of Kyiv

SOCIAL NETWORK WITH GEOLOCATION SEARCH INTEGRATION: AN ALGORITHMIC APPROACH TO SERVICE RANKING

In the contemporary digital landscape, social networks are deeply integrated into daily life, functioning both as communication channels and as tools for local service discovery and community formation. However, existing platforms demonstrate insufficient efficacy in connecting users with relevant local offerings within a defined geographical context.

Traditional social media, while focused on content, lacks the critical features for identifying and filtering services based on geographic proximity. Conversely, dedicated service directories often neglect the need for personalization and community engagement. This resultant functional fragmentation compels users to rely on multiple systems for common local needs, which significantly diminishes overall efficiency.

The objective of this work is to enhance the efficiency of local service discovery and user social interaction by developing a unified platform with integrated geolocation functionality. This platform will ensure the seamless merging of social interaction and service search, offer real-time proximity-based filtering, and incorporate community feedback mechanisms for service rating and validation. Ultimately, it provides a comprehensive solution that converges social communication with practical interaction within the physical world.

The web application is built on a multi-tier architecture that ensures clear separation of concerns and facilitates scalability. The structure comprises the traditional Presentation, Business Logic, Data Access, and Domain Layers.

The client-side employs HTML, CSS, and JavaScript for a responsive and interactive interface. The server-side is implemented in Java utilizing the Spring Boot framework for robust enterprise-grade support. Database interaction is streamlined via Spring Data MongoDB.

MongoDB serves as the primary database, offering flexible data modeling. A key advantage is its native support for geospatial indexing (via 2dsphere indexes), which is critical for efficient coordinate-based queries. GridFS is utilized for storing large binary objects, such as user images.

The central technical challenge in developing a geolocation-aware social platform lies in creating a fair, efficient, and user-centric service ranking algorithm. This algorithm must simultaneously account for multiple heterogeneous factors: geographic proximity (which varies based on user location), service quality (expressed through user ratings and reviews), service availability (online versus offline), and user preferences (which may prioritize either convenience or quality). Moreover, the algorithm must ensure that neither online nor offline services dominate

the search results unfairly, as this would diminish the value of the platform for certain user segments and service providers [1].

Traditional distance-based ranking prioritizes proximity above all other factors, potentially surfacing low-quality services simply because they are nearby. Quality-based ranking sorts services solely by ratings, potentially recommending highly-rated services that are impractically far from the user's location. Furthermore, most existing systems do not adequately handle the coexistence of online and offline services, often treating them as entirely separate categories.

The algorithm developed in this work addresses these limitations through a principled approach based on multi-criteria decision making (MCDM) theory. MCDM provides a framework for combining multiple, often conflicting, objectives into a unified decision metric. In the context of service ranking, the objectives are maximizing service quality (as perceived by previous users) and minimizing distance (for offline services). The challenge lies in normalizing these disparate metrics—ratings on a 1-5 scale and distances measured in kilometers—into comparable values, determining appropriate weights that reflect user priorities, and integrating online services (which lack a distance component) into the same ranking framework.

Each service in the system is represented by a tuple of attributes: $s_i = (t_i, d_i, r_i)$, where $t_i \in \{0, 1\}$ indicates the service type (0 for online, 1 for offline), d_i represents the distance from the user to the service location (defined only for offline services), and $r_i \in [1, 5]$ represents the average user rating (which may be absent for new services). To create a unified scoring metric, normalization is essential to bring all values into a common scale of $[0, 1]$.

For the spatial component, distance normalization uses the formula $D_i = 1 - (d_i / d_{\max})$, where $d_{\max}$ is the maximum distance among all services in the current search result set. This transformation ensures that the closest service receives a spatial score of $D_i = 1$ (when $d_i = 0$), while the farthest service receives $D_i = 0$. Services at intermediate distances receive proportionally scaled scores. This linear normalization is intuitive and computationally efficient, though alternative approaches such as logarithmic scaling or inverse distance weighting could be employed for different distance perception models.

Rating normalization follows a similar principle but must account for the bounded nature of rating scales and the presence of unrated services. For services with established ratings, the normalized rating is computed as $R_i = (r_i - 1) / 4$, which maps the $[1, 5]$ scale to $[0, 1]$. This preserves the relative differences between ratings while ensuring comparability with the normalized distance metric. For new services without ratings, the system assigns $R_i = 0.5$, representing a neutral position that neither penalizes nor artificially promotes unproven services. This approach balances the competing concerns of giving new entrants a fair opportunity while protecting users from potentially low-quality services.

The overall scoring function combines these normalized components through weighted aggregation. For offline services, which possess both spatial and quality dimensions, the score is computed as $S_i = \alpha \cdot D_i + \beta \cdot R_i$, where α and β are weight coefficients satisfying $\alpha + \beta = 1$. These weights encode the relative importance of proximity versus quality in the user's decision-making process. Empirically, values of

$\alpha = 0.4$ and $\beta = 0.6$ have proven effective, reflecting a moderate preference for quality while still giving meaningful consideration to distance. This weighting can be adjusted based on user feedback or made configurable through user preferences in future iterations.

For online services, which lack a geographic component, the scoring function simplifies to $S_i = \gamma \cdot R_i$, where $\gamma = 1.0$ [3]. This formulation ensures that online services are ranked purely by quality and can be directly compared with offline services on the same $[0, 1]$ scale. The choice of $\gamma = 1.0$ means that a perfectly-rated online service ($r_i = 5$, thus $R_i = 1$) receives a score of 1.0, equivalent to a perfectly-rated nearby offline service ($D_i = 1$, $R_i = 1$, yielding $S_i = 0.4 + 0.6 = 1.0$ only when both components are maximal) [2].

An important part of this algorithm is the proportional alternation mechanism, which prevents systematic bias toward either service type in the final ranked list. After computing scores for all services, they are partitioned into two groups: online services O and offline services L . Within each group, services are sorted in descending order of their respective scores. The goal is to merge these sorted lists into a single output list that fairly represents both categories.

The alternation ratio is determined by the relative sizes of the two groups: $r = |O| / |L|$, where $|O|$ and $|L|$ denote the cardinalities of the online and offline service sets respectively [4]. When $r \approx 1$ (i.e., roughly equal numbers of each type), the algorithm interleaves services in a 1:1 pattern: one offline service, one online service, and so forth. When $r > 1$ (more offline than online), the pattern adjusts to approximately $[r]:1$, inserting $[r]$ offline services for each online service. Conversely, when $r < 1$ (more online services), the ratio inverts accordingly.

This mechanism is grounded in algorithmic fairness principles, which have gained prominence in recent years as researchers and practitioners recognize the potential for ranking systems to inadvertently discriminate against certain groups or categories. In this context, fairness means ensuring that both online and offline service providers have equitable opportunities for visibility, rather than allowing one type to monopolize top positions simply due to greater numerical representation in the database. The proportional approach preserves the relative quality ordering within each category (so better services still rank higher than worse ones of the same type) while preventing category-level disparities.

The implementation of proportional alternation is computationally efficient, requiring only a single pass through the two sorted lists with a simple counter mechanism to track the alternation pattern. The time complexity is $O(n \log n)$ dominated by the initial sorting step, where $n = |O| + |L|$. The subsequent merging operation is $O(n)$, making it negligible in comparison.

Efficient implementation of the ranking algorithm depends critically on the underlying geospatial query capabilities. MongoDB's 2dsphere indexes leverage spherical geometry to accurately model the Earth's surface, accounting for its curvature when calculating distances. This is essential for maintaining accuracy over larger geographic regions where flat-plane approximations introduce significant errors.

The system's geospatial queries typically follow a two-stage process. First, a coarse-grained filtering step uses MongoDB's geospatial operators (such as \$near or

\$geoWithin) to retrieve services within a maximum search radius, leveraging the 2dsphere index for efficient spatial pruning. This reduces the candidate set from potentially millions of services to a manageable subset that are geographically relevant. Second, the refined scoring and ranking algorithm is applied to this candidate set, computing normalized distances, combining them with quality metrics, and producing the final sorted output.

The platform supports two distinct user roles. Regular users can create profiles, establish social connections, join interest-based groups, and discover services filtered by distance, category, rating, and price. Service providers can create detailed service listings specifying offerings, pricing, availability, and location (online or physical coordinates), receive reviews, and build reputation through ratings.

The user interface emphasizes clarity with a persistent search bar enabling quick access to discovery functionality. Search results are organized into columns displaying users, groups, and service providers, with the provider column implementing the sophisticated ranking algorithm. Interactive map views support zooming and panning, providing an intuitive geographic perspective on available services.

The developed algorithm successfully addresses the core challenge of fairly and effectively ranking heterogeneous services in a geosocial context. By combining normalized spatial and quality metrics through weighted aggregation and implementing a proportional alternation mechanism, the system ensures that both online and offline services receive equitable representation while respecting user preferences for proximity and quality. The theoretical foundation in multi-criteria decision making and algorithmic fairness provides confidence in the approach's generalizability and adaptability to different domains and user populations.

The implementation demonstrates the feasibility of integrating social networking and geolocation search within a unified platform architecture. MongoDB's geospatial capabilities, combined with Spring Boot's robust application framework and modern web technologies, create a scalable, maintainable, and user-friendly system.

Future work could explore personalized weighting schemes allowing users to adjust parameters based on preferences, incorporation of temporal factors and social signals, and generalization of the proportional alternation mechanism to handle multiple service categories. The results of this work provide a solid foundation for the development of location-aware social platforms that serve the dual purposes of community building and practical service discovery.

References

1. MongoDB. Geospatial Queries in MongoDB [Electronic resource]. – Access mode: <https://docs.mongodb.com/manual/geospatial-queries/>
2. Aggarwal C. C. Recommender Systems: The Textbook. – Springer, 2016. – 499 p.
3. Zhang S., Yao L., Sun A., Tay Y. Deep Learning Based Recommender System: A Survey and New Perspectives // ACM Computing Surveys. – 2019. – Vol. 52, № 1. – P. 1–38.
4. Samet H. Foundations of Multidimensional and Metric Data Structures. – San Francisco: Morgan Kaufmann, 2006. – 1024 p.

Danylo Chumachenko

PhD Student, Software Engineering Department

Vira Liubchenko

Dr. Eng., Professor, Software Engineering Department

Odessa Polytechnic National University, 1 Shevchenka av., Odesa, 65044, Ukraine

A SEVEN-STEP METHOD FOR SELECTING ARCHITECTURAL PATTERNS FOR IOT

Abstract. Selecting suitable architectural patterns in the Internet of Things is challenging due to the heterogeneity of the Edge, Fog, and Cloud layers. This paper proposes MCAL (Modified Completeness Audit Loop), a structured seven-step method that introduces the Layers×Issues matrix, unified weighting and normalization rules, and a single aggregation pipeline. Compared to the baseline CAL approach, MCAL reduces double-counting, improves traceability, and produces more reproducible rankings.

Keywords: Internet of Things; architectural pattern; Edge–Fog–Cloud; Layers×Issues matrix; Modified Completeness Audit Loop.

Introduction. Designing IoT systems requires choosing architectural patterns that balance performance, security, scalability, and maintainability across the Edge–Fog–Cloud layers [1]. However, due to the heterogeneity of these layers and the diversity of non-functional requirements, ad hoc selection often leads to inconsistent and irreproducible results.

A systematic approach to pattern selection, known as the Completeness Audit Loop (CAL) [2], provides a foundation but has several limitations:

- it does not explicitly represent per-layer issues, creating risks of double-counting;
- its normalization of criteria is inconsistent, complicating aggregation;
- its fragmented aggregation steps reduce traceability and reproducibility.

To address these shortcomings, we propose the Modified Completeness Audit Loop (MCAL), a unified seven-step method that integrates the Layers×Issues matrix with consistent weighting and a single aggregation-and-ranking pipeline. The goal of MCAL is to improve the transparency, reproducibility, and robustness of architectural pattern selection in IoT contexts.

A seven-step procedure for selecting architectural patterns. MCAL is presented as a single, self-contained procedure. The modifications to the baseline approach are integrated directly into the following steps.

1. Context and goals. Define the target domain, identify key stakeholders, describe operational scenarios, and specify constraints such as latency, energy consumption, and connectivity requirements. Establish the priorities for non-functional requirements and then record measurable objectives that the chosen architecture must satisfy.

2. Quality attributes and evaluation criteria. Select a set of quality attributes and evaluation criteria, ensuring that each uses clear and consistent scales (ordinal or numeric). Apply unified scale semantics so that all criteria can be aggregated. Clearly

indicate which criteria are benefit criteria (larger values are better) and which are cost criteria (smaller values are better).

3. Construct the Layers×Issues matrix. Organize the matrix with rows representing the layers (Edge, Fog, Cloud) and columns representing Issues, defined as evaluable items such as data locality, robustness under intermittent connectivity, and security requirements. The matrix links criteria and evidence to issues at each layer, prevents double-counting, and makes trade-offs explicit. Edge computing is adopted in the standard sense: computation performed near data sources to meet latency and bandwidth constraints [3].

4. Candidate patterns. Gather a set of architectural patterns relevant to the context. Apply suitability rules and conduct preliminary screening using the Layers×Issues matrix. Patterns that fail to address key issues are excluded early, which helps narrow the search space.

5. Weighting and normalization. Assign non-negative weights to the criteria to reflect stakeholder priorities. Normalize raw scores to a common 0–1 scale using monotonic transformations consistent with the direction of each criterion (benefit or cost). Rules and examples of normalization are given in [2].

6. Aggregation and ranking. Compute aggregate scores for each pattern through weighted aggregation, consistent with [2]. Ensure that each issue in the Layers×Issues matrix is counted exactly once (or explicitly weighted according to layer importance). Produce an ordered ranking of the candidate patterns and present the top-k results, supported by justifications derived from the matrix and criterion definitions.

7. Validation and selection. Conduct a sensitivity analysis on weights and key assumptions to test the robustness of the top-k candidates. Validate the results against the goals defined in Step 1. The ranking can also be used to analyze trade-offs among quality attributes and to identify potential architectural risks [5]. As explicit outputs, MCAL records context-specific claims that justify the final selection.

Case Study: Applying MCAL. This subsection illustrates the application of the MCAL method to the Smart Home domain, using the balanced weights and the pattern catalog from [2] (Smart Home section). Three architectural layers are considered: edge layer (local events, constrained resources, and the need for immediate response), fog layer (device aggregation, buffering, and execution of automation rules), and cloud layer (long-term storage, integrations, and analytics).

1. Criteria and weights. The balanced profile from [2] is adopted: Security 0.18, Responsiveness 0.18, Interoperability 0.18, Reliability 0.16, Scalability 0.12, Resource Efficiency 0.10, and Adaptability 0.08.

2. Layers × Issues matrix. Criterion weights are allocated across Edge, Fog, and Cloud to reflect where attributes are realized and to avoid double-counting (Table 1).

3. Initial pattern pool. Based on [2] and their applicability to Smart Home, the candidate set includes Publish–Subscribe, Token-Based Authentication, RBAC, Input Validation & Sanitization, Local Caching, Edge Analytics, Circuit Breaker, Cloud-on-the-Loop, and Secure OTA Update.

Criterion	Edge	Fog	Cloud
Security	0.30	0.30	0.40
Responsiveness	0.55	0.30	0.15
Interoperability	0.25	0.35	0.40
Reliability	0.35	0.35	0.30
Scalability	0.15	0.30	0.55
Resource Efficiency	0.60	0.25	0.15
Adaptability	0.25	0.30	0.45

Table 1. Layers $\times$ Issues matrix

4. Context linkage. The rationale for selecting candidate patterns in the Smart Home context is as follows:

- Publish–Subscribe connects Edge, Fog, and Cloud; QoS and retained messages provide reliable delivery in sensor–rule–action scenarios.
- Token-Based Authentication provides stateless device/app authentication (short TTLs, token rotation), scalable for large numbers of devices and integrations.
- RBAC supports role-based access control for owners, family members, guests, and admins; supports break-glass in emergencies; enforces least privilege.
- Local Caching ensures hub autonomy during Internet outages; improves latency for local automation; reduces external traffic.
- Input Validation sanitizes and verifies inbound data at the home boundary, reducing the risk of injections and parsing errors at the gateway or in the cloud.

5. Ranking results. MCAL aggregation highlights the importance of Edge and Fog contributions. Emphasis on responsiveness and resource efficiency elevates local mechanisms:

- Publish–Subscribe – event backbone with controlled delivery.
- Token-Based Authentication – scalable authentication for devices/apps.
- RBAC – clear role management and privilege segregation.
- Local Caching – hub autonomy and reliable rule execution.
- Input Validation – ingress-level protection.

6. Aggregation and ranking.

– MCAL ranking: Publish–Subscribe, Token-Based Authentication, RBAC, Local Caching, Input Validation.

– CAL baseline (without Layers $\times$ Issues matrix): Publish–Subscribe, Token-Based Authentication, RBAC, Input Validation, Local Caching.

Let us note that CAL assigns relatively more weight to security at ingress boundaries, raising Input Validation above Local Caching. Instead, MCAL explicitly accounts for edge-level responsiveness and efficiency, lifting Local Caching higher in the ranking. Access-control patterns (Token-Based Authentication and RBAC) remain consistently prominent in both methods.

7. Validation and selection on the Smart Home Case Study. The configuration recommended for the Smart Home scenario is documented together with rollout and robustness checks:

- Recommended configuration. Publish–Subscribe as the event backbone; Token-Based Authentication with RBAC for access and permissions; Local Caching at the hub; Input Validation at ingress points.

- Supplementary mechanisms. Edge Analytics for local automation, Secure OTA Updates with staged rollouts, Circuit Breakers around external APIs, and Cloud-on-the-Loop for cloud-driven scenarios.

- Sensitivity analysis. Varying Responsiveness or Interoperability weights by $\pm 5\%$ leaves the top three positions unchanged. Positions four and five may swap between Local Caching and Input Validation.

- Traceability. A mapping table (“Pattern – affected quality attributes and layers”), together with the weight vector and the Layers $\times$ Issues matrix, is archived to ensure reproducibility.

Conclusion. This work presents MCAL, a modified seven-step method for selecting IoT architectural patterns. By introducing the Layers $\times$ Issues matrix and unifying the weighting, normalization, and aggregation pipeline, MCAL improves traceability, consistency, and reproducibility compared to the baseline CAL approach.

Application to Smart Home and Industry 4.0 contexts demonstrated that MCAL produces stable rankings under sensitivity analysis and supports justifiable architectural decisions.

References:

1. ISO/IEC 30141:2024, Internet of Things (IoT) — Reference Architecture, ISO/IEC JTC 1/SC 41, 2024.
2. D. K. Chumachenko, V. V. Liubchenko, A systematic approach to selecting architectural patterns for IoT development, *Applied Aspects of Information Technology* 8(2) (2025): 178–190. doi:10.15276/aait.08.2025.12.
3. W. Shi, J. Cao, Q. Zhang, Y. Li, L. Xu, Edge Computing: Vision and Challenges, *IEEE Internet of Things Journal* 3(5) (2016): 637–646. doi:10.1109/JIOT.2016.2579198.
4. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications, *IEEE Communications Surveys & Tutorials* 17(4) (2015): 2347–2376. doi:10.1109/COMST.2015.2444095.
5. S. Moaven, J. Habibi, A fuzzy-AHP-based approach to select software architecture based on quality attributes (FASSA), *Knowledge and Information Systems* 62(12) (2020): 4569–4597. doi:10.1007/s10115-020-01496-7.

Yaroslav Hladkyi

Postgraduate, Department of Information Systems and Technologies

Volodymir Druzhynin

ScD (Eng.), Head of Department of Information Systems and Technologies

Myroslava Gladka

PhD (Eng.), Associate Professor, Department of Information Systems and Technologies

Taras Shevchenko National University of Kyiv

IMMERSIVE VR/AR-ENVIRONMENTS FOR NEUROREHABILITATION AND MOTOR FUNCTION RECOVERY

Abstract. This paper examines approaches to the application of immersive VR/AR environments in neurorehabilitation aimed at motor functions recovery in patients after neurological injuries. Mathematical models of recovery dynamics (exponential and logistic) are presented, along with principles of adaptive regulation of rehabilitation task complexity and methods for therapy personalization based on biomonitring and machine learning. The advantages of VR/AR solutions are demonstrated in promoting neuroplasticity, increasing motivation, and enhancing rehabilitation efficiency are demonstrated. Illustration of functional recovery models and task complexity adaptation are provided.

Keywords: neurorehabilitation, virtual reality, augmented reality, motor function recovery.

Immersive VR/AR environments provide a controlled, safe, and repeatable space for intensive training therapy that promotes neuroplasticity and facilitates motor function recovery after stroke, spinal cord injury, or neurological disorders [1, 2].

The combination of exercises with controlled feedback (haptic/visual/audio) and individualized complexity adaptation enhances recovery efficiency compared to traditional therapy methods [3].

Motivational game-based content, scenarios reinforcing self-efficacy and explicit progress metrics increase patient engagement and adherence to training [4].

The technical components of the system include:

- tracking of limb position and orientation (IMU, optical systems);
- real-time motion processing and pose reconstruction (Kalman filters or particle smoothers);
- adaptative difficulty modules (Rule-based, adaptive staircase, RL).

Let us present the main mathematical models of motor recovery that enables visualization of the environment's operation.

The exponential model of convergence toward the maximum can be expressed as:

$$F(t) = F_{\infty} (1 - e^{-kt}) \quad (1)$$

here $F(t)$ — the functional level over time (or rehabilitation sessions), F_{∞} — the maximum potential recovery, k — the recovery rate constant.

The logistic (sigmoid) model is applied in cases characterized by phase-based recovery dynamics:

$$F(t) = \frac{F_{\max}}{1 + e^{-k(t-t_0)}} \quad (2)$$

here t_0 — the time point corresponding to the accelerated growth phase of the function.

The algorithm for adapting the training difficulty according to the patient's condition depends on the rehabilitation stage and training phase. Let the difficulty level D increase if the performance P_y of the last n attempts $\geq P_{up}$ and decrease if $\leq P_{down}$. Formally:

$$D_{t+1} = D_t + \begin{cases} +1, & \text{якщо } \bar{P}_{t-n+1:t} \geq P_{up}, \\ -1, & \text{якщо } \bar{P}_{t-n+1:t} \leq P_{down}, \\ 0, & \text{інакше.} \end{cases} \quad (3)$$

Effectiveness and exercise performance monitoring metrics are provided by the following indicators [5]:

- kinematic metrics: speed, trajectory accuracy, smoothness (jerk), and reaction time;
- functional metrics: Fugl–Meyer Assessment (FMA), Timed Up and Go (TUG), Barthel Index;
- biomarkers and neurofeedback: EEG/EMG signals iced for VR/AR-based interventions.

The primary approaches to model training and personalization include:

- regression/time-series models for predicting individual progress $F(t)$;
- RL agents for optimal exercise selection policies, considering long-term rewards (maximize cumulative functional gain).

To demonstrate the functional capabilities of the environment, we highlight from the software implementation the modules responsible for generating the logistic recovery curve of motor function (a) (Fig. 1) and the adaptive exercise difficulty adjustment according to the patient's performance (b) (Fig. 2).

a) Motor recovery - logistic model

```
t = np.linspace(0, 30, 301) # therapy sessions
```

```
F_max = 100
```

```
k = 0.25
```

```
t0 = 8
```

```
F = F_max / (1 + np.exp(-k*(t - t0))) # logistic/sigmoid recovery
```

```
plt.figure(figsize=(8,4.5))
```

```
plt.plot(t, F)
```

```
plt.xlabel('Sessions (t)')
```

```
plt.ylabel('Motor Function Level F(t)')
```

```
plt.title('Motor Function Recovery Model - Logistic Curve')
```

```
plt.grid(True)
```

```
plt.tight_layout()
```

```
file1 = out_dir / 'motor_recovery_logistic.png'
```

```
plt.savefig(file1, dpi=150)
```

```
plt.close()
```

b) Adaptive task difficulty vs performance - staircase/adaptive algorithm (example)

```

sessions = np.arange(1,31)
difficulty = np.minimum(1 + np.floor(sessions/3), 10) # difficulty steps
increase
# simulate performance with noise, improving with sessions but dropping when
difficulty jumps
np.random.seed(1)
base_perf = 40 + 1.8*sessions # baseline improvement
perf = base_perf - 3*(difficulty - np.minimum(1, difficulty)) +
np.random.normal(0,4,size=sessions.shape)
perf = np.clip(perf, 0, 100)
plt.figure(figsize=(8,4.5))
plt.plot(sessions, perf, marker='o', linestyle='-')
plt.xlabel('Session')
plt.ylabel('Percentage of Task Completion (%)')
plt.title('Adaptive Task Difficulty vs. Patient Performance')
plt.grid(True)
plt.tight_layout()
file2 = out_dir / 'adaptive_difficulty_performance.png'
plt.savefig(file2, dpi=150)
plt.close()

```

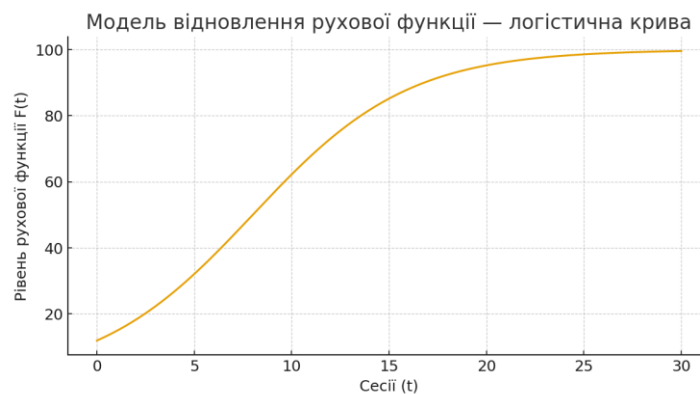


Figure 1 – Motor function recovery model – logistic curve



Figure 2 – Adaptive exercise difficulty vs. Patient performance

Immersive VR/AR environments represent a promising direction in the development of modern neurorehabilitation, integrating the principles of neuroplasticity, gamification, and personalized load adaptation. The usage of virtual simulators with biofeedback enables increased patient efficiency and motivation, objective progress monitoring, and optimization of the recovery program. Mathematical models of functional recovery dynamics and adaptive complexity control algorithms form the foundation for developing intelligent systems to support therapeutic decision-making. Future research should focus on integrating artificial intelligence technologies, biosensory systems, and telerehabilitation platforms into a unified digital medicine ecosystem.

References:

1. Wen, Dong; Fan, Yali; Hsu, Sheng-Hsiou; Xu, Jian; Zhou, Yanhong; Tao, Jianxin; Lan, Xifa; Li, Fengnian (2021). Combining brain–computer interface and virtual reality for rehabilitation in neurological diseases: A narrative review. *Annals of Physical and Rehabilitation Medicine*. T. 64, № 1. c. 101404. doi:10.1016/j.rehab.2020.03.015. ISSN 1877-0657
2. The use of virtual reality in neurorehabilitation, [online] Available at: <https://neurooptima.in.ua/vikoristannya-virtualnoyi-realnosti-u-nejroreabilitaciyi> (Accessed: 1 November 2025).
3. Gupta, Ankur; Vardalakis, Nikolaos; Wagner, Fabien B. (2023). Neuroprosthetics: from sensorimotor to cognitive disorders. *Communications Biology* Vol.6, No.1. pp. 1—17. doi:10.1038/s42003-022-04390-w
4. Modern technologies in neurorehabilitation: a teaching aid for practical classes for applicants of the 1st (bachelor's) level of training in speciality 227 "Physical therapy, occupational therapy" / V. I. Dariy, M. V. Sikorska, I. V. Vizir, A. O. Dronova. – Zaporizhzhia: ZDMFU, 2025. – 92 p.
5. Shepel, A. I., & Goroshko, V. I. (2023). The use of innovative methods of virtual reality in physical therapy of patients with musculoskeletal injuries. *Rehabilitation and Recreation*, (17), 150–158. <https://doi.org/10.32782/2522-1795.2023.17.18>
6. VRNOW (2025) VRNOW [online]. Available at: <https://www.adv.ua/vr-now/> (Accessed: 1 November 2025).
7. Medical aspects of the use of virtual reality in the rehabilitation of patients (2025) [online]. Available at: <https://mister-blister.com/> (Accessed: 1 November 2025).

Kyrylo Karmazin

Postgraduate student

Julia Pisarenko

Candidate of technical sciences

Viktor Glushkov Institute of Cybernetics of the NAS of Ukraine

OVERVIEW OF INTELLIGENT METHODS IN ADAPTIVE VIDEO STREAMING SYSTEMS

Abstract. The paper provides a systematic review of modern intelligent algorithms applied in adaptive video streaming systems to enhance the Quality of Experience (QoE). The relevance of the topic is driven by the rapid growth of video traffic and the need for flexible mechanisms of adaptation to variable network conditions, end-user device limitations, and viewing context. The purpose of this work is to form a systematic review and analysis of methods, algorithms, and architectures of intelligent adaptive video systems, as well as to highlight the prospects for further research in this area.

Keywords: Adaptive video streaming; Quality of Experience; Adaptive Bitrate; Machine Learning; Reinforcement Learning; hybrid algorithms; network prediction; edge computing; intelligent systems; buffering optimization.

Adaptive Bitrate (ABR) systems segment video streams and select optimal quality levels based on buffer status, bandwidth, and network parameters. Classical algorithms (throughput-, buffer-, and hybrid-based) [1] are simple and effective in stable networks but perform poorly under unpredictable conditions, leading to lower QoE stability [2].

Intelligent approaches such as Machine Learning (ML), Reinforcement Learning (RL), and hybrid models improve adaptability. ML predicts network states using historical data and user behavior, employing GRU- or CNN-RNN-based models [3] for bitrate optimization. Yet, their deployment demands large datasets and careful adaptation to limited resources. RL enables agents to learn bitrate selection policies that balance quality and buffering. Algorithms like Pensieve dynamically coordinate between smooth playback and quality but remain complex, resource-intensive, and often lack generalization. Hybrid models combine ML prediction with RL or classical ABR, achieving better QoE stability and fewer quality switches. However, their design must account for latency, scalability, energy efficiency, and the distribution of computation between client, edge, and server.

Architectural designs of adaptive systems – client, server, and edge-oriented offer trade-offs: client systems are simpler, server ones enable centralized control, while edge systems provide the best adaptability through reduced latency and contextual analytics. Additional challenges include buffering, multi-device operation, and viewing context. Optimizing startup buffering, prefetching, and device-aware quality adjustment improves performance. A combined architecture with client ABR and edge prediction supports synchronization and contextual QoE optimization.

Each method class has strengths and limitations. Current research confirms that hybrid models achieve the best balance of QoE and resource use. Practical deployment, however, still depends on robust data collection, scalable model training,

and seamless integration with client infrastructures. Comparison of video processing and transmission approaches is given in Table 1.

Approach	QoE Stability	Buffering	Complexity
Heuristic ABR	Moderate: works well in stable networks but degrades under high variability.	Simple buffer policy based on buffer length or current state; limited active control.	Low: simple to implement and maintain, minimal training requirements.
Reinforcement Learning (RL)	Higher potential QoE stability; adapts to complex scenarios, improves QoE but may fluctuate under unknown conditions.	Advanced buffering control; RL agents consider buffer state and optimize pre-buffering.	High: requires data collection, training, complex policy design, and risk of overfitting.
Hybrid Models	Best stability potential, combining advantages of classical and intelligent methods for better QoE balance.	Good buffer control & bitrate selection capabilities: hybrid allows for more flexible buffer and bitrate control at the same time.	Medium-high: more complex than classical ABR but easier than full RL; requires more resources and infrastructure.

Table 1. Comparison of Video Processing and Transmission Approaches

In conclusion, intelligent algorithms based on ML, RL, and hybrid approaches significantly improve the stability and efficiency of adaptive video streaming. Their successful implementation requires coordinated data processing and model training across client, server, and edge levels. The results are valuable for video service developers and CDN engineers, enabling better QoE, reduced delays, and optimized use of network and device resources.

References:

1. Babak Taraghi, Abdelhak Bentaleb, Christian Timmerer, Roger Zimmermann, and Hermann Hellwagner. Understanding quality of experience of heuristic-based HTTP adaptive bitrate algorithms // Proceedings of the 31st ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '21). – Association for Computing Machinery, New York, NY, USA. – 2021. – PP. 82-89. [DOI: 10.1145/3458306.3458875].
2. L. Peroni and S. Gorinsky. Quality of Experience in Video Streaming: Status Quo, Pitfalls, and Guidelines // Proceedings of the 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India. – 2024. – PP. 558-567. [DOI: 10.1109/COMSNETS59351.2024.10427330].
3. Darwich, M., Bayoumi, M. Video quality adaptation using CNN and RNN models for cost-effective and scalable video streaming Services // Cluster Computing 27(5):1-21. – 2024. – PP. 6355-6375. [DOI: 10.1007/s10586-024-04315-8].

Yuliia Mavrodi

Master's degree student of the Department of Technologies Management

Tetiana Latysheva

Candidate of Technical Sciences, Assistant of the Department of Management Technologies

Taras Shevchenko National University of Kyiv

PROJECT OF WEB SERVICE DEVELOPMENT FOR MONITORING POWER OUTAGES IN UKRAINE

Abstract. The ongoing energy crisis in Ukraine, driven by continuous infrastructure damage, necessitates rapid and accurate public information about power supply status. Existing solutions are fragmented, often showing only planned schedules instead of actual emergency outages. This research presents the development of a comprehensive, integrated web service designed for real-time monitoring and visualization of power outages. The key scientific novelty lies in the systematic combination of: 1) Automated aggregation of official energy provider API data; 2) Crowdsourced, user-generated outage reports with automated validation; and 3) A dynamic, interactive map displaying the actual energy status and the locations of critical resilience infrastructure ("Resilience Points") with detailed service and loading information. This solution enhances public adaptation to energy constraints by providing personalized, geolocated multi-channel notifications and serves as a vital tool for local authorities in crisis management and resource planning.

Keywords: Power Outage Monitoring, Web Service, Crisis Management, Crowdsourcing, Resilience Points, Interactive Mapping, Information System, Predictive Analytics.

1. Introduction

The reliability of information regarding power outages has become a critical social and infrastructural challenge in Ukraine. The gap between scheduled energy curtailment and unexpected emergency disconnections creates a high degree of uncertainty for millions of citizens and vital services. Current informational tools lack the integration, real-time accuracy, and user-centric features required for effective crisis adaptation. The aim of this project is to bridge this information gap by creating a unified, resilient, and scalable web service [1].

2. System Architecture and Methodology

The proposed web service adopts a robust client-server architecture. The backend is implemented using Python (Flask), chosen for its flexibility and strong support for integrating machine learning libraries necessary for future predictive analysis [2]. The data layer employs MongoDB, a NoSQL document database, which provides the necessary schema flexibility and high performance for managing heterogeneous data streams (structured outage schedules, semi-structured user reports, and geospatial coordinates).

The core functional modules include five main components:

1. **Data Collection and Validation Module:** Responsible for asynchronous aggregation, cleaning, and strict verification of data streams from official provider APIs and the crowdsourcing mechanism, including automated spam filtering and geographic consistency checks [3].

2. **User Management Module:** Handles user lifecycle, including registration/authorization, profile management, and persistent storage of user-defined monitoring addresses.

3. **Notification Module:** Generates personalized, geolocated alerts (Push, SMS, Email) based on the user's distance from the outage and the event's severity, calculated using a dedicated mathematical priority function:

$$Priority(u_i, p_i) = \frac{1}{distance(u_i, p_i)} * Severity(p_i)$$

4. **Interactive Map Module:** Utilizes Leaflet.js and geospatial queries to display real-time, filtered data on planned/emergency outages and the locations of "Resilience Points", including their service and load status.

5. **Analytics and Reporting Module:** Focuses on advanced functions, including predictive modeling for potential future outages and the generation of structured reports for administrator and analytical use [4].

3. Scientific Contribution

The primary contribution of this research is the development of a systematic approach to combine disparate information streams (official vs. crowdsourced) and visualize them in a single, verified environment. Notably, the system moves beyond passive data display by formalizing predictive modeling using historical data (P_{hist}) and external factors (W , e.g., weather, infrastructure attacks) to forecast potential future outages (P_{future}) [5]:

$$P_{future}(r_i) = \alpha * \sum_{t=1}^n P_{hist}(r_i, t) + \beta * W(r_i)$$

This framework allows users to view data based on the actual, current state of the power grid, providing immediate, actionable intelligence that is currently unavailable through official channels alone.

4. Conclusion

The developed web service concept provides a critically needed, integrated information solution for adapting to the energy crisis. Its combination of robust crowdsourcing, API integration, and personalized notification mechanisms represents a significant step forward in crisis information management. Future work will focus

on integrating the predictive model and developing a dedicated mobile application with offline functionality to further enhance accessibility and resilience.

References:

1. Pressman R. S., Maxim B. R. Software Engineering: A Practitioner's Approach / R. S. Pressman, B. R. Maxim. – 9th ed. – New York: McGraw-Hill Education, 2020. – 671 p.
2. Kirichek G. G., Kopytchenko O. V. Implementation of Web System Optimization Method / G. G. Kirichek, O. V. Kopytchenko // CEUR Workshop Proceedings. – 2020. – Vol. 2608. – P. 153–159. – Available at: <https://ceur-ws.org/Vol-2608/paper16.pdf>
3. Shepetukha Y., Koval S. Data integration for critical infrastructure protection / Y. Shepetukha, S. Koval // Proceedings of the 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), 2020. – P. 1000–1004. – Available at: <https://ieeexplore.ieee.org/document/9043785>
4. Artificial Intelligence in the Energy Sector: Analytical Report / National Institute for Strategic Studies. – Kyiv, 2022. – 70 p. – Available at: https://niss.gov.ua/sites/default/files/2022-07/dopovid-ai-v-energetici-red_01-pogodzheno-sukhodolya_02-1.pdf
5. Li X., Zhang Y., Wang Y. Research on renewable energy power demand forecasting method based on IWOA-SA-BILSTM modeling / X. Li, Y. Zhang, Y. Wang // Frontiers in Energy Research. – 2023. – Available at: <https://www.frontiersin.org/articles/10.3389/fenrg.2023.1331076/full>.

Danylo Skydan

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Yurii Shcheblanin

Candidate of technical sciences, senior researcher, associate professor. Department of Cyber Security and Information Protection

Taras Shevchenko National University of Kyiv

ANALYSIS OF PERFORMANCE AND STABILITY OF COMBINED METHODS IN WI-FI PROTOCOLS

Abstract. The mechanisms for enhancing Wi-Fi protocol security using combined encryption schemes are analyzed. A comparative analysis of different approaches to hybrid encryption is carried out in terms of performance and resistance to attacks. Today, it is relevant to implement combined encryption methods, which can significantly improve data protection in wireless networks and reduce the risk of modern attacks, such as dictionary attacks, replay attacks, and man-in-the-middle (MitM) attacks.

Keywords: Wi-Fi security, hybrid encryption, cryptography.

Wireless access has become a key part of modern information systems. Its convenience and mobility have made Wi-Fi the dominant technology for both users and corporate networks. The IEEE 802.11 standards support a wide range of applications, from home networks to complex infrastructures. However, the widespread use of Wi-Fi also brings risks, as attacks on wireless protocols can threaten data confidentiality and integrity [1].

The WPA2 standard, still used by millions of devices, is vulnerable to several attacks, including dictionary attacks on passwords, replay attacks that repeat legitimate traffic, and man-in-the-middle (MitM) attacks that intercept or modify transmitted data. Even strong symmetric encryption like AES-256 cannot fully prevent these threats. WPA3 aims to improve security, but adoption is limited due to device compatibility and deployment complexity [2].

One promising approach is combining symmetric and asymmetric encryption. Using asymmetric cryptography for authentication and key exchange along with symmetric ciphers for data protection improves protocol resistance to attacks. This study focuses on schemes combining Elliptic Curve Cryptography (ECC) with AES-256 and ChaCha20, including a three-layer variant for additional protection.

Methodology and Experiments

To evaluate the performance and security of the combined encryption schemes, controlled tests were conducted in a Wi-Fi (Wireless Fidelity) network environment. Three encryption schemes were analyzed: ECC (Elliptic Curve Cryptography) + AES-256, ECC + ChaCha20, and a three-layer scheme ECC + AES-256 + ChaCha20 [3].

The tests involved transmitting 100 MB of data over TCP (Transmission Control Protocol) connections to measure throughput and latency. CPU (Central

Processing Unit) usage and energy consumption were monitored during continuous encryption to assess how resource-intensive each scheme is. Key generation and secure key exchange were included in the measurements to simulate real-world operation conditions.

Each test was repeated multiple times to calculate average values and ensure the reliability of the results. The methodology allowed for a comparison of the schemes in terms of speed, efficiency, and security, highlighting trade-offs between encryption strength and system performance.

Additionally, network conditions were controlled to maintain consistency in measurements, and all schemes were tested under similar loads to ensure fair comparison. This setup enabled a practical evaluation of how each encryption method behaves in typical wireless network scenarios, helping to determine the most effective approach for protecting data in transit [4].

Conclusions

The study showed that the performance and effectiveness of combined encryption schemes largely depend on the device architecture and the chosen algorithms. ECC + ChaCha20 provides a good balance of speed, energy efficiency, and security. ECC + AES-256 is less efficient on devices without hardware AES acceleration. The three-layer scheme offers the highest security but at the cost of performance. Overall, using ECC improves resistance to attacks and strengthens Wi-Fi security. The choice of scheme should consider the balance between speed, energy use, and protection level [5].

References

1. Bernstein, D. J. (2008). The Salsa20 family of stream ciphers. In *New Stream Cipher Designs* (pp. 84–97). Springer.
2. Langley, A., Chang, W., & Strombergson, J. (2016). ChaCha20 and Poly1305 for IETF Protocols. RFC 8439. IETF.
3. Dworkin, M. J. (2001). Recommendation for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC. NIST SP 800-38D.
4. Gupta, V., & Gupta, S. (2002). Elliptic Curve Cryptography (ECC) and its applications. In *Proc. 6th IEEE International Symposium on Consumer Electronics* (pp. 209–212).
5. Vanhoef, M., & Piessens, F. (2017). Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1313–1328. <https://doi.org/10.1145/3133956.3134027>

¹ Ihor Surovtsev

DSc (Engineering), Senior Researcher, Department Head

² Valentyna Galimova

PhD (Chemistry), Associate Professor

³ Yevheniya Savchenko-Syniakova

PhD (Engineering), Senior Researcher, Leading Researcher

⁴ Volodymyr Liakhov

Research Assistant, PhD student,

⁵ Andriy Khalevych

Research Assistant, PhD student,

⁶ Yaroslav Antoniuk

Senior Researcher

^{1,3-6} *Institute of Information Technologies and Systems of the NAS of Ukraine*

³ *National University of Life and Environmental Sciences of Ukraine*

SYSTEMS FOR RAPID DETERMINATION OF THE ECOLOGICAL STATE OF WATER RESOURCES

Determining the environmental status of water resources during wartime is critical for ensuring public safety, supporting the economy, and preventing environmental disasters. Russia's military aggression has caused significant damage to Ukraine's water infrastructure, leading to the destruction of water supply systems, contamination of rivers and groundwater, and a shortage of high-quality drinking water. This is primarily due to artillery shelling, missile and drone attacks, as well as direct destruction caused by the explosion of the Kakhovka Dam in June 2023. As a result of military action, significant quantities of heavy metals, petroleum products, and other contaminants enter surface and groundwater.

Expanding the capabilities of environmental monitoring of water bodies through the development of new technologies and systems for rapidly determining the concentrations of toxic substances entering water resources is currently a pressing issue [1]. The use of such systems for drinking and domestic water for military units and affected populations will allow for rapid assessment of the environmental status of water in the field, determining the degree of danger, and the risk of disease from consuming contaminated drinking water [2]. Rapid water quality monitoring systems must be portable, self-powered, and capable of reliable real-time water monitoring and meet modern requirements for network and cloud-based environmental data storage.

The use of portable systems for rapid detection of toxic substances in water bodies on-site for water quality and safety assessments solves the problem of field environmental monitoring, which requires efficient data management for collecting, storing, processing, and analyzing environmental data [3]. Network and cloud solutions [4, 5] ensure reliability, scalability, and data availability, enabling the effective integration of field concentration measurement systems into automated environmental data collection and processing systems, reducing the risk of

information loss, increasing productivity, and enabling real-time data analysis. The degree of hazard and risk of disease from consuming contaminated water can be assessed using heavy metal pollution indices (HPIs) [6] or other pollution indices.

FUNDING

This research has carried out under grant support of the National Research Fund of Ukraine within the framework of “Portable software and hardware system for rapid determination of toxic substances in water bodies contaminated as a result of military operations” project (registration number 2023.04/0128) implemented under the direction “Science to Strengthen Defense Capabilities of Ukraine”.

References:

1. E. Bernalte, S. Arevalo, J. Perez-Taborda, J. Wenk, P. Estrela, A. Avila, M. Di Lorenzo. Rapid and on-site simultaneous electrochemical detection of copper, lead and mercury in the Amazon river. *Sens. Actuators, B* 2020, 307, No. 127620. <https://doi.org/10.1016/j.snb.2019.127620>
2. I.V. Surovtsev, P.Y. Velykyi, M. Hrytsaienko, V.M. Galimova. Analytical system for environmental monitoring and risk assessing of drinking water consumption. *Cybernetics and Computer Engineering*, 2021, 4(206), 17-38. <https://doi.org/10.15407/kvt206.04.017>
3. Y. Zhou, H.L. Wang, D. Song, Z.G. Li, S.T. Han, F. Long, A.N. Zhu. Simple, rapid, and sensitive on-site detection of Hg²⁺ in water samples through combining portable evanescent wave optofluidic biosensor and fluorescence resonance energy transfer principle. *Anal. Chim. Acta* 2021, 1155, 338351
4. I.V. Surovtsev, Ye.A. Savchenko-Syniakova, V.M. Galimova, V.S. Liakhov. Approaches to the development of information technology for rapid identification of the environmental objects ecological state. *Cybernetics and Computer Engineering*, 2024, 4(218), 17-38. <https://doi.org/10.15407/kvt218.04.005>
5. Nadica Stojanovic, Dr. Sunita Chaudhary. Real-Time Water Quality Monitoring in Aquaculture using IoT Sensors and Cloud-Based Analytics. Volume 4 Issue 2(2023). P. 174–187.
6. R. Appiah-Opong, A. Ofori, M. Ofosuene, E. Ofori-Attah, F.K.E. Nunoo, I. Tuffour, C. Gordon, D.K. Arhinful, A.K. Nyarko, B.Y. Fosu-Mensah. 2021. Heavy metals concentration and pollution index (HPI) in drinking water along the southwest coast of Ghana. *Appl Water Sci.* 11(3): 1–10. doi: <https://doi.org/10.1007/s13201-021-01386-5>

Orest Vovchak

Ph.D. Student

Zenoviy Veres

Ph.D. Associate professor

Lviv Polytechnic National University (Lviv, Ukraine), Department of Computerized Automatic Systems

COMPARATIVE ANALYSIS OF DATA STORAGE METHODS FOR IOT APPLICATIONS: ADVANTAGES AND LIMITATIONS

Abstract. This paper provides a comparative analysis of existing data storage types applicable to modern Internet of Things (IoT) systems. The study focuses on relational, non-relational, time-series, cloud, distributed file, and blockchain-based storages, evaluating their features, scalability, performance, and data integrity. The results highlight that while traditional relational databases remain mature and reliable for structured data, emerging paradigms such as time-series and blockchain storages provide advantages in latency, traceability, and auditability for IoT telemetry. The study emphasizes the need for hybrid approaches that combine flexibility, cost-efficiency, and verifiable data integrity across environments.

Keywords. IoT, data storage, blockchain, NoSQL, time-series database, cloud storage

In modern digital systems, especially those related to the Internet of Things (IoT), data is generated continuously by sensors, devices, and services. These data streams can include environmental readings, system logs, GPS positions, or transaction records. The volume and velocity of such data are constantly increasing, which makes efficient data storage and management a key challenge. Selecting the right data storage type directly affects the performance, scalability, and reliability of IoT solutions [1]. Traditional relational databases, while mature and consistent, were not designed for the real-time, high-frequency nature of IoT data. This has led to the development and adoption of alternative systems such as NoSQL, time-series databases, and blockchain-based storage [2, 3]. The goal of this paper is to compare and summarize the main types of data storage systems used today, focusing on their suitability for IoT scenarios. The comparison is based on a structured analysis of their features, advantages, disadvantages, and typical use cases.

Relational databases such as MySQL, PostgreSQL, or SQL Server are widely used for structured data and support transactions with strong consistency (ACID). They are mature and well-documented, but they scale poorly with very large or fast-changing IoT data.

Non-relational (NoSQL) databases like MongoDB, Redis, Neo4j, or Cassandra provide flexible structures and high speed. They are good for telemetry and JSON-based data but do not always guarantee transaction safety or strong data protection [4].

Time-series databases (InfluxDB, TimescaleDB, OpenTSDB) are built for sensor and time-based data. They allow fast data writing, low latency, and easy aggregation over time. However, their query functions are limited and often specialized [5].

Cloud storages such as AWS S3, Azure Blob, or Google Cloud Storage are cost-effective and easy to scale. They are suitable for unstructured data but depend on access policies and lack strong internal data integrity tools. Distributed file systems (Hadoop HDFS, Ceph, IPFS) are designed for storing very large datasets and ensuring fault tolerance. IPFS also adds immutability through hashing, which improves data trust. Still, such systems can be complex to manage and slower for real-time access.

Blockchain storages like Ethereum, Hyperledger Fabric, or Corda use decentralization and cryptography to protect data. They provide full immutability, audit trails, and trust between parties [6]. Their main drawbacks are low speed, high resource usage, and integration complexity.

In conclusion, each storage method has its own role in IoT data management. Time-series and NoSQL systems are efficient for real-time data, while blockchain storage ensures trust and traceability. Among all approaches, blockchain stands out as the most secure and verifiable option, offering complete data immutability and transparent auditing. Despite its current performance limitations, its reliability and built-in trust mechanisms make it the most promising foundation for critical IoT infrastructures where data integrity and accountability are essential. Future systems are expected to rely increasingly on blockchain to guarantee trustworthy and tamper-proof IoT operations.

References:

1. Lee, C., Kim, J., Ko, H., & Yoo, B. (2023). Addressing IoT storage constraints: A hybrid architecture for decentralized data storage and centralized management. *Internet Things*, 25, 101014. <https://doi.org/10.1016/j.iot.2023.101014>.
2. Rohal, D., Lwakatare, L., Wu, Y., Haataja, J., Nurminen, J., & Kangasluoma, J. (2025). Comparative Analysis of Three IoT Data Storage System Architectures on AWS Cloud. 2025 IEEE 22nd International Conference on Software Architecture Companion (ICSA-C), 51-59. <https://doi.org/10.1109/icsa-c65153.2025.00016>.
3. R., Zhou, W., & , Z. (2023). An Efficient NoSQL-Based Storage Schema for Large-Scale Time Series Data. *J. Database Manag.*, 35, 1-21. <https://doi.org/10.4018/jdm.339915>.
4. Maamari, S., & Nasar, M. (2025). A Comparative Analysis of NoSQL and SQL Databases: Performance, Consistency, and Suitability for Modern Applications with a Focus on IoT. *East Journal of Computer Science*. <https://doi.org/10.63496/ejcs.vol1.iss2.76>.
5. Hao, Y., Qin, X., Chen, Y., Li, Y., Sun, X., Tao, Y., Zhang, X., & Du, X. (2021). TS-Benchmark: A Benchmark for Time Series Databases. 2021 IEEE 37th International Conference on Data Engineering (ICDE), 588-599. <https://doi.org/10.1109/icde51399.2021.00057>.
6. Heo, J., Ramachandran, G., Dorri, A., & Jurdak, R. (2024). Blockchain Data Storage Optimisations: A Comprehensive Survey. *ACM Computing Surveys*, 56, 1 - 27. <https://doi.org/10.1145/3645104>.

SECTION 5: SOFTWARE ENGINEERING, DATA STRUCTURES, AND COMPUTING

Maksym Demydov

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Yurii Babenko

Assistant at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

Taras Shevchenko National University of Kyiv

HIERARCHICAL ROLE MODEL IN CRM

Abstract. The growing use of CRM systems requires reliable mechanisms for managing user access. This paper examines the hierarchical role model (RBAC) in CRM systems, focusing on its importance for data security and centralized access control. A comparison with OAuth Scopes highlights their distinct advantages and application areas. While OAuth Scopes enable fine-grained, temporary permissions for APIs, the hierarchical role model remains essential for stable, scalable, and structured internal access management.

Keywords: hierarchical role model, CRM, RBAC, OAuth Scopes, access control, data security, authorization system

A hierarchical role model in CRM means that roles within the system have “top-down” relationships: a higher-level role can view the data of lower-level roles, but not vice versa. The key components of this model include: user, role, profile, and permissions & visibility. [1]

Roles determine which records (for example, deals or contacts) a user can see — for instance, a regional manager can view data from subordinates within their region but not from others. The profile defines what actions a user can perform with those records (create, read, update, delete, import, export) across different CRM modules (leads, contacts, deals, etc.). [1]

Permissions and visibility establish the base level of access within an organization — for example, a module may be “Public” (visible to everyone) or “Private” (visible only to the owner and their superiors in the hierarchy) (*Figure 1*). [1]

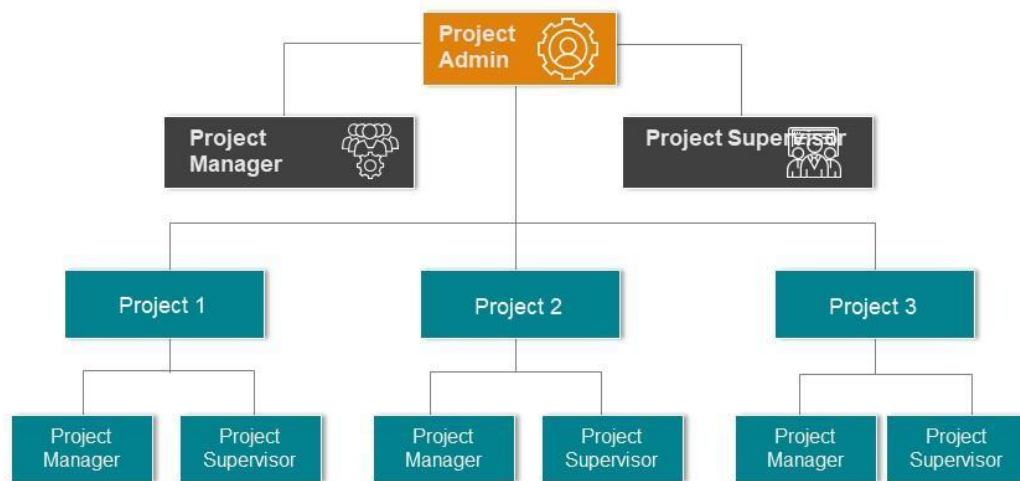
In modern information systems, there are several approaches to managing user access. The most common among them are OAuth Scopes and RBAC (Role-Based Access Control), which differ in principles of operation and areas of application (*Table 1*).

OAuth Scopes and RBAC (Role-Based Access Control) are two methods of managing access, but they serve different purposes and work in distinct ways. [2]

- **OAuth Scopes:** These are permission strings tied to access tokens that define what an application can do with a user’s data. They are temporary, fine-grained, and best suited for external API integrations that require specific, short-term permissions.[2]

- **RBAC:** This method assigns permissions based on user roles, grouping actions according to job functions. It is static, broader in scope, and best suited for internal systems requiring long-term access [2]

Project Role Hierarchy Sample Of PPT



This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Figure 1. Operation scheme of the Hierarchical Role Model

Table 1 – Comparison of Access Management Approaches

Feature	OAuth Scopes	RBAC
Duration	Temporary (token-based)	Static (role-based)
Granularity	Fine-grained (resource-level)	Broader (role-level)
Flexibility	High, dynamically adjustable	Lower, requires manual updates
Use Case	External API integrations	Internal user access

Despite the emergence of modern authorization methods, the hierarchical role model (RBAC) remains relevant because it provides a clear access structure, reflects the organizational hierarchy (for example, “admin → manager → user”), and allows centralized management of permissions for a large number of employees. In systems where access is defined by job responsibilities, RBAC simplifies administration, ensures transparency, and provides stable long-term access management. [3–4] This model is especially effective in large enterprises where user responsibilities are clearly defined and do not change frequently. By assigning permissions based on organizational hierarchy, it minimizes the risk of unauthorized access and makes the

management of employee transitions—such as promotions or departmental transfers—more efficient.

Although OAuth scopes provide high precision and security in access control, their role is mainly limited to managing specific permissions within resources or API integrations. They are effective for temporary, technical use cases but do not replace the need for a stable hierarchical structure of access provided by RBAC. The hierarchical model maintains consistency, centralized control, and logical subordination of rights, which are critical for large organizations and internal systems. In hybrid environments, both models can complement each other: OAuth scopes provide flexibility for external integrations, while RBAC ensures order and accountability within the internal infrastructure. This combination allows companies to balance operational agility with long-term access stability and governance.

Conclusion

The hierarchical role model remains an essential component of modern security systems because it ensures a clear access structure that mirrors real organizational hierarchies. It simplifies user permission management, reduces the risk of access errors, and enhances information control. Thanks to its clarity and scalability, this model continues to serve as an effective foundation for building secure corporate systems.

References

1. Marketing P. C. A complete guide to Roles Hierarchy and Access Controls for sales organisations. *CRM Blog | Pepper Cloud*. Available at: <https://blog.peppercloud.com/roles-hierarchy-and-access-controls-for-sales-organisations/>
2. Using OAuth 2.0 scopes vs. permissions for app authorization. Using OAuth 2.0 scopes vs. permissions for app authorization. Available at: <https://www.aserto.com/blog/scopes-vs-permissions-authorization>
3. Bennett T. OAuth Scopes vs RBAC: Key Differences. Blog. Available at: <https://blog.dreamfactory.com/oauth-scopes-vs-rbac-key-differences>.
4. Mosyan D. 3 API Authorization Mechanisms. Medium. Available at: <https://medium.com/@dmosyan/3-api-authorization-mechanisms-e84ab589d76c>.

Georgy Gaina

PhD in Technical Sciences, Professor, Professor of the Department of Intelligent Technologies

Hrygorii Hnatiienko

PhD in technical sciences, deputy dean of the faculty of information technology

Iryna Domanetska

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Oleh Ilarionov

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Hanna Krasovska

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Taras Shevchenko National University of Kyiv

PROFUNCTOR SEMANTICS IN CATEGORICAL LOGIC FOR THE VERIFICATION OF QUERIES AND DATA INTEGRITY IN RELATIONAL DATABASES

Abstract. The paper proposes a categorical logical model for the formal verification of query correctness and data integrity in relational databases. Unlike traditional logical–algebraic approaches, the model employs the apparatus of categorical logic to represent schemas, constraints, and queries in the form of commutative diagrams. The approach to profunctor semantics is extended to enable the formal description of “one-to-many” and “many-to-many” relationships within a bicategorical framework. A unified formalism is developed for proving query soundness and referential integrity at both the structural and logical levels.

Keywords: categorical logic, relational databases, query verification, data integrity, profunctors, commutative diagrams, formal methods.

The complexity of relational database schemas and their associated queries poses significant challenges for ensuring the correctness of data processing and the preservation of data integrity. Traditional methods—such as relational algebra and logical inference—often fail to provide a formal basis for query verification in the presence of complex inter-table relationships, leading to potential errors or violations of referential integrity.

A new approach to the formal verification of query correctness and data integrity based on categorical logic is proposed. In this framework, database schemas, constraints, and queries are represented through commutative diagrams, providing a unified and rigorous formalism. The object of the research is the processes of data manipulation within relational databases, while the subject of the study is the categorical-logical methods for their modeling and verification.

This study presents a novel categorical logical model that, for the first time, integrates schemas, queries, and constraints into a unified framework for the formal verification of query correctness. The proposed approach introduces a profunctor-based semantics to formally represent non-uniform “one-to-many” and “many-to-

many” relationships, thereby generalizing the traditional JOIN operation in relational databases. Furthermore, a formal criterion for preserving referential integrity through the verification of diagram commutativity is developed, which enables the automation of the verification process. The overarching objective of the research is to construct a unified categorical model that supports the formal verification of both query correctness and data integrity. The model contributes to improving the reliability of information systems and opens new prospects for the automation of verification procedures in complex database schemas. Its practical implementation can be realized using tools such as Catlab.jl and Python–CatTheory.

A review of previous research reveals significant progress in categorical approaches to database modeling (Spivak, Fong [1, 2]; Schultz, Wisnesky [3]; CQL [4]). However, these studies primarily focus on theoretical aspects and provide limited attention to the practical verification of query correctness and data integrity, particularly in the presence of complex “many-to-many” relationships. Traditional methods, such as Datalog [5] and query containment analysis [6], are effective for verifying query equivalence but are limited in their ability to formally represent non-uniform relationships.

The proposed approach addresses this gap by integrating categorical logic with profunctor semantics, thereby providing a mechanism for verifying both soundness and referential integrity through the commutativity of diagrams.

In the proposed model, a database schema is represented as a small category, where objects correspond to tables, morphisms represent foreign keys, and an attribute functor maps tables to the sets of their attributes. Integrity constraints (foreign keys) are formalized as commutative diagrams, ensuring consistency among relational mappings.

As an illustrative example, consider the typical STUDENTS–COURSES–ENROLLMENTS schema, comprising three interconnected tables:

- Students: id (PK), name, email;
- Courses: id (PK), title, description;
- Enrollments: enrollment_id (PK), student_id (FK $\rightarrow$ Students.id), course_id (FK $\rightarrow$ Courses.id), enrollment_date.

The “many-to-many” relationship between students and courses is implemented through an intermediate table, Enrollments.

In the categorical model:

- Objects: *Students*, *Courses*, *Enrollments*;
- Morphisms: *student*: Enrollments $\rightarrow$ Students, *course*: Enrollments $\rightarrow$ Courses;
- The attribute functor F maps each table to the set of its attributes.

The referential integrity of the foreign key *student_id* is formalized by means of a pullback diagram (Fig.1), which ensures that every enrollment entry consistently corresponds to an existing student record, preserving the integrity of the relational mapping between entities.

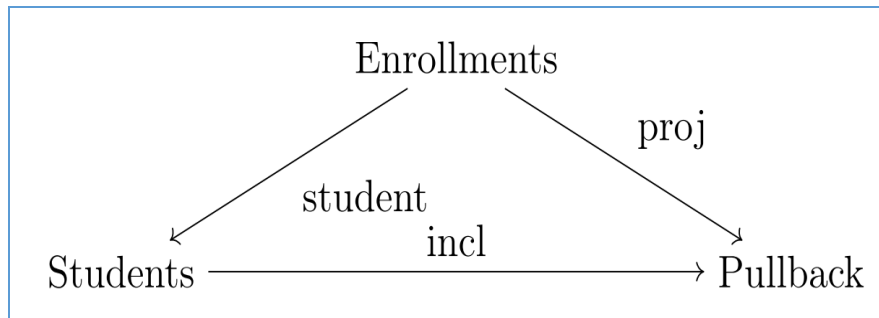


Figure 1 – Pullback- diagram for a foreign key student_id

Queries are modeled as profunctors. For example, a query

```
SELECT Students.name, Courses.title  
FROM   Students JOIN Enrollments ON Students.id =  
Enrollments.student_id  
JOIN Courses ON Courses.id = Enrollments.course_id;
```

can be represented as a profunctor Φ , defined by a span structure:



Figure 2 – Profunctor Span- structure for a query

The proposed approach demonstrates several advantages, including the inherent flexibility of profunctors, the natural generalization of the JOIN operation within a categorical framework, and the ability to perform rigorous formal verification of queries and integrity constraints. However, the method also entails certain limitations, primarily related to computational complexity.

Future research will focus on integrating the model with database management systems, improving computational efficiency through optimization techniques, and extending the approach to support graph database architectures.

References:

1. Spivak D., Fong B. Seven Sketches in Compositionality: An Invitation to Applied Category Theory. MIT Press, 2019.
2. Spivak D. Compositional Structures in Database Semantics. arXiv:2302.08101, 2023.
3. Schultz P., Wisnesky R. Categorical Databases. MIT Press, 2021.
4. Wisnesky R. The Categorical Query Language: Data Integration via Functorial Mappings. Theory and Practice of Logic Programming, 2020.
5. Cohen S., Nutt W., Sagie Y. Deciding equivalences among conjunctive aggregate queries // J. ACM. – 2007. – Vol. 54. – doi:10.1145/1219092.1219093.
6. Kosiur D. Understanding Policy-Based Networking, 2nd ed. Wiley, New York, NY, 2001.

Maryna Hladych

4th year bachelor of computer science student

Iryna Domanetska

Ph.D., associate professor, associate professor at the Department of Intelligent Technologies

Taras Shevchenko National University of Kyiv

ANALYSIS OF EXISTING APPROACHES TO TASK DURATION PREDICTION

Abstract. The study reviews current methods for predicting task duration, with a particular focus on modern machine learning approaches. It analyzes regression models, recurrent neural networks (LSTM), attention-based architectures, and graph neural networks (GNN), which enable modeling of context, sequence, and inter-task dependencies. Real-world datasets such as MS-LaTTE and anonymized Microsoft calendar events are discussed as training sources for these models. The advantages of hybrid solutions that combine classical estimates (e.g., PERT) with ML-based predictions are highlighted. The paper outlines the potential for integrating such models into UX-oriented to-do systems with high levels of personalization and explainability.

Keywords: task duration prediction, machine learning, neural networks, MS-LaTTE, graph models, to-do systems, contextual data.

Accurate prediction of task duration is a critical component of effective planning in both project management and personal to-do systems. Despite the widespread use of digital tools for organizing activities, estimating time remains a challenging problem due to task variability, contextual factors, and limited historical data.

Classical models originating from project management include the PERT method (Program Evaluation and Review Technique), analog estimation, and network-based scheduling. PERT relies on three time estimates—optimistic, most likely, and pessimistic—and calculates expected duration. This approach accounts for uncertainty and risk, making it suitable for complex projects with high unpredictability. Analog estimation, on the other hand, uses historical data from similar tasks to estimate the duration of new ones. While simple to implement, its accuracy depends heavily on the relevance and quality of the data. Network-based scheduling, including Gantt charts and the Critical Path Method (CPM), visualizes task dependencies, identifies time buffers, and optimizes calendar planning.

Modern approaches increasingly rely on machine learning (ML) methods. One example is a gradient boosting model developed by Microsoft Research, which uses task descriptions, user history, and execution context to predict duration. This model demonstrates high accuracy and adaptability in personal to-do systems [4]. It was trained on anonymized calendar events from Microsoft users, including task descriptions, creation and completion timestamps, and durations.

Another study presented at an IEEE conference compares the performance of various ML algorithms—such as Random Forest, XGBoost, and SVR—for predicting task duration in software development. The authors emphasize the importance of feature selection, including task type, assignee role, and historical patterns [5]. These

models were trained on proprietary datasets collected from task management systems (e.g., Jira), containing information on task type, status, assignee, duration, and inter-task relationships.

Neural networks, particularly recurrent models (RNN, LSTM) and attention-based architectures, are gaining traction due to their ability to model task sequences, contextual changes, and unstructured data. These models effectively process textual task descriptions and enable personalized predictions [6]. Training such models requires both private and public datasets. For instance, MS-LaTTE is an open dataset containing over 1 million records of completed tasks from mobile to-do apps, including timestamps, locations, task types, and durations [7]. This dataset supports behavioral modeling and context-aware prediction based on time of day, location, and device type.

A promising direction involves hybrid models that combine classical estimates (e.g., PERT) with ML-based predictions. These approaches offer interpretability and adaptability to dynamic conditions. Graph Neural Networks (GNNs) are particularly noteworthy, as they model task dependencies as graphs—a natural representation for multi-level plans [8]. GNNs are trained on structured data from enterprise systems where tasks have well-defined relationships, statuses, and metadata.

In summary, classical methods remain relevant for structured projects but lack flexibility. Machine learning models offer high accuracy and adaptability, especially when trained on rich datasets. Hybrid and graph-based approaches are the most promising for integration into modern to-do systems, particularly when considering UX requirements and the need for explainable predictions.

References

1. Kim, J., Teevan, J., & Cutrell, E. (2019). Task Duration Estimation. Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM '19). <https://www.microsoft.com/en-us/research/wp-content/uploads/2019/01/TaskDuration-WSDM2019.pdf>
2. Alenezi, M., & Banitaan, S. (2023). Applying Machine Learning to Estimate the Effort and Duration of Software Tasks. *IEEE Access*, 11, 10227275. <https://ieeexplore.ieee.org/document/10227275>
3. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
4. Li, Y., Zhang, Y., & Wang, Y. (2021). MS-LaTTE: A Dataset of Where and When To-do Tasks are Completed. arXiv preprint arXiv:2111.06902. <https://arxiv.org/pdf/2111.06902>
5. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>

Hryhorii Hnatiienko

Ph.D., Associate Professor, deputy dean of the faculty of information technology

Oleh Ilarionov

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Hanna Krasovska

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies

Taras Shevchenko National University of Kyiv

MULTI-ATTRIBUTE DETERMINATION OF MEASURES OF SIMILARITY BETWEEN BRANCHES OF KNOWLEDGE AND SPECIALTIES IN THE TASK OF DOCUMENTARY SELECTION OF CANDIDATES FOR POSITIONS

Abstract. This paper describes the prerequisites and features of determining the similarity measures between fields of knowledge and specialties in the task of documentary selection of candidates in personnel management. It is also proposed to determine the values of the elements of the matrix of similarity of fields of knowledge and the matrix of similarity of specialties by expert means. Based on the use of these matrices, a substantiated determination of the distances between job profiles and candidate profiles built on the basis of the analysis of education documents is provided.

Keywords: selection of candidates for a position, fields of knowledge, specialties, similarity measures, candidate's career

In [1], the problem statement and mathematical model of automated candidate selection using a documentary approach, which is used to assess their competence, are presented. This paper also presents the stages of candidate selection and the tools for determining the similarity measures between job requirement profiles and documentary candidate profiles.

Let us assume that in a decision-making situation related to personnel management, a requirement for the level of education of candidates and the possibility and necessity of formalizing the requirements for work experience required of a candidate are established. Let this requirement be characterized by a vector consisting of two elements

$$x=(x_1, x_2). \quad (1)$$

where x_1 corresponds to the field or fields of knowledge in which the candidate received education;

x_2 – work experience.

That is, developing a profile of the ideal candidate means formalizing their skill set, assessing competencies, determining the level of experience, researching the educational trajectory, etc.

The methods of specifying requirements for the documentary level of candidates have the same format as the requirements for the position or for the profile

of the ideal candidate of type (1). That is, the recruiter or the candidates themselves provide feedback in the form of vectors of type

$$q_i = (q_{i1}, q_{i2}), i \in I = \{1, \dots, n\}, \quad (2)$$

where q_{i1} is a vector consisting of three elements corresponding to the specialties according to diplomas, respectively, 1st, 2nd and 3rd levels of education, which the i -th candidate received;

q_{i2} is the work experience of the i -th candidate for the position.

It is necessary to determine the degree of proximity of the requirements of the selection organizer and the candidate profiles based on the analysis of vectors (1) and (2). Subsequently, after calculating these degrees of proximity, a rating is compiled and the selection is made.

In this case, metrics should be used to measure the distances between the job profiles (1) and the candidate profiles (2):

$$d_i = d(x, q_i), i \in I. \quad (3)$$

To do this, first, an expert-based matrix of similarity of fields of knowledge Mg and a matrix of similarity of specialties Ms is created.

The Resolution of the Cabinet of Ministers of Ukraine [2] approved the List of fields of knowledge and specialties in which applicants for higher and professional pre-higher education are trained:

- A. Education
- B. Culture, arts and humanities
- C. Social sciences, journalism, information and international relations
- D. Business, administration and law
- E. Natural sciences, mathematics and statistics
- F. Information technology
- G. Engineering, production and construction
- H. Agriculture, forestry, fisheries and veterinary medicine
- I. Health and social security
- J. Transport and services
- K. Security and defense

Based on this list of branches, we create a square matrix Mg with dimensions 10×10 . The elements of this matrix are the measures of similarity between branches of knowledge, the codes and names of which are given above. To calculate the metric values of the measures of similarity between branches of knowledge, it is advisable to apply multi-attribute selection methods. The attributes in this situation can be, in particular, the following areas:

- m_1 — Artificial Intelligence
- m_2 — Digital Transformation
- m_3 — Cybersecurity
- m_4 — Data Use
- m_5 — Mutual Influence

- m₆– Online Learning and Remote Work
- m₇– Lifelong Learning
- m₈– Globalization
- m₉– Tool Development
- m₁₀– Social Networks and Media
- m₁₁– Technology Use
- m₁₂– Critical Thinking

Remark 1. It is obvious that the number of attributes on the basis of which the expert assessment of common features and relationships between different specialties is carried out can be supplemented. The similarity measures of branches in multi-attribute selection can be determined using ordinal and cardinal scales and using different similarity measures [3].

Similarly, the elements of the similarity matrix of specialties Ms within one or different branches of knowledge are also determined. The main attributes in this case can be:

- s₁– Common theoretical basis
- s₂– Intersection of skills
- s₃– Intersection of competencies
- s₄– Functional connection
- s₅– Emphasis on practical or theoretical aspects of education

s₆– Emphasis on theory, fundamental principles, mathematics, algorithms, artificial intelligence, etc.

- s₆– Commonality or difference of academic and professional space

Remark 2. When calculating the values of similarity measures of the type (3) based on the similarity matrix of specialties, the values of the elements of the matrix Mg and the matrix Ms determined by experts are also taken into account.

In addition to the expert formation and use of the similarity matrix of knowledge areas Mg and the similarity matrix of specialties Ms , it is also logical to use a multi-attribute comparison of educational trajectories to determine the relationship of preference between applicants for positions.

For example, we will give a scheme for quantitative assessment of two options for educational trajectories within one field of knowledge F (information technologies):

option 1 – bachelor $F4$ (systems analysis), master $F3$ (computer science), postgraduate $F6$ (information systems and technologies);

option 2 – bachelor $F5$ (Cybersecurity and information protection), master $F5$ (Cybersecurity and information protection), postgraduate $F5$ (Cybersecurity and information protection).

It is clear that the quantitative assessment of each educational trajectory option depends on the level of the position and the goal of recruitment.

Qualitative analysis indicates that option 1 forms a universal strategist-researcher with a wide coverage of IT disciplines, and option 2 forms a highly specialized expert-practitioner in the field of security.

Expert assessments, which were carried out on a 100-point scale, are presented in Table 1.

Table 1

Results of expert evaluation of educational trajectories (*F4-F3-F6*) and (*F5-F5-F5*) according to various criteria

Evaluation criteria	Option 1: Universal (F4-F3-F6)	Option 2: Specialist (F5-F5-F5)
For a career as a top manager/CIO/CDO/RND	90-95 points	60-65 points
For a career as a scientist/researcher	90-95 points	65-70 points
For a career as a narrow technical expert in KB	65-70 points	80-95 points
For flexibility in the labor market (changing fields)	80-90 points	65-70 points
Resulting quantitative assessment	90 points	65 points

Remark 3. The resulting quantitative scores will be calculated based on the layering method described in [4].

At the next stage of selection, based on the results of the technical interview and assessments of the proximity of the candidates' experience and its duration, the obtained values (3) are refined, depending on the level of the position. In this case, emphasis may be placed on creativity, diversity of interests, depth of specialized knowledge, breadth of views on future activities, etc.

Thus, this paper considers the formulation of the problem of preliminary documentary selection of candidates for a position. Two approaches to the formation of a matrix of similarity of fields of knowledge and a matrix of similarity of specialties are presented. In addition, an approach to multi-attribute comparison of educational trajectories of different candidates is proposed, which can become the basis for additional analysis and comparison of candidates. Preliminary selection of candidates for positions based on a documentary approach has received significant development through the approaches introduced by the authors to formalize this area of research.

References:

1. Hryhorii Hnatiienko, Oleg Ilarionov, Hanna Krasovska, Yevhen Patkin. Mathematical model of automated selection of candidates with a documentary approach to assessing their competence // Applied information systems and technologies in the digital society: collection of abstracts and scientific reports of participants of the IX International Scientific and Practical Conference (Kyiv, October 1, 2025) / edited by V. Pleskach, O. Fendyo. Kyiv: Taras Shevchenko National University, 2025. P. 66-76. (in Ukrainian)
2. Resolution of the Cabinet of Ministers of Ukraine dated February 21, 2025 No. 188 on approval of the List of branches of knowledge and specialties in which applicants for higher and professional pre-higher education are trained. <https://zakon.rada.gov.ua/laws/show/188-2025-%D0%BF#Text>. (in Ukrainian)
3. Hnatiienko, H., Hnatiienko, O., Tmienova, N., Snytyuk, V. Mathematical Model of Management of the Corporate Culture of the Organizational System / CEUR Workshop Proceedings, Volume 3624, Pages 250-265, 2023 // Selected Papers Selected Papers of the X International Scientific Conference "Information Technology and Implementation" (IT&I-2023). Conference Proceedings. Kyiv, Ukraine, November 20 - 21, 2023.
4. Hnatiienko, H. M., Hnatiienko, O. H. (2024). Method for determining fuzzy values of the relative importance of characteristics of alternatives using the layering method // Scientific Bulletin of Uzhgorod University. Mathematics and Computer Science Series, 45(2), 172–187. DOI [https://doi.org/10.24144/2616-7700.2024.45\(2\).172-187](https://doi.org/10.24144/2616-7700.2024.45(2).172-187) (in Ukrainian)

Khodatskyi Volodymyr

Student

Anastasiia Nikolaenko

Ph.D., Assistant of the Department of Software Systems and Technologies

Taras Shevchenko National University of Kyiv

VISUALIZATION AND COMPARISON OF PATHFINDING ALGORITHMS BETWEEN TWO POINTS

Abstract. This paper investigates algorithms for finding optimal paths between two points in a two-dimensional grid space. A comparative analysis of four algorithms was conducted: depth-first search, breadth-first search, Dijkstra's algorithm, and A*.

Keywords: depth-first search, breadth-first search, Dijkstra's algorithm, A* search, recursive backtracking.

Introduction. The rapid development of digital technologies has led to widespread application of optimal route-finding algorithms in various areas of human activity. However, for effective use of these methods, it is important not only to be able to apply them, but also to deeply understand the principles of their operation, peculiarities of behavior under different conditions, and criteria for choosing the optimal approach for a specific task. This research is dedicated to developing a tool for analyzing and visually demonstrating the operation of four fundamental algorithms – depth-first search, breadth-first search, Dijkstra's algorithm, and A* heuristic search – in a two-dimensional grid environment with the ability to model different scenarios.

Literature Review. Pathfinding algorithms are a fundamental topic in graph theory and computer science. Such algorithms can be divided into uninformed and informed depending on the use of additional information to guide the search [1]. Depth-first search and breadth-first search are basic methods of graph traversal. Depth-first search uses a stack to explore vertices, making it simple to implement, but it does not guarantee finding the shortest path [2]. In contrast, breadth-first search uses a queue for sequential exploration of all neighboring vertices at each level of the graph, which ensures the shortest path in unweighted graphs [1].

Dijkstra's algorithm is an optimal solution for weighted graphs with non-negative weights, using a priority queue to select the vertex with the smallest current path cost [1]. However, in large graphs, this algorithm can be inefficient due to the need to examine all neighboring vertices. The A* algorithm is a heuristic extension of Dijkstra's algorithm [3]. It uses an evaluation function (heuristic) that allows reducing the number of explored vertices by directing the search toward the target point. The heuristic must be admissible (not exceeding the true path cost) to ensure A* optimality [4].

Objective. The objective of this work is to research and analytically compare pathfinding algorithms in a two-dimensional grid space considering obstacles and weight coefficients, as well as to develop software for their visualization.

Research objectives. To achieve the objective, the following tasks were set: to conduct an analysis of existing pathfinding algorithms, to investigate their theoretical foundations including operating principles, mathematical models and limitations, to design the architecture of a software solution, to develop an application with algorithm visualization support, to conduct a comparative analysis of algorithm efficiency in different graph configurations.

Research Methods. Theoretical analysis of literature sources allowed studying existing approaches to implementing pathfinding algorithms and methods of their visualization. Object-oriented programming was applied to create a modular program structure that ensures flexibility and the possibility of expanding functionality. Python language with the Pygame library was used for implementation, providing capabilities for creating a graphical interface and handling user events.

Comparative analysis of algorithm efficiency was conducted based on four metrics: algorithm execution time, number of visited cells, length of the found path, and total path weight. Testing was performed in three different scenarios: a grid without weights and obstacles, a grid with weights and obstacles, and a maze without weights. Recursive backtracking algorithm was used for maze generation. All results were saved as screenshots with metrics for further analysis.

Research Results. Fig. 1 shows a typical result of the pathfinding algorithm visualization on a grid with generated obstacles and weights.

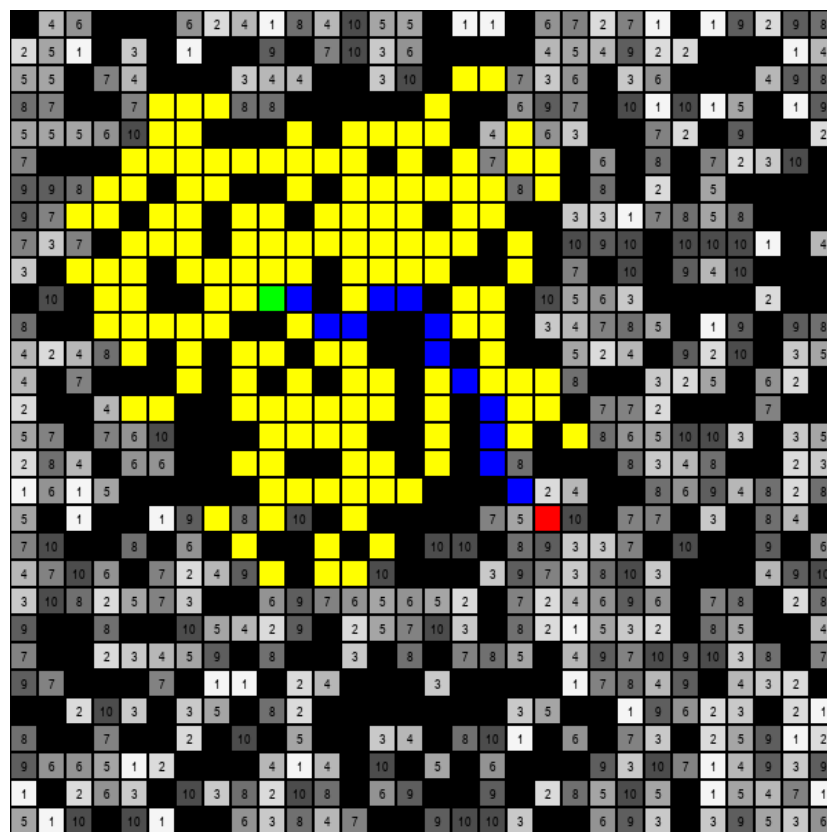


Figure 1 – Result of A* search execution on a grid with weights and obstacles

In the testing scenario on an empty grid without obstacles and weights, A* search demonstrated the highest efficiency, taking only 0.13 seconds and visiting only 20 cells. In comparison, Dijkstra's algorithm completed the task in 2.11 seconds, visiting 450 cells, breadth-first search took 10.55 seconds with 396 visited cells, and depth-first search took 5.94 seconds with 588 cells. Regarding path length and weight, breadth-first search, Dijkstra's algorithm, and A* search found the optimal path of 11 cells with weight 13.31, while depth-first search found a suboptimal path of 583 cells with weight 816.44.

In the grid with weights and obstacles, A* search again proved most efficient in execution time (0.65 s) and number of visited cells (160). Dijkstra's algorithm performed the search in 1.24 seconds, visiting 268 cells, breadth-first search took 5.89 seconds with 314 visited cells, and depth-first search took 4.89 seconds with 415 cells. Breadth-first search, Dijkstra's algorithm, and A* search found the optimal path of 14 cells with weight 56.36, while depth-first search showed a path of 280 cells with weight 2039.12, which is significantly worse than the optimal result.

In the maze without weights, A* search took 1.12 seconds, visiting 263 cells, Dijkstra's algorithm took 1.44 seconds with 340 visited cells, breadth-first search took 3.68 seconds with 340 cells, and depth-first search took 3.86 seconds with 405 cells. Breadth-first search, Dijkstra's algorithm, and A* search found the optimal path of 140 cells with weight 165.92, while depth-first search found a path of 159 cells with weight 185.34.

Conclusions

As a result of the research, a comparative analysis was conducted, which showed that the A* algorithm is the most efficient in all investigated scenarios due to the use of a heuristic function. In the grid with weights and obstacles, it proved to be 47.58% faster than Dijkstra's algorithm, 88.96% faster than breadth-first search, and 86.71% faster than depth-first search. Additionally, A* search visited the smallest number of cells (160) compared to other algorithms.

Dijkstra's algorithm ensures path optimality in weighted graphs but is inferior to A* search in speed due to the absence of heuristics. Breadth-first search is appropriate for use in grids without weights, where it guarantees the shortest path, but it is less efficient in execution time. Depth-first search proved to be the least efficient in all scenarios due to a suboptimal search strategy, although it can be useful for demonstrating the principle of deep graph traversal.

References

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C. Introduction to Algorithms. – K.: BHV Publishing Group, 2010. – 1296 p.
2. Skiena, S. The Algorithm Design Manual. – Lviv: SPOLOM, 2012. – 720 p.
3. Knuth, D. The Art of Computer Programming. Volume 3: Sorting and Searching. – K.: Williams, 2000. – 800 p.
4. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. – K.: Williams, 2006. – 1152 p.

Oleksandr Kolchyn

PhD, senior researcher

Vlad Volkov

PhD, senior researcher

Dariia Rudenko

PhD student

V.M. Glushkov Institute of cybernetics NAS of Ukraine

TOWARDS REVEALING GAPS IN INPUT VALIDATION LOGIC VIA EVOLUTIONARY CONSTRAINT RELAXATION

Abstract. In this work, we propose a novel evolutionary approach to identifying gaps in input validation logic—execution paths where checks on untrusted external input are either missing or significantly relaxed. Our key insight is to frame the detection problem as an evolutionary search for constraint relaxation: by exploring alternative execution paths that reach the same sink location, we attempt to discover weaker or bypassed validations on the same input-dependable variable.

Keywords: input validation, testing, taint analysis

Robust input validation is a foundational aspect of cybersecurity. Improper handling of external or untrusted input remains a leading cause of security vulnerabilities — for example, “Improper Input Validation” is consistently ranked among the most critical issues in the CWE Most Dangerous Software Weaknesses [1]. Programs often contain inconsistencies or gaps in input validation logic: different execution paths may apply different constraints to the same input variable [2] before it reaches a security-critical operation.

Let a program contain a sink location s , dependent on an input variable x . For a given execution path t that reaches this sink, let $R(t, s, x)$ denote the set of symbolic constraints imposed on x along this path at s . These constraints typically originate from input checks, conditional statements, or assertions in the code.

Our goal is to find an alternative execution path t' to the same sink such that:

$$R(t, s, x) \subset R(t', s, x)$$

That is, the set of inputs reaching the sink along t' strictly includes those reachable via t , indicating a potential validation gap: some inputs are permitted in certain paths but blocked in others.

We employ a genetic algorithm to search for such relaxed paths t' , guided by a custom-designed fitness function that quantifies how much the constraints on x are weakened compared to a reference path t . We extract symbolic constraints $R(t, s, x)$ and $R(t', s, x)$ using symbolic modeling [3] with constraint reconstruction. Rather than merely counting removed constraints, we assign each constraint c a weight $w(c) \in \mathbb{R}^+$ based on its restrictiveness. For example, exact inequality ($0 < x < 10$) is treated as more restrictive than disequality ($x \neq 0$).

The fitness function is defined as:

$$\text{fitness}(t') = \sum_{c \in R(t,x) \setminus R(t',x)} w(c) + \sum_{c \in R(t,x) \cap R(t',x)} \delta(c, c') \cdot w(c)$$

Where:

- $\delta(c, c') \in [0,1]$ measures the degree to which constraint c is weakened in t' (with $\delta = 1$ meaning completely removed),
- $w(c)$ reflects the semantic strictness of each constraint.

The algorithm performs a search over the program's execution paths with the goal of identifying those that impose weaker constraints on a target input variable at a given sink. It follows a population-based evolutionary strategy, where each candidate solution corresponds to a feasible path through the program. At each iteration, the algorithm evaluates the fitness of each path based on the restrictiveness of the input validation logic it encounters. Paths that allow broader input ranges receive higher fitness scores. To focus the search on relevant parts of the control-flow graph, we employ backward slicing with respect to the pair (s, x) . Moreover, to better target relaxed input validations, we refine the slicing process by excluding branches that enforce overly restrictive constraints on x . For instance, a condition such as *if* ($x == 0$) admits only a singleton value and therefore does not contribute to discovering weaker validations. As a result, such control-flow dependencies are not included in the slice at all, and execution paths passing through them are eliminated from consideration.

Initial experiments on synthetic programs show that the method can find paths with relaxed constraints on the sink variable and generate inputs that bypass checks, revealing validation gaps. The proposed approach can also be applied to augment existing test suites by identifying alternative execution paths with less restrictive constraints on taint-tracked variables. Such paths implicitly define wider value ranges, exposing new boundary conditions that may not be covered by the original tests. These newly revealed boundaries can then serve as targets for additional test inputs.

An important direction for future work is to refine the constraint weighting function $w(c)$ by learning from empirical data. For example, by correlating constraint types with observed input rejection or acceptance behavior across execution traces, the system could adaptively assign weights that reflect the actual semantic impact of each constraint.

References:

1. Common Weakness Enumeration (CWE) Top 25 Most Dangerous Software Weaknesses. https://cwe.mitre.org/top25/archive/2024/2024_kev_list.html
2. Kolchin, A., Potiyenko, S. (2022). Extending Data Flow Coverage to Test Constraint Refinements. In: ter Beek, M.H., Monahan, R. (eds) Integrated Formal Methods. IFM 2022. Lecture Notes in Computer Science, vol 13274.
3. Letichevsky, A. A., Godlevsky, A. B., Letychevsky Jr, A. A., Potiyenko, S. V., & Peschanenko, V. S. (2010). Properties of a predicate transformer of the VRS system. *Cybernetics and Systems Analysis*, 46(4), 521-532.

Oleh Kotliar

2nd year Master's student, Department of Software Systems and Technologies

Kseniia Dukhnovska

PhD in Technical sciences, Assoc. Prof. of Software Systems and Technologies

Department

Faculty of Information Technology, Taras Shevchenko National University of Kyiv

SOFTWARE FOR PERSONALIZED HRIS REPORT GENERATION: INTERACTIVE REPORT BUILDER WITH SERVER-SIDE PDF RENDERING

Abstract. We present a software architecture and method for fast, personalized HR reporting directly from an HRIS. The approach combines a client-side report builder for column selection and multi-level filters with a server pipeline that compiles dynamic queries, generates HTML layouts, and renders print-perfect PDFs. The stack uses ASP.NET Core, Entity Framework Core for projection-only queries, and a headless browser for consistent PDF export. Caching of presets and results reduces latency under repeated workloads. The result is a deployable client-server system that produces tailored reports with low response times and without external BI platforms.

Keywords: HRIS, dynamic reporting; interactive filters, projection queries, ASP.NET Core, Entity Framework Core.

Introduction. Organizations need on-demand HR reports that match evolving business questions [1]. Off-the-shelf HRIS reports are rigid, while full BI stacks add cost and integration overhead. The goal is an in-product report builder that lets users select fields, define complex predicates, preview data, and export uniform PDFs, all under role-based security and audit.

Background. HR datasets are broad, deeply relational, and privacy-critical [2]. Typical HRIS tables mix slowly changing attributes (grade, department, manager) with event streams (hires, transfers, leaves, payroll runs). Naive “dump then format” workflows overload databases, inflate memory in application tiers, and leak personally identifiable information during ad-hoc handling. Effective reporting requires three properties to coexist: minimal data movement through strict projection, deterministic presentation for reproducibility, and enforceable governance over every attribute and row. Projection means returning only the columns selected by the user rather than materializing full entities or DTOs. Governance means row-level and column-level security that tracks business rules and legal bases, not only coarse roles. Determinism means the same inputs produce byte-identical layouts regardless of client environment, which eliminates “works on my machine” artifacts and stabilizes audit trails.

Many vendor solutions deliver one or two of these properties but rarely the full set within a single transactional path. BI stacks optimize visualization and caching but add integration latency, duplicated security models, and divergent data semantics. Report generators inside ERPs often render predictably but force rigid schemas and prebuilt reports that do not fit emergent questions from HR and finance [3]. In practice, teams either accept slow, manual exports or proliferate shadow spreadsheets that bypass access control. The background for this work is a need to collapse query compilation, layouting, and export into one governed pipeline embedded in the HRIS,

so performance is dictated by database selectivity and template rendering, not by client hardware or manual reformatting.

Method. The system accepts a declarative report schema from the UI, validates the caller's permissions, and compiles the specification into LINQ expression trees that describe projections, predicates, joins, paging, and sorting. A column mapper translates user-selected fields into projection expressions so the database returns only necessary attributes, while a predicate factory assembles composable filters that short-circuit and avoid unnecessary joins. The API executes the compiled query with mandatory pagination and produces a stable preview. For distribution, the same projected dataset is injected into an HTML template with localized headers, group totals, page numbers, and signature blocks; a headless browser such as Playwright renders this HTML to a PDF with fixed margins and consistent typography. Optional CSV or Excel exports stream from the identical projection to keep outputs consistent. Performance is maintained through targeted indexing, capped column counts, strict timeouts, and caching of both result sets and common report presets; reliability follows from an asynchronous pipeline with back-pressure and full audit logs of schemas, parameters, and generated files. Security is enforced through role- and attribute-based rules at compilation time, input validation for the schema and filters, and redaction of personally identifiable information in telemetry.

The sequence diagram (Fig. 1) illustrates the interaction process among system components during report generation. The process begins when the user, through their web client, opens the report configuration page. The server receives the request and displays the corresponding interface with a set of available columns and filtering options.

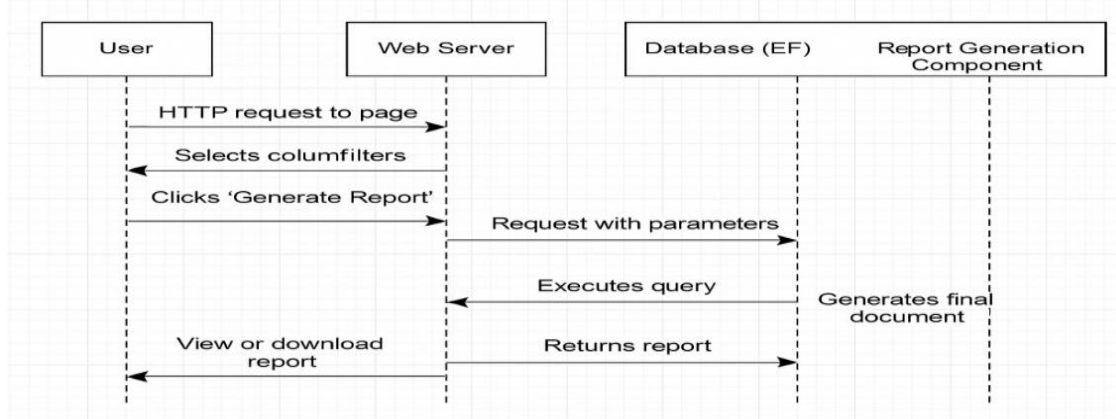


Figure 1 – Sequence diagram

Conclusion. The architecture delivers low-latency, policy-compliant HR reporting without external BI. Dynamic projections minimize data movement, template HTML ensures consistent presentation, and headless PDF export standardizes delivery. The design scales with data width and user concurrency while preserving governance.

References:

1. Armstrong M., Taylor S. Armstrong's Handbook of Human Resource Management Practice. – 15th ed. – Kogan Page Publishers, 2020. – 800 p.
2. Dessler G. Human Resource Management. – 16th ed. – Pearson Education, 2020. – 720 p.
3. Kavanagh M. J., Johnson R. D. Human Resource Information Systems: Basics, Applications, and Future Directions. – 4th ed. – SAGE Publications, Inc., 2021. – 592 p.

Yevhen Kutniak

master's degree student of the Department of Management Technologies

Oleksandr Timinskyi

PhD, associated professor of the Department of Management Technologies

Taras Shevchenko National University of Kyiv

EFFECTIVE CONTROL OF TECHNICAL DEBT IN IT PRODUCTS

Abstract. Technical debt, a metaphor originally coined by Ward Cunningham, has become a central concern within modern IT product lifecycle control. Effective technical debt control is essential for sustaining software quality, accelerating feature delivery, and protecting long-term business value. This paper outlines current definitions, underlying causes, quantitative metrics, and contemporary strategies for technical debt control within IT products, focusing particularly on Agile and DevOps environments. Recommendations are formulated based on academic research and industry case studies.

Keywords: Technical debt, software engineering, Agile, DevOps, refactoring, metrics, risk control.

Introduction. Technical debt (TD) refers to the implied cost of additional rework caused by expedient decisions during software development, typically made to meet short-term goals at the expense of long-term maintainability and adaptability. Over the past decade, technical debt has grown from a theoretical concept into a concrete business risk, influencing everything from project timelines and operational costs to software stability and organizational competitiveness.

Recent global surveys indicate that technical debt represents up to 40% of IT product assets, and organizations can spend 10–20% extra per project addressing debt-related issues. Left uncontrolled, TD can significantly constrain innovation, slow business processes, and endanger system security and user experience.

Ward Cunningham first described technical debt in 1992, observing that “a little debt speeds development as long as it is paid back promptly with a rewrite.” He likened the cost of imperfect code to monetary debt, warning that accumulating interest (in the form of increased future effort) can hinder progress if not controlled. Modern academic definitions stress that technical debt consists of “design or implementation constructs that are expedient in the short term, but set up a technical context that makes future changes more costly or impossible”. Debt can originate from not only coding shortcuts but also architectural compromises, documentation omissions, and suboptimal testing practices.

Martin Fowler's “Technical Debt Quadrant” remains influential, distinguishing TD along two axes – intentionality (deliberate vs accidental) and recklessness (prudent vs reckless):

- Deliberate, Prudent Debt: Strategic trade-offs consciously made and tracked for short-term gain (e.g., rapid prototyping).
- Deliberate, Reckless Debt: Conscious shortcuts without proper consideration of long-term risk.

- Inadvertent, Prudent Debt: Debt arising from unforeseen consequences or newly discovered best practices.
- Inadvertent, Reckless Debt: Accidental, high-risk debt due to ignorance or lack of discipline.

Common types of technical debt include:

- Architectural debt: Inefficient system structures hindering scalability and maintainability.
- Code debt: Code smells, lack of refactoring, or excessive complexity.
- Test debt: Insufficient or outdated test coverage.
- Documentation debt: Incomplete, missing, or obsolete documentation.
- Security debt: Vulnerabilities introduced by delaying security controls.

Root Causes of Technical Debt. Main causes include business pressure for rapid delivery, lack of testing and automation, evolving requirements, legacy components, insufficient developer experience, and inadequate documentation. Organizational culture may also discourage investment in quality improvement, leading to mounting interest on unpaid debt.

Quantifying Technical Debt. Key technical debt metrics:

- Tech Debt Ratio (TDR): Ratio of effort required for remediation to system development effort.
- Code Coverage: Percentage of code exercised by automated tests.
- Cyclomatic and Cognitive Complexity: Indicate difficulty of code maintenance.
- Number and Age of Known Issues: Track unresolved defects over time.
- Defect density: Frequency of bugs per lines of code.

SQALE and Gartner frameworks provide structured means of scoring and visualizing systemic debt, incorporating business impact and remediation cost.

Preventive Approaches

- Intentional planning and documentation: Record all design compromises and their rationale.
- Continuous refactoring: Allocate 15–20% of development cycles to codebase improvement and simplification.
- Increasing code quality: Enforce code standards, peer reviews, and modular architecture.
- Expanding automated testing: Include comprehensive unit, integration, and end-to-end tests.
- Agile backlog control: Treat high-impact debt as backlog items and prioritize based on business value and risk.

Reactive Approaches

- Remediation sprints: Dedicate specific development cycles to repay debt.
- Incremental improvements: Embed small refactoring tasks within feature work.
- DevOps automation: Use CI/CD and infrastructure-as-code to enforce best practices and quickly surface regressions.

Tooling

- Static analysis: Tools like SonarQube and CAST automate detection of code smells, complexity, and vulnerabilities.
- Issue tracking and dashboards: Jira, Azure Boards, and GitHub Issues support the visibility and prioritization of technical debt.
- Automated metrics gathering: Monitor debt and remediation progress to inform strategic decisions.

Business Impact and Risks

Failure to control technical debt can lead to escalating maintenance costs, project delays, and an erosion of developer productivity and morale. Systems burdened by growing debt are vulnerable to instability, security breaches, and obsolescence, ultimately undermining business objectives.

Organizations that actively control technical debt achieve faster feature delivery, improved quality, lower costs, and greater resilience during periods of rapid change.

Practical Recommendations

- Make technical debt visible through tracking tools and dashboards.
- Align refactoring and remediation with business priorities.
- Foster a culture of code quality and accountability.
- Regularly review and update dependencies to minimize security debt.
- Document all technical compromises and reasoning.
- Promote business-technology communication to justify investment in quality over the long term.

Future Work

Future studies on technical debt control should aim to establish more objective and automated methods for its detection and quantification. While current practices rely heavily on expert judgment and manual assessment, machine learning and natural language processing can be leveraged to analyze source code, documentation, and issue-tracking data to identify patterns that indicate emerging debt. Such approaches could significantly reduce human bias and improve early warning capabilities, allowing organizations to act before the debt accumulates beyond control.

Another promising direction involves the development of predictive and economic models that estimate the financial and operational impact of technical debt over time. Integrating these models into project control and DevOps tools would allow organizations to visualize long-term trade-offs between short-term delivery speed and software maintainability. Simulation-based approaches could further enhance this analysis by forecasting the effect of different repayment or prevention strategies on future software evolution, cost efficiency, and business value.

At the same time, the rapid advancement of generative AI tools and so-called “vibe coding” practices—where developers or even non-developers rely heavily on AI-generated suggestions without fully understanding or validating the produced code—poses new challenges. These technologies accelerate delivery but can unintentionally amplify existing problems of poor structure, duplicated logic, and missing

documentation. As a result, the scope and complexity of technical debt are likely to intensify in the near future. Future research should therefore focus on creating frameworks and tools capable of identifying AI-generated code artifacts, quantifying their risk, and enforcing explainability and accountability in automated coding workflows.

Beyond technical aspects, future investigations could examine organizational and educational responses to this shift. Studies may explore how developer training, team communication, and control awareness can adapt to a hybrid human-AI programming environment. Finally, building comprehensive governance models that integrate technical, cultural, and economic perspectives will be crucial for sustaining software quality in an era where automation and creativity coexist. Such models should not only measure technical debt but also provide actionable strategies to balance innovation speed with long-term sustainability.

Conclusions.

Effective control of technical debt in IT products demands both systemic prevention and active remediation, supported by clear metrics and a culture that values long-term software health. In Agile and DevOps-driven environments, technical debt should be treated as a first-class consideration – tracked, quantified, prioritized, and continuously paid down. Doing so yields sustainable innovation, reduced risk, and tangible gains in competitive advantage.

References:

1. Bavota, Gabriele, and Barbara Russo. "A large-scale empirical study on self-admitted technical debt." *Proceedings of the 13th International Conference on Mining Software Repositories (MSR '16)*. ACM, 2016, pp. 315–326. DOI: [10.1145/2901739.2901742](https://doi.org/10.1145/2901739.2901742).
2. Buschmann, Frank. "To Pay or Not to Pay Technical Debt." *IEEE Software* 28.6 (2011): 29–31.
3. Woubshet Nema Behutiye, Pilar Rodríguez, Markku Oivo, and Ayşe Tosun. "Analyzing the concept of technical debt in the context of agile software development: A systematic literature review." *Information and Software Technology*, vol. 82, 2017, pp. 139-158. DOI: [10.1016/j.infsof.2016.10.004](https://doi.org/10.1016/j.infsof.2016.10.004)
4. Brown, Nanette, Yuanfang Cai, Yuepu Guo, Rick Kazman, Miryung Kim, Philippe Kruchten, Erin Lim, Alan MacCormack, Robert Nord, Ipek Ozkaya, Raghvinder Sangwan, Carolyn Seaman, Kevin Sullivan, and Nico Zazworka. "Managing Technical Debt in Software-Reliant Systems." *Proceedings of the 18th FSE/SDP Workshop on the Future of Software Engineering Research (FoSER '10)*, ACM, 2010, pp. 47-52. DOI: [10.1145/1882362.1882373](https://doi.org/10.1145/1882362.1882373).
5. Nugroho, Ariadi, Joost Visser, and Tobias Kuipers. "An empirical model of technical debt and interest." *Proceedings of the 2nd International Workshop on Managing Technical Debt (MTD '11)*. ACM, 2011, pp. 1–8. DOI: [10.1145/1985362.1985364](https://doi.org/10.1145/1985362.1985364).

¹ **Nataliia Pashynska**

PhD, Assistant

² **Viktor Putrenko**

Doctor of Technical Sciences, Professor

¹ *Taras Shevchenko National University of Kyiv,*

² *American University Kyiv*

THE SYSTEM APPROACH FOR DEVELOPMENT OF A DATABASE FOR THE PURPOSES OF REGIONAL CLIMATE CHANGE MODELING

Abstract. The study used data from global and regional climate change modelling in the Zakarpattia region for the period up to 2050 using various climate models and scenarios. The study also organized data on the administrative-territorial division of the territory, land cover monitoring, analysis of the territory's relief, statistics on natural hazards and socio-economic monitoring of the region's development on the base of geospatial database.

Keywords: Climate Change, Databases, Geoinformation Modelling.

One of the greatest challenges of our time is the process of global climate change, which directly affects all spheres of life. However, the effects of climate change have very different results depending on the spatial and temporal section. Therefore, for the adaptation needs of individual communities, it is necessary to develop information systems that are capable of covering large diverse data sets not only in the field of climatology, but also in the social, economic, environmental, management and security domains. These tasks are solved on the basis of the formation of geoinformation systems with distributed databases, applied modeling algorithms and decision support modules.

Methodology will be adjusted according to conditions in Zakarpattya oblast in Ukraine and respecting expert opinions of project team members or subcontractors [1]. Analysis of exposure, sensitivity, impact, adaptive capacity, and vulnerability is relevant for the preparation process of adaptation strategy. We are not solving problems of emissions, mitigation, climate variability etc., which are subjects of IPPC, energy efficiency strategies, sustainable transport plans or reduction of carbon emissions strategies (Figure 1).

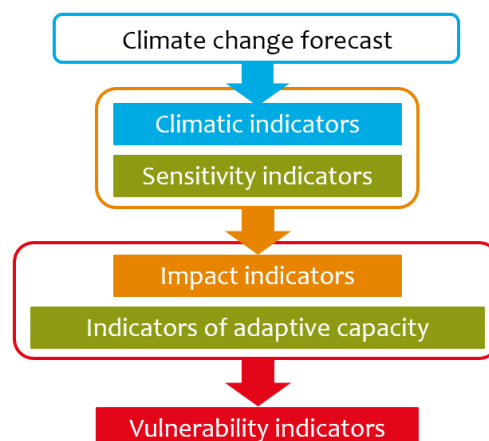


Figure 1. A general model for assessing the impact of climate change

At the first stage, the basic administrative units of data collection and unification were determined. Such units are the territorial communities that were formed in the Zakarpattia region in 2019 in the course of the decentralization reform.

Individual indicators are expressed in numerical value for each territory. One numerical value is assigned for each indicator to each territory (municipality). For better clarity, indicators are usually expressed in the form of a map (scheme) where each area of the municipality is assigned to one color depending on the value of the indicator [2].

We can represent the structure of a geodatabase model as a system of interconnected entities and relationships.

Formally, the complex multifactor system can be described as basis sets, variables, operators/mappings, and objectives.

System:

$$\mathcal{S} = \langle \Omega, T, \mathcal{A}, \mathcal{N}, X, U, Y, P, Z, \mathcal{C}, \mathcal{J}, F, G, H \rangle;$$

Base sets:

$\Omega \subset \mathbb{R}^2$ - spatial domain; $T = \{t_0, \dots, t_K\}$ - time.

$\mathcal{A} = \{A_1, \dots, A_M\}$ - administrative units; $\mathcal{N} : \Omega \rightarrow 2^\Omega$ - neighborhood map.

Variables:

$x_t : \Omega \rightarrow [0,1]^m$ - state vector (normalized indicators).

$u_t : \Omega \rightarrow \mathbb{R}^r$ - inputs/scenarios; $y_t \in Y$ - outputs/products.

$z_t : \Omega \rightarrow \mathbb{R}^q$ - observables.

$P \in \mathbb{R}^p$ - parameters (weights, thresholds).

Operators / mappings:

$H : \mathcal{D} \rightarrow Z$ - measurement / data extraction.

$\phi_i : \mathbb{R} \rightarrow [0,1]$ - normalization; $n_{\{i,t\}}(p) = \phi_i(z_{\{i,t\}}(p))$.

$\text{Agg}_A : \{f : \Omega \rightarrow \mathbb{R}\} \rightarrow \mathbb{R}$; $n_{\{i,t\}}(A) = \text{Agg}_A(n_{\{i,t\}})$.

$F : X \times U \times P \rightarrow X$; $x_{\{t+1\}} = F(x_t, u_t; P)$.

$G : X \times P \rightarrow Y$; $y_t = G(x_t; P)$.

Quality / constraints / objectives

$\mathcal{C} = \{K_{\text{top}}, K_{\text{norm}}, K_{\text{data}}\}$ - validity predicates ($K(\cdot)=\text{true} \Rightarrow$ admissible).

$\mathcal{J} = \{J_\ell\}$; e.g., minimize

$\min J = \alpha \cdot E_t[\sum_{\{A \in \mathcal{A}\}} V_t(A)] + \beta \cdot R_{(w)}$ subject to $K \in \mathcal{C}$.

To conduct the study, data from global and regional modeling of climate change in the territory of the Transcarpathian region for the period until 2030 were used using various climate models and scenarios. The work also used data from the administrative-territorial division of the territory, land cover monitoring, analysis of the relief of the territory, statistics of dangerous natural situations, and socio-economic monitoring of the development of the region. All data were structured on the basis of the formation of a number of geospatial databases [3].

Data analysis was carried out on the basis of modeling and overlaying data layers with unified spatial and temporal resolution. For the purpose of joint analysis

of data series, methods of normalization of numerical values are used. The main result of the modeling is the prediction of changes in the main parameters of the socio-economic and ecological state of the territory, identification of the most promising adaptation measures.

Based on the collected data, a geospatial data base was formed, which has a rather complex structure. The database includes raster and vector data sets and describes the distribution of the main factors influencing climate change, administrative structure. The sequence of data processing includes primary data from international climate databases, secondary data from input processing, tertiary data that are summarized on the basis of the administrative system, and normalized data that are used in subsequent calculations [4].

The result of using the geographic information system is the development of mathematically based proposals for the formation of a climate change adaptation strategy in the Transcarpathian region of Ukraine for the time horizon until 2050, which is based on the use of a set of analytical tools and geospatial databases.

Conclusion. According to the results of the research, an applied geospatial database for analyzing the impact of climate change on the development of all components of sustainable regional development was developed with the formation of relevant recommendations for mitigating and preventing negative consequences in the foreseeable future.

References:

1. Kingston, R., Cavan, G., Butlin, T., Gill, S., & Lindley, S. (2014). Web-GIS tools for climate change adaptation planning in cities. In *Handbook of Climate Change Adaptation: Climate Change Adaptation Technologies* Springer Nature
2. Hawchar Lara, Naughton Owen, Nolan Paul, Mark G. Stewart Mark G., Paraic C. Ryan Paraic C. (2020) A GIS-based framework for high-level climate change risk assessment of critical infrastructure. *Climate Risk Management* Volume 29, 100235
3. Cavan G, Kingston R (2012) Development of a climate change risk and vulnerability assessment tool for urban areas. *Int J Disaster Resilience Built Environ* 3(3):253–269
4. Putrenko, V.V., Pashynska, N.M. (2017). The use of remote sensing data for modeling air quality in the cities. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(5W1), pp. 57–62

Oleksandra Pryshchepa

2nd year master's student in project management

Lyubov Kubiavka

Ph.D., Associate Professor, Department of Management Technologies

Faculty of Information Technologies, Taras Shevchenko National University of Kyiv

INSURANCE BROKER ASSISTANT SYSTEM (IBAS): FROM IDEA TO IMPLEMENTATION

Abstract. This paper presents the process of creating the Insurance Broker Assistant System (IBAS) information system — from the formation of the idea to the development of the software solution. The project aims to automate the process of concluding insurance contracts by brokers and agents. Based on an analysis of the problems of the modern insurance market in Ukraine, regulatory requirements, and digitalization trends, the need to create a web-based system to improve the efficiency of insurance intermediaries is justified.

Keywords: IBAS, insurance intermediaries, information system, automation, insurance, digitalization, design.

The insurance services market in Ukraine is undergoing active transformation, caused by a simultaneous reduction in the number of insurance companies, the growing role of the National Bank of Ukraine as a regulator, and increased demand for digital services. According to the NBU, in 2025 there will be about 60 insurance companies and more than 19,000 insurance intermediaries operating in the market. In a competitive environment and with constant changes in legislation, brokers and agents need digital tools that ensure fast and error-free contract processing.

The relevance of the Insurance Broker Assistant System (IBAS) project lies in the creation of a unified information system that automates the process of concluding insurance contracts and managing customer data. The system is designed to eliminate duplication of information, reduce the number of manual operations, increase data accuracy, and ensure compliance with legal requirements.

The purpose of the study is to justify the concept and develop the project basis for the IBAS information system, which will automate the activities of insurance intermediaries and become the basis for further implementation.

The idea of creating IBAS arose as a response to the key problems faced by insurance intermediaries:

- repeated manual data entry into documents;
- lack of a centralized database of contracts;
- difficulty in adapting to changes in legislation;
- inefficient control of renewals and commissions;
- lack of integration with state registries.

A PEST analysis was conducted during the project justification stage. The results of the SWOT analysis showed that the key opportunities for innovation are the development of insurtech solutions, the introduction of microservice architectures, and increasing customer confidence in electronic insurance formats.

To structure the objectives, a problem tree and an objective tree were constructed, which made it possible to determine the directions of development.

The system was designed according to the principles of microservice architecture, which ensures scalability, flexibility, and reliability. The implementation involves the use of the following stack: Angular (frontend), .NET/Node.js (backend), PostgreSQL (DB), Azure Cloud with CI/CD support, and logging systems.

The initial work on the implementation of IBAS includes the creation of an MVP (minimum viable product) version, which consists of basic modules: user registration, contract template management, and a bonus calculator. Within this stage, the main non-functional requirements have also been defined: security (data encryption), scalability, user-friendly interface, adaptability to different devices, and reliability of the cloud infrastructure.

The expected results of the initial implementation are:

- a 70% reduction in the average time required to draw up a contract;
- automatic creation of supporting documents (invoices, agreements, attachments);
- centralization of customer and contract data;
- transparency of commission accounting and reporting;
- the possibility of integration with state registries and insurance company APIs.

Thus, IBAS is an example of the practical implementation of a modern approach to the digitalization of the insurance business in Ukraine. At the stage from idea to implementation, a complete concept was developed, goals, functional and technical requirements, an architectural model, and the mathematical basis of the system were defined.

The proposed solution reduces the burden on insurance intermediaries, minimizes errors in contracts, increases the transparency of financial transactions, and speeds up customer service. IBAS has the potential to become the basis for the digital transformation of brokerage companies and increase the competitiveness of the Ukrainian insurance market.

References:

1. National Bank of Ukraine. National Bank of Ukraine. URL: <https://bank.gov.ua>
2. Register of Insurance Intermediaries. <https://kis.bank.gov.ua/>. URL: <https://kis.bank.gov.ua/search-fu>
3. About insurance. Official web portal of the Ukrainian Parliament. URL: <https://zakon.rada.gov.ua/laws/show/1909-20#Text>

Mykhailo Rudyk

Second-year master's degree student

Yana Bielozerova

Scientific supervisor, Ph.D., associate professor

State University "Kyiv Aviation Institute"

ARCHITECTURAL STRATEGIES AND TECHNICAL CHALLENGES OF DATABASE SHARDING IN HIGH-LOAD SYSTEMS

Abstract. This article explores the architectural approaches, strategies and technical challenges of database sharding in high-load information systems. The advantages and limitations of primary data distribution strategies, the specifics of selecting a sharding key and the impact of architectural decisions on consistency, transactional integrity and operational management are analyzed. Performance modeling of three typical sharding strategies was conducted, the results of which allow for the formulation of practical recommendations for selecting an architectural approach. It is shown that the effectiveness of sharding is determined by the balance between performance, availability, and administrative complexity.

Keywords: database, sharding, horizontal scaling, distributed systems, performance, architecture.

Main content

Modern information systems operate in an environment of rapid growth in data volumes and user counts. Traditional vertical scaling of databases, which involves increasing the resources of a single server, is gradually losing its effectiveness due to physical and economic limitations. In such conditions, sharding - a technology for horizontally distributing data across independent nodes - has become one of the most effective solutions. Each shard stores a portion of the information and processes queries autonomously, ensuring stable response times and system fault tolerance. Sharding architecture implements the shared-nothing principle [1], meaning complete resource isolation between nodes, and is used in most modern high-load platforms such as Google, Facebook, Amazon, Vitess, or Citus Data.

The main idea behind sharding is to increase query processing efficiency and ensure continuous scalability [2]. By distributing data across multiple servers, the load on a single node is reduced, which positively impacts performance. Additionally, such an architecture improves reliability - the failure of one shard does not halt the entire system but only limits access to a portion of the data.

Despite the clear advantages, implementing sharding creates a number of architectural and technical challenges that complicate database design and administration. The most important among these is the choice of a data distribution strategy [3]. The three most common approaches are range-based, hash-based, and directory-based.

In the first case, data is distributed based on key value ranges (e.g., user ID or transaction date). This approach is notable for its simple implementation and efficiency when executing queries within a specific range, but it is prone to creating "hotspots" when the latest shards become overloaded with new data.

Hash-based sharding ensures a uniform distribution of records among nodes, which is particularly effective for high volumes of write operations, but it complicates the execution of aggregate and analytical queries. Directory-based sharding, which relies on a lookup table mapping keys to shards, offers high flexibility during data migration but creates the risk of a centralized point of failure.

An equally important factor is the choice of the sharding key. An ideal key is characterized by high cardinality, uniform distribution, and immutability over time. Keys with low variability (e.g., status or record type) lead to data concentration on specific shards, causing load imbalance and degradation of system performance [3].

Technical Aspects and Challenges

Beyond performance issues, sharding introduces a range of technical and organizational challenges related to data consistency, transactional integrity, and query routing.

In distributed systems, it is impossible to simultaneously ensure consistency, availability, and partition tolerance (CAP theorem) [4]. A strong consistency model, based on the synchronous two-phase commit protocol, guarantees data integrity but leads to increased latency and reduced availability. In contrast, eventual consistency involves asynchronous updating of shards, which provides high performance and flexibility.

The problem of cross-shard transactions is one of the most complex. Traditional protocols like the two-phase commit do not always scale well due to resource locking. An alternative is the "Saga" pattern, which implements a sequence of local transactions with compensating actions in case of failure. This approach ensures fault tolerance and simplifies the rollback of failed operations without global locking [5].

Modern NewSQL solutions (e.g., Google Spanner, CockroachDB) integrate distributed coordination with the ACID model, combining the advantages of traditional and distributed databases.

Another complex problem is query routing. In the absence of a sharding key, the query must be sent to all shards (a scatter query), which significantly reduces system performance. A similar situation arises when performing JOIN operations between tables on different shards, requiring additional data transfer over the network. This demands careful data schema design to minimize cross-shard operations.

From an operational perspective, sharding increases the complexity of system maintenance. The need for rebalancing, synchronous updating of table schemas, node monitoring, and backup procedures raises operational costs. Therefore, modern platforms like Vitess, Citus, and YugabyteDB are actively implementing mechanisms to automate sharding processes, query routing, and system health monitoring.

Research Results

For the practical analysis, the operation of a sharded system based on PostgreSQL was modeled using the Citus extension [6], which provides logical data

sharding. A test database with 10 million transactional records was created, distributed across six logical shards. The goal of the experiment was to evaluate the effectiveness of three sharding strategies in terms of query response time, load uniformity, and system stability during rebalancing.

The results showed that hash-based sharding provides the most uniform data distribution and stable performance under high loads. However, in queries that span a large number of records (e.g., analytical queries), a 25-30% increase in response time was observed due to the need to access all shards.

Range-based sharding proved more effective for time-series or sequential data but led to the formation of “hot shards,” which accounted for over 40% of queries.

The directory-based strategy demonstrated the highest flexibility during data migration between nodes but required additional mechanisms for replicating the central directory.

The analysis also confirmed the critical impact of the sharding key choice: when using attributes with low cardinality, system performance decreased by 20-35% due to uneven load distribution. In contrast, keys with unique or nearly unique values ensured stable distribution and minimal latency.

The results obtained indicate that the effectiveness of sharding is determined not only by the architectural strategy but also by the nature of the data, the frequency of updates, and the structure of the queries. This points to the need for an adaptive approach to selecting a distribution scheme that considers the specifics of the particular system.

Conclusions

Database sharding is a key tool for ensuring the scalability and performance of modern high-load systems. Within this work, architectural strategies were analyzed, and their effectiveness was modeled. The results showed that hash-based sharding provides the best load uniformity, range-based is optimal for time-series data, and directory-based is the most convenient for migrations. The effectiveness of sharding is determined by the correct choice of key, the balance between consistency and availability, and adaptation to the load profile. The further development of this technology is associated with process automation, the integration of intelligent rebalancing mechanisms, and the search for a compromise between ACID properties and the horizontal scalability of distributed systems.

References:

1. Sikha Bagui, Loi Tang Nguyen (2015). Database Sharding: To Provide Fault Tolerance and Scalability of Big Data on the Cloud.
2. Srinivasan Jayaraman (2024). Efficient Data Sharding Techniques for High-Scalability Applications.
3. Siamak Solat (2024). Sharding Distributed Databases: A Critical Review.
4. Alessandro Margara et al. (2022). A Model and Survey of Distributed Data-Intensive Systems.
5. Erfan Zamanian et al. (2016). The End of a Myth: Distributed Transactions Can Scale.
6. Citus Data. Citus Data Documentation, Stable version.

Danylo Rybachok

Bachelor in Telecommunications and Radio Engineering, Tactical Level Officer,
Master Student

Andrii Yarmilko

PhD in Information Technologies, Docent, Associate Professor at Department of
Automated Systems Software

Bohdan Khmelnytskyi National University of Cherkasy

ARCHITECTURAL PARADIGMS FOR CONTESTED ENVIRONMENTS: A COMPARATIVE ANALYSIS OF MONOLITHIC AND MICROSERVICE DESIGNS FOR MODERN MILITARY LOGISTICS

Abstract. This paper provides a comparative analysis of monolithic and microservice architectures for tactical military logistics systems. It examines architectural suitability in Disconnected, Intermittent, Low-Bandwidth (DIL) conditions and explores advanced resilience patterns such as Domain-Driven Design, Saga, and CQRS. The study proposes a resilient microservice architecture with an AI Orchestrator, aligning with the JADC2 concept, to enhance decision advantage, adaptability, and operational tempo at the tactical edge.

Keywords: military logistics, microservice architecture, monolithic architecture, contested environments, JADC2, Domain-Driven Design, system resilience.

The selection of a software architecture is a foundational decision that dictates a system's capacity for resilience, adaptation, and scalability. In the context of tactical military logistics, these are not abstract qualities but direct determinants of operational success.

A monolithic architecture is the traditional model for building applications, where all functionality is developed, deployed, and scaled as a single, tightly coupled unit [1]. In a logistics context, such a system would be a single, large codebase responsible for order intake, inventory management, vehicle routing, and real-time tracking. While this approach can simplify initial development and deployment, its inherent limitations become critical liabilities in a dynamic military environment [1]. Most critically, it represents a single point of failure; a bug in a non-essential module could potentially crash the entire logistics planning engine, halting all operations.

In response to these challenges, the microservice architectural style emerged. It structures an application as a collection of small, autonomous, and loosely coupled services, each responsible for a distinct business capability [2]. These services are developed, deployed, and scaled independently, organized around business functions rather than technical layers. Key patterns include "Database per Service," where each service manages its own data to ensure isolation, and asynchronous communication via message brokers to decouple services and enhance resilience in environments with intermittent connectivity.

However, the transition to microservices introduces significant complexity inherent in any distributed system. This is evidenced by the experiences of companies like Amazon Prime Video and Segment, which reverted to monolithic architectures after facing challenges with operational overhead, performance bottlenecks, and unmanageable complexity [3]. These cases highlight that a "Monolith First" strategy

can be a pragmatic approach, allowing a team to understand domain boundaries before undertaking a complex and risky decomposition.

The choice of architecture is a fundamental sociotechnical decision that dictates organizational structure, aligning with Conway's law, which states that organizations design systems that mirror their communication structures.

In the military context, the Armed Forces of Ukraine are actively modernizing by adopting NATO-standard enterprise systems like SAP and LOGFAS. While powerful at the strategic level, these systems are not designed for the granular, real-time challenges of the tactical echelon, creating a "strategic-tactical gap". The tactical environment is characterized by Disconnected, Intermittent, and Low-Bandwidth (DIL) conditions, where network connectivity is a contested resource, not a guarantee [4]. In such environments, a naively implemented microservice architecture reliant on synchronous communication would suffer catastrophic failures. A monolith, with its reliable in-process calls, offers superior autonomy when disconnected. Therefore, microservices for tactical use must be designed for graceful degradation and operational autonomy, a principle the monolith cannot support.

From a security perspective, microservices create an "explosion of the attack surface," with each API endpoint becoming a potential vulnerability. The internal "east-west" traffic between services is a primary concern, often being unencrypted and unmonitored by default. However, this forces the adoption of a more robust, Zero Trust security model, which is operationally superior in a contested cyber environment where perimeter breaches are a given.

To build a resilient microservice system for this environment, advanced engineering patterns are not optional but mandatory. Domain-Driven Design (DDD) provides a methodology for defining logical service boundaries through its concept of the Bounded Context, ensuring services are aligned with business capabilities [5]. To maintain data consistency across these distributed services without resorting to fragile distributed transactions, the Saga pattern is employed. A saga is a sequence of local transactions coordinated through either choreography (event-driven) or orchestration (centrally managed). The Orchestration model aligns perfectly with the proposed AI Orchestrator concept. To ensure the reliability of these sagas, the Transactional Outbox pattern is used to guarantee that a business event is published if and only if the corresponding database transaction commits, preventing data loss in unreliable networks.

For performance and segregation, the Command Query Responsibility Segregation (CQRS) pattern separates the models for updating data (Commands) from the models for reading data (Queries), allowing each to be optimized and scaled independently. This is often paired with Event Sourcing, where the state of an entity is stored as an immutable sequence of events, providing a perfect audit trail and a robust foundation for the write-side of a CQRS system. To handle network unreliability, the Circuit Breaker pattern prevents cascading failures by temporarily blocking calls to a failing service, allowing it time to recover.

The synergy of these patterns culminates in an architecture centered on an AI Orchestrator. This component acts as the cognitive core of the system, implementing

the Saga Orchestration pattern to manage complex workflows. It leverages specialized microservices for predictive demand forecasting, dynamic vehicle routing optimization (VRP) using techniques like Reinforcement Learning, and management of autonomous assets like delivery drones. This architecture is a tactical embodiment of the broader Joint All-Domain Command and Control (JADC2) concept, which aims to accelerate the "sense, make sense, act" decision cycle by integrating sensors and actors across all domains [6].

Future development trajectories should explore the integration of Digital Twins – high-fidelity virtual replicas of the logistics network – to enable advanced simulation, validation, and wargaming of logistical plans. This creates a continuous learning loop where the AI Orchestrator can test strategies in a virtual environment before execution and refine its models based on real-world outcomes. Furthermore, the adoption of such agile, evolving systems necessitates a shift in doctrinal frameworks like the NATO Architecture Framework (NAF) and defense procurement models, moving away from rigid, long-cycle processes toward continuous certification models like DevSecOps and Continuous Authority to Operate (C-ATO).

In conclusion, while a monolithic architecture has its place, the extreme demands of the modern battlefield necessitate a resilient, adaptable, and intelligent system. A well-engineered microservice architecture, built upon a foundation of advanced design patterns and centered on an AI Orchestrator, provides a viable and superior path to bridging the strategic-tactical gap and achieving decision advantage in military logistics.

References:

1. Atlassian, Microservices vs. monolithic architecture. URL: <https://www.atlassian.com/microservices/microservices-architecture/microservices-vs-monolith>.
2. M. Fowler, Microservices. URL: <https://martinfowler.com/articles/microservices.html>.
3. P. Verma, When Monoliths Triumph: Case Studies of Reversing Microservices.... URL: <https://medium.com/@princevermasrcc/when-monoliths-triumph-case-studies-of-reversing-microservices-adoption-6f0bb3fd055e>.
4. REDCOM, Communications Challenges in DIL. URL: <https://www.redcom.com/dil-environments/>.
5. M. Fowler, Bounded Context. URL: <https://martinfowler.com/bliki/BoundedContext.html>.
6. DoD, Summary of the Joint All-Domain Command and Control Strategy, 2022. URL: <https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.pdf>.

Bohdan Semonov

PhD Student at the Department of Computer Engineering, Faculty of Radiophysics, Electronics and Computer Systems

Sergiy Pogorilyy

Doctor of Technical Sciences, Professor at the Department of Computer Engineering, Faculty of Radiophysics, Electronics and Computer Systems

Taras Shevchenko National University of Kyiv

SEMANTIC ENRICHMENT OF SOURCE CODE IDENTIFIERS THROUGH AUTOMATIC ABBREVIATION EXPANSION FOR COMMIT MESSAGE GENERATION

Abstract. Automatic generation of commit messages for source code modifications is an important task in software engineering, aimed at improving documentation quality and facilitating software maintenance. This paper presents an approach to semantic enrichment of identifiers in source code through automatic abbreviation expansion, which enhances the quality of input data for generation models.

Keywords: BERTScore, CodeT5, abbreviation expansion, commit message generation, semantic enrichment, semantic evaluation, source code, transformer models.

1. Introduction

Although commit messages play a crucial role in documenting code changes, there is currently no unified standard for how they should be written. As a result, developers use a wide range of approaches, which often leads to inconsistencies and reduces the clarity and completeness of the recorded information. Therefore, creating automated approaches for producing clear and standardized commit messages is an important task. Transformer-based models represent a promising direction for this, as they have demonstrated strong performance in applying natural language processing methods to code-related problems [1].

Guidelines for presenting information in commit messages [2]:

1. commit messages should specify the reasons for the changes, describe their expected impact, and, in the case of a bug fix, indicate the specific issue that was resolved;
2. commit messages must be understandable and convey the changes' effects, regardless of the reader's familiarity with the code or problem.

Examples of commit messages that comply with Rule 2: "Add JWT-based session handling to replace deprecated cookie auth"; "Fix division by zero in invoice calculation to prevent crash".

The main objective of this study is to develop and experimentally validate an approach for automated generation of code change messages using transformer-based models. An important part of this approach involves semantic enrichment of identifiers in source code by automatically expanding abbreviations, thereby providing additional contextual information that can improve the informativeness of the generated messages. To formally describe the proposed approach, let the corpus of examples be: $D = \{(x_i, y_i)\}_{i=1}^N$, where x_i – the textual representation of changes in the source code;

y_i – the corresponding commit message.

Identifiers in the source code are represented as a set of tokens: $x_i = \{t_1, t_2, \dots, t_m\}, t_j \in V$, where V – a vocabulary that includes abbreviations, acronyms, reserved words, and so on.

For the semantic enrichment of identifiers, a mapping is introduced: $E: V \rightarrow V'$, where $E(t_j)$ – the expanded form of the token.

Thus, the enriched input text: $x'_i = \{E(t_1), E(t_2), \dots, E(t_m)\}$.

A transformer model f_θ with parameters θ is used as the generation model, approximating the conditional distribution: $f_\theta: x'_i \rightarrow \hat{y}_i$, where $\hat{y}_i$ – the generated description of the changes.

The model is trained by minimizing the cross-entropy loss:

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^{|y_i|} \log P_\theta(y_{i,t} | y_{i,<t}, x'_i), \quad (1)$$

Where $y_{i,t}$ – the t^{th} token in the target message.

The quality of the results is evaluated using metrics $M: Q = M(\hat{y}_i, y_i)$, and the overall metric value is calculated as the average:

$$\bar{Q} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} M(\hat{y}_i, y_i). \quad (2)$$

Thus, the objective of the study is to find such θ^* that:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta), \quad (3)$$

incorporating the semantic enrichment of identifiers E , so that $\bar{Q}$ is maximized.

2. Data Collection and Preprocessing

To construct the training dataset, one of the largest web services for hosting IT projects and collaborative software development – GitHub – was utilized. This platform provides a dedicated application programming interface known as the GitHub REST API [3].

After data acquisition, the records were preliminarily filtered, as the change descriptions varied in style, length, and level of detail. All previously defined rules can be implemented programmatically, except for Rule 4, for which formal methods are insufficient. Due to the absence of clear structured criteria for evaluating the semantic content of changes, fulfilling this rule requires the use of machine learning methods capable of modeling the contextual semantics of commit messages. To address this, a filtering method for change descriptions based on MLP and RNN models was developed [4].

3. Method of Semantic Enrichment of Identifiers in Source Code

To address the task of generating commit messages, this study employs a single transformer-based model – Salesforce/codet5p-770m – which is specifically adapted for code-related tasks as part of the CodeT5+ family. The model was fine-tuned on a precompiled corpus of “change–message” pairs covering several programming languages (C/C++, Ruby, Python, Java, JavaScript, PHP), using its

native tokenizer (AutoTokenizer from the Hugging Face Transformers library) during fine-tuning. The Hugging Face Transformers framework was chosen due to the availability of models pre-trained on source code and the ability to efficiently run fine-tuning and inference workflows in the Google Colab environment.

However, one factor that reduces the quality of automatically generated commit messages is the low level of semantic informativeness of identifiers in the source code: many identifiers consist of abbreviations or acronyms meaningful only to their authors and insufficiently expressive in terms of semantic content. As a result, the message generation system may lack essential semantic cues, which can lead to overly general or inaccurate descriptions of changes.

To address this issue, we propose a semantic enrichment pipeline for identifiers through the automatic expansion of abbreviations. Its integration into the code preprocessing stage prior to message generation enhances the expressiveness and informativeness of textual representations fed into natural language models [5].

The proposed method consists of the following stages (see Figure 1):

1. identifier extraction: identifiers are retrieved from modified code fragments through the construction of an abstract syntax tree (using the Python programming language via the ast module);
2. segmentation: identifiers are decomposed into constituent lexemes according to their naming conventions (snake_case, camelCase, PascalCase);
3. abbreviation expansion: for each extracted lexeme, potential expansions are generated based on:
 - a. dictionary-based abbreviation – full form pairs;
 - b. the contextual CodeBERT model [6], which can identify the most probable correspondences;
4. candidate selection: semantic ranking methods (cosine similarity in vector spaces) are applied, along with consistency checks against the context of code changes;
5. formation of semantically enriched representations: the expanded identifiers are integrated into the input data for message generation models.

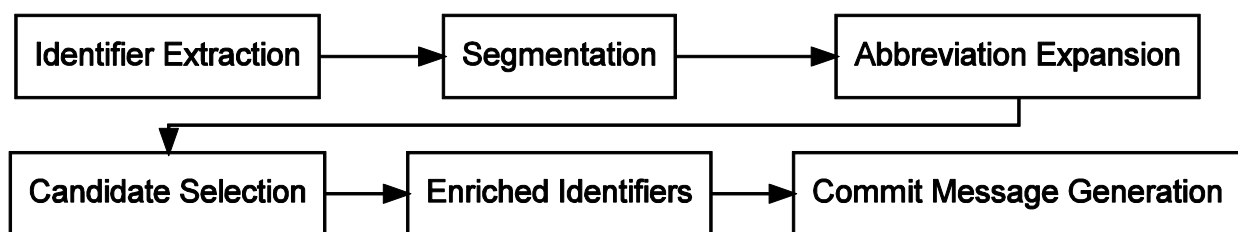


Figure 1 – A schematic representation of the method for semantic enrichment

This enables the models to produce more specific and comprehensible descriptions, as they gain access to terms closer to natural language.

Thus, the automatic expansion of abbreviations enhances the quality of the semantic space in which commit message generation is performed. The application of this method helps reduce the number of incorrect or overly general descriptions and improves the comprehensibility of the development history for different team members. The results of the experimental evaluation indicate that the integration of the proposed pipeline leads to improvements in the main quality metrics, as presented in Table 1.

Table 1 – Effect of semantic enrichment on model performance

Prog. lang. Metrics	C/C++		Ruby		Python		Java		JavaScript		PHP	
	Base-line	Sem. En-rich.	Base-line	Sem. En-rich.	Base-line	Sem. En-rich.	Base-line	Sem. En-rich.	Base-line	Sem. En-rich.	Base-line	Sem. En-rich.
BLEU	3.92	5.47	6.16	8.55	6.22	9.1	5.32	8.06	14.27	16.47	4.3	6.29
ROUGE-1	24.34	29.72	26.08	30.17	27.87	32.27	28.17	33.51	35.64	41.59	24.64	28.9
ROUGE-2	12.36	18.37	11.79	15.84	15.04	20.94	18.03	21.13	25.53	27.72	9.09	13.37
ROUGE-L	22.54	27.33	23.67	29.2	25.8	31.47	26.59	32.78	34.01	40.82	23	27.5
BERTScore	0.88	0.91	0.86	0.89	0.87	0.9	0.88	0.91	0.9	0.92	0.87	0.9
BLEURT	0.35	0.36	0.35	0.36	0.36	0.37	0.36	0.36	0.4	0.42	0.32	0.34

4. Conclusion

In this study, the transformer-based model CodeT5+ (770M) was applied to the task of automatic commit message generation. The model was fine-tuned on a domain-specific corpus of code changes, and both n-gram-based and semantic metrics were used for evaluation. To enhance the informativeness of the input, semantic enrichment of identifiers was performed through automatic abbreviation expansion.

Experimental evaluation demonstrated that the integration of the proposed pipeline led to measurable improvements in quality metrics: BLEU increased by 2–3 percentage points, ROUGE by 4–7 percentage points, BERTScore by approximately 0.03, and BLEURT by 0.01–0.02.

These results confirm the effectiveness of combining transformer-based generation with identifier preprocessing to improve the clarity and relevance of commit messages.

References:

1. Semonov, B. O., Pogorilyy, S. D. Research of the application of GPGPU and TPU technologies for ensuring comment quality in version control systems. PROBLEMS IN PROGRAMMING. National Academy of Sciences of Ukraine (Co. LTD Ukrinformnauka) (Publications). doi:10.15407/pp2025.01.024.
2. How to Write a Git Commit Message. cbeams. URL: <https://cbea.ms/git-commit/#seven-rules>.
3. GitHub Docs. GitHub REST API. GitHub Docs. URL: <https://docs.github.com/en/rest?apiVersion=2022-11-28>.
4. Pogorilyy S., Semonov B. The Implementation of a Commit Messages Filter for Software Version Control Systems. The 9th International Conference on Control and Optimization with Industrial Applications.
5. Newman, C., Decker, M., AlSuhaibani, R., et al. An Empirical Study of Abbreviations and Expansions in Software Artifacts. doi:10.1109/ICSME.2019.00040.
6. Feng, Z., Guo, D., Tang, D., et al. CodeBERT: A Pre-Trained Model for Programming and Natural Languages. URL: <https://arxiv.org/abs/2002.08155>.

Oleksandr Timinskyi

PhD, associated professor of the Department of Management Technologies

Denys Liashenko

PhD student of the Department of Management Technologies

Taras Shevchenko National University of Kyiv

ANALYSIS OF THE DEVELOPMENT DIRECTIONS OF COMPUTER SCIENCE MODELS IN THE FIELD OF IT PROJECT MANAGEMENT

Abstract. This paper examines the impact of the development of computer science models and methods on the formation of modern concepts of IT project management. The role of systems analysis, data analytics, artificial intelligence, cognitive technologies, and risk theory in transforming approaches to planning, control, and decision-making in IT projects is analyzed. It is shown that the integration of these domains has led to the creation of adaptive, analytical, and risk-oriented models of IT project management. The paper also outlines directions for further research and formulates the main conclusions.

Keywords: computer science, IT project management, systems analysis, data analytics, risk-oriented approach.

Introduction. The field of IT project management is inseparably connected with the development of computer science and the models and methods that are used and continuously improved within it. The emergence and active advancement of such areas as systems analysis, artificial intelligence, machine learning, and data analytics have transformed approaches to planning, control, and decision-making in IT projects. Today, IT project management relies not only on managerial practices but also on intelligent models and corresponding algorithms.

The evolution of computer science contributes to the formation of new management concepts characterized by adaptability, data orientation, and self-improvement capability. Managerial processes are increasingly automated, while data analytics and risk prediction models have become integral parts of IT team operations. As a result, project management is becoming more precise, flexible, and risk-oriented.

The purpose of this publication is to determine the role of computer science in shaping modern concepts of IT project management and to outline the main directions of its influence on the development of relevant models and methods.

Evolution of Computer Science in the Context of IT Project Management.

The development of computer science has directly influenced the establishment of IT project management. In the 1940s–1950s, John von Neumann formulated the computer architecture, while Norbert Wiener, in his work *Cybernetics* (1948), described the principles of feedback that became the foundation of automated control systems. In the 1960s, Jay Forrester developed the concept of system dynamics, which laid the groundwork for modelling project management processes. During the same decade, the PERT (1958) and CPM (1959) methods emerged, initiating the formalization of project scheduling and resource planning.

In the 1970s–1980s, Management Information Systems (MIS) and Decision Support Systems (DSS) appeared, focused on data processing and managerial

decision support. Their evolution led to the emergence of corporate Enterprise Resource Planning (ERP) systems (SAP R/3, Oracle Applications, etc.) in the 1990s, which integrated planning, control, and reporting functions, significantly enhancing project management systems.

During this period, project management approaches were also standardized through the publication of the following standards: PMBOK (first edition – 1987; current – 2021), PRINCE2 (first edition – 1996; current – 2023), and ISO/IEC 21500 (first edition – 2012; current – 2021). These standards established and continue to advance the systems approach to project management, its models and methods, including project life cycle models. Since the early 2000s, the development of the Agile methodology (2001), its frameworks Scrum and Kanban, as well as the implementation of Data Science, Machine Learning, and ISO 31000, has formed the foundation for modern flexible, adaptive, and risk-oriented models of IT project management.

The Influence of Computer Science Development on Modern IT Project Management Concepts. The development of computer science has determined the direction of transformation in IT project management – from intuitive methods to formalized, algorithmically supported models. The formation of modern concepts has resulted from the integration of systems analysis, data analytics, artificial intelligence, cognitive technologies, and risk theory into planning, monitoring, and decision-making processes. As a result, IT project management has acquired the characteristics of an applied computer science discipline, in which managerial actions are based on data, forecasting, and the formalized evaluation and management of risks.

The systems approach laid the foundation for considering a project as a complex dynamic system with numerous interdependent elements. The concepts of systems thinking formulated by Norbert Wiener (Cybernetics, 1948) and Ludwig von Bertalanffy (General System Theory, 1968) provided the impetus for developing methods of goal decomposition, life cycle modelling, and feedback analysis. These principles are reflected in standards such as PMBOK (2021), ISO/IEC 21500:2021, and CMMI, which define management as a continuous process of collecting, analyzing, interpreting, and applying information to achieve defined objectives. At the practical level, systems analysis is implemented through business process modelling (in BPMN and UML notations), configuration management, hierarchical WBS structures, and Network Analysis methods (PERT, CPM).

The next stage of evolution is associated with the development of data analytics and Business Intelligence. The concept of data-driven management presupposes the use of data as the basis for managerial decision-making. The application of Big Data, Process Mining, Predictive Analytics, and BI systems (Power BI, Tableau, Qlik, etc.) has enabled the tracking of project performance indicators (KPI, SPI, CPI), the forecasting of delays, the identification of risks, and the optimization of resource allocation. Within digital management platforms (Jira, Asana, ClickUp, MS Project Online), these tools are integrated as part of the project's information and analytical framework. Thus, computer science has transformed traditional monitoring into a process of continuous real-time data analysis and utilization.

The integration of Artificial Intelligence (AI) and Machine Learning (ML) has initiated the concept of AI-assisted project management, within which systems are capable of automatically generating predictive scenarios, identifying critical tasks, and assessing the impact of deviations. Clustering models, neural networks, natural language processing (NLP), and reinforcement learning are used to automate risk management, task distribution, and release planning. Such capabilities are implemented in tools including Oracle AI for Project Management, Forecast, Jira Advanced Roadmaps, and the intelligent modules of SAP Project Intelligence Network. As a result, IT projects have acquired self-improving properties – the system's ability to adjust planning and resource allocation without direct managerial involvement.

The development of cognitive technologies and computational intelligence has contributed to the emergence of adaptive management – a concept that enables flexible responses to environmental changes. The use of fuzzy logic, evolutionary algorithms, and cognitive modelling makes it possible to account for environmental uncertainty and multifactorial complexity. In contemporary frameworks, such as the Scaled Agile Framework (SAFe) and Disciplined Agile Delivery (DAD), adaptability is combined with feedback data and analytical indicators, enabling the creation of controllable, self-learning systems.

A distinct direction of development has been the integration of risk and uncertainty theory into project management processes. Based on the ISO 31000:2018 and COSO ERM standards, risk-oriented management combines systems analysis, data analytics, and risk modelling. Quantitative models and methods such as the Monte Carlo method, Bayesian modelling, the Value-at-Risk (VaR) model, and scenario analysis make it possible to evaluate the influence of random factors and determine acceptable risk thresholds. The advancement of computational power and machine learning models has enabled automated risk forecasting, ranking, and the improvement of response plans. Consequently, the risk-oriented approach has evolved from an isolated management element into an integrated component of the project's informational ecosystem.

The generalised analysis of the development directions of computer science models and the key characteristics of their influence on the formation of modern IT project management concepts is presented in Table 1.

The comparison of computer science directions and management concepts demonstrates that modern IT project management is the result of a synthesis of methods from various scientific schools. While systems analysis ensures structural integrity, data analytics provides the justification of decisions, and artificial intelligence delivers automation, the risk-oriented approach adds a dimension of uncertainty, enhancing the scientific accuracy and adaptability of management systems to the real conditions of the digital society. Collectively, these directions form the methodological foundation of the current stage of IT management evolution – intellectual-analytical management – within which project activities are based on the integration of knowledge, data, and risk management models.

Table 1.

The Impact of Computer Science Models Development on the Formation of Modern IT Project Management Concepts

Direction of Computer Science	Key Methods or Technologies	Impact on Management Concepts	Examples of Implementation or Standards
Systems Analysis and Modelling	System Dynamics (Forrester), UML, BPMN, Network Analysis (PERT/CPM)	Formation of system-oriented management; decomposition, life cycle, feedback control	PMBOK (2021), ISO/IEC 21500:2021, CMMI
Data Science and Business Intelligence	Big Data, Predictive Analytics, Process Mining, BI Platforms	Data-driven management; performance analytics, forecasting of deadlines and risks	Power BI, Tableau, Jira Analytics
Artificial Intelligence and Machine Learning	Neural Networks, NLP, Clustering, Reinforcement Learning	AI-assisted management; automation of planning, forecasting, and performance evaluation	Oracle AI, Forecast, SAP Project Intelligence
Cognitive Technologies and Computational Intelligence	Fuzzy Logic, Evolutionary Algorithms, Cognitive Modelling	Adaptive management; self-learning and adaptation to environmental change	SAFe, DAD, Adaptive Project Control
Risk and Uncertainty Theory	Bayesian Modelling, Monte Carlo Simulation, Scenario Analysis, VaR	Risk-oriented management; quantitative risk assessment and decision optimization	ISO 31000:2018, COSO ERM, Quantitative Risk Tools

Conclusions. The development of computer science has become a key factor in the evolution of IT project management. The integration of models and methods from systems analysis, data analytics, artificial intelligence, cognitive technologies, and risk theory has facilitated the transition from traditional to analytical and adaptive management models. Modern approaches – system-oriented, data-driven, AI-assisted, and risk-oriented management – form the methodological foundation within which managerial decisions in projects are based on data and the modelling of uncertainties.

Further research should focus on formalization of models and methods of risk-oriented management in IT projects, the development of mathematical models for risk forecasting, and the creation of intelligent decision support systems that incorporate risk considerations within the chosen research domain.

References:

1. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition. USA, PMI, 2021, 250 p.
2. International Organization for Standardization, ISO 31000:2018 – Risk Management – Guidelines, Geneva, Switzerland: ISO, 2018. [Online]. Available: <https://www.iso.org/standard/65694.html>
3. International Organization for Standardization, ISO/IEC 21500:2021 – Project, Programme and Portfolio Management – Context and Concepts, Geneva, Switzerland: ISO, 2021. [Online]. Available: <https://www.iso.org/standard/75704.html>
4. Hillson D. Managing Risk in Projects, 2nd ed.: Routledge, 2024. – 170 p.
5. Ziuziun V., Timinskyi O., Kolomiiets A., Liashenko D., Yurechko A. Research of management models for commercial IT project development in a remote team environment. Management of Development of Complex Systems, 2025, Iss. 61, pp. 26–34, <https://doi.org/10.32347/2412-9933.2025.61.26-34>.

Oleksandr Timinskyi

PhD, associated professor of the Department of Management Technologies

Artem Yurechko

PhD student of the Department of Management Technologies

Taras Shevchenko National University of Kyiv

COMPARATIVE ANALYSIS OF IT PROJECT MANAGEMENT METHODOLOGIES IN THE DEVELOPMENT OF WEB- AND CONTENT-ORIENTED SYSTEMS

Abstract. This study examines traditional IT project management models, including life cycle models, as well as the Agile methodology and its frameworks, Scrum and Kanban. The advantages and disadvantages of both traditional and agile models are identified, along with the domains in which their application is recommended. A comparative analysis of project management approaches is conducted with a focus on their use in Headless CMS development projects. Hybrid management models suitable for Headless CMS development are characterized. The directions for further research are outlined, and conclusions are formulated.

Keywords: computer science, IT project management, life cycle models, Agile, Headless CMS, hybrid management approaches.

The development of the digital society and the emergence of architectures such as Headless CMS (a content management system that operates solely on the server side and functions as a content repository; a Headless CMS makes content accessible via an API for display on any device, without an integrated frontend or presentation layer) have necessitated new approaches to IT project management. Traditional methods based on sequential execution of project phases are increasingly giving way to agile approaches that ensure a high degree of responsiveness to changing business requirements [1].

The management of web and content-oriented system development requires the adaptation of methodologies to multi-channel ecosystems – ranging from traditional websites to mobile applications, APIs, digital signage, and smart TVs. Under these conditions, it is crucial to select an approach that enables effective team coordination, stable content integration, and rapid product scalability [2].

Traditional IT Project Management Models: Structure and Limitations.

Classical approaches to IT project management – particularly life cycle models such as Waterfall, Spiral, and the V-model – have historically formed the foundation of the systems engineering approach to software development [3]. Their core principle lies in the sequential execution of phases – from requirements analysis to design, development, testing, and implementation. This structure ensures a high level of process control, clear delineation of responsibilities, and predictable timelines. In many industries – including defense, aviation, and finance – where compliance with regulatory and certification standards is essential, these models remain relevant due to their formalization and rigorous documentation at every stage.

However, in the development of web- and content-oriented systems, which are characterized by rapidly changing business requirements, the Waterfall model

demonstrates significant limitations. Any modification introduced after the technical specification has been approved demands substantial time and resources, as the model does not provide for cyclic or iterative processes. Moreover, the lack of continuous user feedback often leads to situations where the final product may ultimately fail to meet real market needs [4].

In the context of Headless CMS development projects, where the frontend and backend are developed in parallel by different teams, the use of the Waterfall model introduces additional risks – such as asynchronous workflows, complex API updates, and limited capacity for quick integration of new content delivery channels. The Spiral and V-model partially mitigate these risks by introducing intermediate testing stages; however, they remain constrained in flexibility and responsiveness to change. These processes are also inconsistent with the CI/CD (Continuous Integration / Continuous Delivery) paradigm, which requires frequent, automated updates and releases [5].

Therefore, traditional life cycle models are advisable in IT project management only when product requirements are clearly defined, stable, and subject to strict control – for instance, in the development of system cores, base APIs, or security components, where flexibility is not critical and reliability and standards compliance are the primary priorities.

Agile, Scrum, and Kanban as the Foundation of Modern IT Project Management.

The Agile methodology and its numerous frameworks emerged as a response to the rigidity of traditional waterfall processes. Their primary goal is to create an environment in which teams can rapidly adapt to change, reduce risks, and maintain product quality throughout the IT project lifecycle [6].

The application of Agile to the management of Headless CMS development projects offers several clear advantages:

- parallel work of frontend and backend teams;
- continuous integration with APIs;
- rapid testing of modules and content components;
- the ability to incrementally deploy system versions.

The Scrum framework, derived from Agile methodology, structures the process through fixed iterations (sprints), daily meetings (daily stand-ups), and quality control mechanisms such as reviews and retrospectives [7]. This ensures transparency of the development process but requires a high level of team maturity and discipline.

The Kanban framework, which also falls under the Agile umbrella, by contrast, does not impose strict time constraints and focuses on a continuous flow of tasks. It is particularly suitable for maintenance activities, content integration, DevOps cycles, and small teams working on incremental improvements to a Headless CMS without large-scale planning [8].

For clarity, the main characteristics of the most common project management approaches are compared below (Table 1).

Table 1.

**Comparison of IT Project Management Approaches with Emphasis
on Their Application in Headless CMS Development Projects**

Parameter	Waterfall	Agile	Scrum	Kanban
Type of management	Sequential	Iterative	Iterative and regular	Continuous flow
Client feedback	After completion	At every stage	Within each sprint	Continuous
Flexibility to change	Low	High	High	Very high
Schedule control	Rigid	Adaptive	Strictly regulated	Flexible
Best suited conditions	Stable requirements, large-scale systems	Dynamic markets, rapid change	Products with shifting priorities	Continuous maintenance and updates
Suitability for Headless CMS development projects	Limited	High	High	High

Hybrid Management Models in Headless CMS Projects

Practical experience in IT project implementation demonstrates that the optimal management strategy for complex IT projects lies in the combination of different approaches, commonly referred to as hybrid management models [9]. For instance, strategic planning may follow the principles of the Waterfall model (defining project stages, budgets, and risks), while implementation may proceed in an Agile or Scrum format.

In Headless CMS development projects, such models enable coordination among multiple teams (responsible respectively for API development, content creation, UI/UX design, and integrations) operating on distinct work cycles. Components with fixed requirements (such as databases and API architecture) are implemented through sequential processes, while those with dynamic requirements (interfaces, content modules) follow flexible iterative cycles [10].

This combined approach enhances process coherence, ensures the stability of core system components, and at the same time maintains the agility required for rapid content updates – a critical factor for multichannel solutions.

As promising directions for further research in this field, the following should be identified:

- the formalization of Headless CMS development projects, including the definition of project parameters and analysis of their variability depending on the industry and application domain;

- the creation and formalization of hybrid management models for IT projects involving Headless CMS development.

Conclusions

The conducted analysis has demonstrated that traditional waterfall life cycle models of IT projects, despite their predictability, fail to provide the flexibility and responsiveness required for the development of web- and content-oriented systems. The Agile methodology and its frameworks, Scrum and Kanban, create an environment conducive to rapid adaptation, effective team collaboration, and continuous integration.

In the domain of Headless CMS development projects, the application of hybrid models – both for the project life cycle and overall project management – is advisable. Such models combine the structural clarity of the Waterfall approach with the adaptability of Agile methodologies. This integration reduces risks associated with architectural complexity while ensuring process stability and development predictability.

Therefore, the integration of Agile and traditional approaches represents a promising direction for enhancing the effectiveness of IT project management within multichannel content ecosystems.

References:

1. Kerzner H. Project Management: A Systems Approach to Planning, Scheduling, and Controlling. – Wiley, 2022.
2. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition. USA, PMI, 2021.
3. Sommerville I. Software Engineering. – Pearson Education, 2020.
4. Boehm B. Software Engineering Economics. – Prentice Hall, 1981.
5. Royce W. Managing the Development of Large Software Systems. – Proceedings of IEEE WESCON, 1970.
6. Beck K. et al. Manifesto for Agile Software Development. – Agile Alliance, 2001.
7. Schwaber K., Sutherland J. The Scrum Guide. – 2020.
8. Anderson D. Kanban: Successful Evolutionary Change for Your Technology Business. – Blue Hole Press, 2010.
9. Deloitte Insights. Future of Hybrid Project Management. – Deloitte, 2023.
10. ISO/IEC 21500:2021. Project, Programme and Portfolio Management – Context and Concepts. – ISO, 2021.

Diana Vashchilina

Bachelor Student, Department of Software Systems and Technologies

Olena Vashchilina

Phd, Associate Professor, Department of Applied Information Systems

Anastasiia Nikolaienko

Phd, Assistant Lecturer, Department of Software Systems and Technologies

Taras Shevchenko National University of Kyiv

STRUCTURING EDUCATIONAL IT PROJECTS WITH NPM WORKSPACES

Abstract. This paper presents the experience of using NPM Workspaces to implement a full-stack web application within an educational student project. The tool enabled the creation of a unified repository structure that includes a client-side interface, server-side logic, and a shared module containing TypeScript types. This approach ensured consistency across components, simplified configuration, and enabled automated testing, code style validation, and deployment. The project demonstrates the advantages of monorepository architecture in individual educational development, including rapid setup, flexible scalability, and code organization that supports effective learning and presentation in academic environments.

Keywords: full-stack web application, educational student project, monorepository, NPM Workspaces

Introduction. Developing a full-stack web application in an educational environment involves addressing a range of technical and organizational challenges. Key tasks include clearly separating the client-side and server-side logic, managing dependencies, reusing types, and ensuring consistency across modules. One of the modern solutions for optimizing the development process is organizing the source code according to the monorepository principle – an approach in which all application components are stored within a single repository. This enables efficient project structuring, centralized dependency management, avoidance of code duplication, and consistency across components [1, 2].

Monorepository Architecture. In the JavaScript development environment, a convenient way to implement a monorepository is through NPM Workspaces – a built-in Node.js feature that enables managing multiple packages within a single project. This mechanism is officially supported, does not require additional libraries, and integrates well with popular frameworks such as React, Express.js, and TypeScript [3].

As part of an individual educational project, a monorepository architecture based on NPM Workspaces was implemented [4]. Figure 1 presents the repository structure, which includes three main components:

- client – the client-side module (React), responsible for user interaction, routing, interface rendering, and handling requests to the server.
- server – the server-side module (Express.js), which implements business logic, processes HTTP requests, manages sessions, and provides access to the database.

- shared – the shared module containing TypeScript types, utilities, constants, API interfaces, and functions used by both the client and server components.

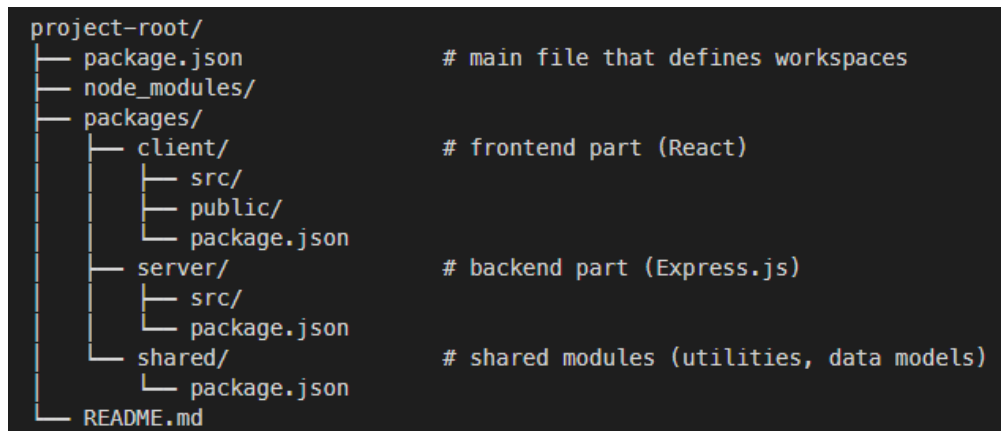


Figure 1 – Repository structure based on NPM Workspaces

A key element of the project is the `package.json` file located in the root directory, which declares the workspaces. This enables centralized management of all modules: installing shared dependencies, running scripts, and executing tests without switching between folders. The shared module ensures consistency across application parts, reduces errors, and improves implementation quality. It serves as a central point for reusable code shared between the frontend and backend, implemented according to principles of contract-based programming and type safety. It includes domain-specific types and DTOs (places, tags, tours, trips, users, auth), enums for statuses, API paths, HTTP codes, and messages, unified error classes, and Zod validation schemas used on both client and server sides.

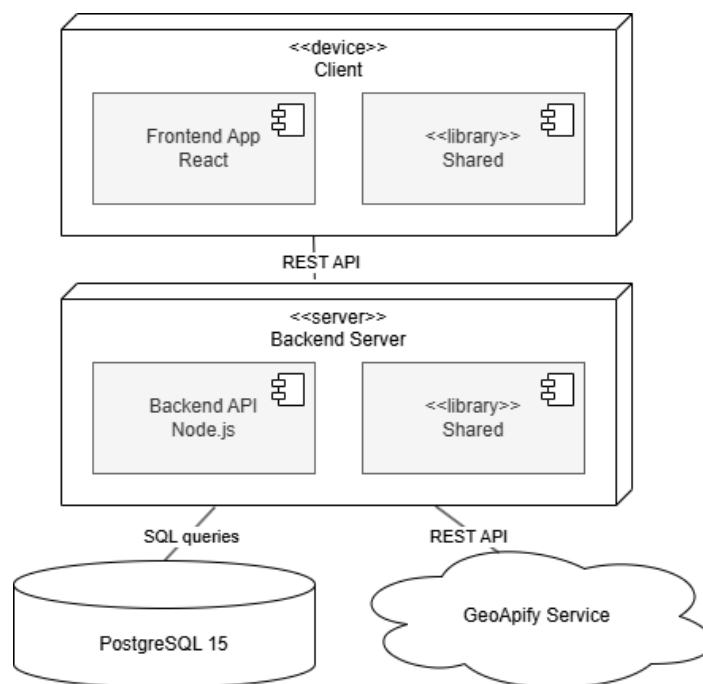


Figure 2 – Deployment diagram of the educational web application

Deployment Architecture. To generalize the architectural logic of the educational application, Figure 2 presents a deployment diagram that illustrates the interaction between the system's core components. The diagram reflects the deployment architecture of an application built according to monorepository principles. The client-side (React) and server-side (Node.js) components both utilize a shared library of types and utilities, ensuring logical consistency. The server interacts with a PostgreSQL 15 database via SQL queries and with the external GeoApify REST API for geolocation services. Communication between the client and server is handled through HTTP requests, following REST architecture principles. This structure demonstrates end-to-end logic, component isolation, and alignment with modern full-stack development practices [5].

Conclusions. The use of a monorepository architecture based on NPM Workspaces in an educational student project enabled efficient organization of the web application's structure by integrating the client-side interface, server-side logic, and shared resources within a single repository. This approach ensured type consistency, centralized dependency management, and code reuse, contributing to reduced error rates and simplified configuration. The implementation of end-to-end logic, adherence to REST architecture, and integration with external services reflect alignment with modern full-stack development practices. The project structure is well-suited for demonstration, validation, and future scalability, making it appropriate for individual educational use and presentation in academic settings.

References:

1. Prus O.V., Maidaniuk V.P. Analysis of the main principles of working with monorepositories: features, advantages, and disadvantages. Information Technologies and Automation – 2023: Proceedings of the 16th International Scientific and Practical Conference. 2023. P. 267–270.
2. Google Open Source. Monorepo: Advantages and Challenges: website. URL: <https://opensource.google/documentation/reference/repository-structure/>
3. Node.js Documentation. Using NPM Workspaces: website. URL: <https://docs.npmjs.com/cli/v7/using-npm/workspaces/?v=true>
4. Vashchilina D.O., Nikolaienko A.Yu. Design and development of an interactive web application architecture for personalized excursions. Mathematical and Software Technologies of the Internet of Everything: Proceedings of the 12th Eastern European Conference MSTIoE (April 08–09, 2025, Kyiv). Kyiv: Taras Shevchenko National University of Kyiv, 2025. P. 77–79.
5. Fowler M. Patterns of Enterprise Application Architecture. Boston: Addison-Wesley, 2003. 533 p.

Oleksii Vedmiediev

PhD student of the Project Management Department

Kyiv National University of Construction and Architecture

ASPECTS OF PROACTIVE ORGANIZATIONAL DEVELOPMENT OF CONSTRUCTION PROJECT TEAMS WHEN IMPLEMENTING VALUE HARMONIZATION WITH THE IT CONTEXT

Abstract. The subject area of construction project management is considered. The approaches of value-based management, project team management and stakeholder management are analyzed. A method of proactive organizational development of construction project teams when implementing value harmonization is proposed. Its implementation on the IT platform of a construction company is recommended. The method contains three main stages: identification and mapping of stakeholder values; formation of a project team with an orientation towards value harmonization; monitoring, adjustment and support of value harmonization during the project implementation process. Directions for further research are considered, conclusions are formulated.

Keywords: project management, IT platform, project team, stakeholders, value harmonization.

Introduction. In the modern construction industry, projects are increasingly characterized by increased complexity (including in the IT part), multiplicity and multi-vectority (consisting of different values) of stakeholders and significant dynamism of the external environment. In such conditions, the success of the implementation of a construction project is no longer limited to meeting deadlines, budgets and quality standards; it is increasingly important to create value for different groups of stakeholders, as well as to ensure the consistency of their expectations and values. By value harmonization we mean the process in which key stakeholders of the project (customer, contractors, users of the construction object being created, local community) achieve a certain level of consistency in their values – that is, they mutually share or accept common value guidelines that determine what is meaningful and useful within the project.

At the same time, the project team, as an organizational unit, plays a crucial role in translating these values into specific actions, processes and behavior. Proactive organizational development of the team means that the team does not simply react to external influences, but independently initiates changes, adapts to the value environment, creates conditions for its internal transformation in the direction of supporting value harmonization. An obligatory component of any modern project is the IT component, and therefore, to implement innovative approaches and methods, appropriate IT artifacts of the construction project must be created.

The purpose of this publication is to highlight the developed method of proactive organizational development of construction project teams, which is aimed at implementing value harmonization between key stakeholders through team development. Methodologically, it is based on research on value-based project management, stakeholder management theory and organizational team development.

A value-based approach in construction project management. The traditional approach in construction focused on the “three-in-one” criterion of success: time, cost, quality. However, in modern realities, more attention is paid to

“value-based” approaches, i.e. creating value for the customer and other stakeholders. For example, the study [1] noted that the focus should shift from simply reducing costs to ensuring the best value for the entire life cycle of the facility. Ukrainian studies also emphasized that project scope management in construction should take into account the value components of stakeholders’ needs and expectations, which can become the basis of a new methodological component of scope management [2].

Stakeholder management in construction projects. Effective stakeholder management is considered a critical condition for the success of a construction project: balancing the interests of different groups, timely involvement, communication, monitoring expectations. For example, in [3] highlights that over the past two decades, the number of studies focused on stakeholder relations in construction has been increasing, and that this topic is multidimensional. Another study proves that stakeholder management directly affects the creation of project value through such parameters as cost optimization, time savings, improvement of the quality of project management and product, environmental and social benefits [4].

Organizational development of the project team. A project team is not just a set of people performing tasks, but a social system with its own norms, values, culture, and interaction processes. Organizational development of the team includes motivation, training, communication, leadership, and effective interaction. In the context of implementing value harmonization, it is important that the team is able to not only perform tasks, but also understand and support the values of stakeholders, participate in their coordination and transmission.

Thus, the combination of three components – value-based approach, stakeholder management and team development – creates a theoretical basis for developing a method of proactive organizational development of teams in construction projects.

Description of the method. The proposed method is based on further development of research [5, 6] and consists of three main stages, which are sequentially and logically connected, and ensure proactivity in team development and the implementation of value harmonization. IT support for the implementation of such a method can be the ERP system of a construction company.

Stage 1. Identification and mapping of stakeholder values.

This stage involves:

- Identification of key project stakeholders (e.g., customer, main contractor, users of the construction object being created, local community, etc.).
- Conducting surveys, facilitation sessions or interviews to identify their values: what they seek to gain from the project (for example: minimal costs, high quality, environmental responsibility, social impact, long-term operational benefit, etc.).
- Mapping values, taking into account: the degree of importance for each stakeholder, the level of agreement between stakeholders (whether there are conflicts/disagreements).
- Formation of a basic “value profile” of the project - a guideline for further actions of the team.

Stage 2. Formation of a project team with an orientation towards value harmonization.

At this stage, the following are proposed as the main actions:

- Defining the team's organizational structure, roles, distribution of responsibilities (the RACI matrix is proposed as a leading model), communication channels that support participation and mutual understanding.
- Conducting a training or workshop with team members, where they are introduced to the results of Stage 1 (stakeholder values, their priority, level of agreement, other metrics).
- Introducing interaction mechanisms that promote proactive behavior: regular meetings, feedback sessions, coaching groups, facilitation, discussion of value dilemmas, etc.
- Creating a culture in which the team independently monitors its own understanding of the project's values, identifies possible discrepancies and initiates changes.

Stage 3. Monitoring, adjusting and supporting value harmonization during the project implementation process.

At this stage, the following is carried out:

- Setting key performance indicators (KPIs) with a focus on value alignment (e.g.: % of agreed values, level of stakeholder satisfaction, number of internal conflicts due to value inconsistency, number of team initiatives to change processes due to values).
- Regular assessments of the state of the team and stakeholders: through surveys, facilitation, discussions, change analysis.
- Taking corrective measures when a decrease in value alignment or an increase in risks is detected: adapting the team structure, changing communication channels, repeated trainings, changing interaction processes.
- Supporting a learning culture: team reflection, sharing experiences, "lessons learned from the project", allowing the team to transform and improve throughout the project life cycle.

Justification of the effectiveness of the method.

The method of proactive organizational development of the construction project team when implementing value harmonization has the following advantages:

- *Increasing consistency between stakeholders.* Through values mapping and the team's active work on their coordination, the risks of conflicts, misunderstandings and delays are reduced. Studies show that effective stakeholder management contributes to the creation of project value through such parameters as cost optimization, time savings, and quality improvement [4-6].
- *Formation of a proactive team.* The team does not simply perform the specified tasks, but is aware of the project's value orientations, participates in their transmission, activates changes within the project, adjusts the construction project management system. This contributes to higher motivation, involvement and responsibility of team members.
- *Dynamic change management.* The method involves regular monitoring, adjustment and training, which allows you to adapt to changes in stakeholders, the environment, expectations – this is especially important in construction projects with

a long life cycle (for example, in complex infrastructure restoration projects) and many uncertainties, among which the most serious source of uncertainty is the war in Ukraine, caused by the aggression of the Russian Federation.

- *Increased project performance.* Value alignment means more holistic management, better decisions, less risky changes, lower costs of conflicts or rework, and as a result – increased stakeholder satisfaction, a higher quality construction project product, a better company reputation, and an increased likelihood of concluding agreements with stakeholders for new construction projects.

Conclusions. In the context of construction projects, which are increasingly focused on creating value and involving a large number of multi-vector stakeholders, the method of proactive organizational development of the team when implementing value harmonization becomes particularly relevant. The proposed method allows integrating the value aspects of stakeholders into the work processes of the project team - from mapping values and forming a team to monitoring and adjusting project artifacts during its implementation. The method must be implemented on the IT platform of the construction company. The use of such a method contributes to a more conscious and adaptive work of the team, strengthens mutual understanding between stakeholders, increases the quality of management and project effectiveness.

For further research [7], it is possible to determine the direction of empirical implementation of the method on real construction projects (for example, public infrastructure or housing), with measurement of the impact on performance indicators (value alignment, stakeholder satisfaction, adherence to budgets/deadlines, number of conflicts, etc.). Also, as a relevant direction, it is possible to identify a comparison of the proposed method with traditional (in particular, value-oriented) methods of team and stakeholder management to identify differences and advantages of the new approach.

References:

1. Forgues D. A Value-Based Approach to Managing Construction Projects: Trends and Research. Paper presented at PMI® Global Congress 2005 – North America, Toronto, Ontario, Canada. Newtown Square, PA: Project Management Institute, 2005. https://www.pmi.org/learning/library/2019/04/07/15/34/value-based-approach-managing-construction-projects-7456?utm_source=chatgpt.com
2. Danchenko E. B., Nakhimi M. M., Savina O. Y. Value-oriented management of the content of the construction project. *Bulletin of NTU "KhPI". Series: Strategic management, portfolio, program and project management*, 2020, №2, pp.23-31.
3. Yang Y., Wei Z., Zhang Z. Stakeholder Relationship in Construction Projects: A Mixed Methods Review. *Buildings*, 2023, 13(12), 3122.
4. Gaur S., Tawalare A. G. Stakeholder management and its role on value creation in construction projects: A cross-case analysis. *Asian Journal of Civil Engineering*, 2024, 25(4), 1-14.
5. Alhiddi A., Osborne A.N., Anyigor K.T. (2019). Organizational Culture and Stakeholder Success Criteria in Construction Projects. *Periodica Polytechnica Architecture*, 50(2), 148–154.
6. Vijayan P., Deepan G. Effectiveness Of Stakeholder Management In Construction Industry. *International Journal of Intellectual Advancements and Research in Engineering Computations*, 2021, 9(3), 60–68.
7. Bushuyev S., Vedmiediev O. Models of value measurement in a construction project in the context of strategic communications with key stakeholders. *Management of Development of Complex Systems*, 2025, 63, 6–13,

Oleksandr Voitenko

PhD, associated professor of the Project Management Department

Valerii Hudov

PhD student of the Project Management Department

Kyiv National University of Construction and Architecture

DAMPEN-TURBULENCE MODELS OF PROJECT PORTFOLIO MANAGEMENT TEAM OF A GENERAL CONTRACTING ORGANIZATION AND THEIR IT SUPPORT

Abstract. It is determined that the construction industry operates in a turbulent environment, where high rates of change in the economy, technology, and socio-political factors significantly affect projects. The dampen-turbulent model of construction project portfolio management is proposed, based on flexible approaches and the use of IT tools for monitoring, adaptation, and real-time decision support. The model ensures process stability, improves communication efficiency, and develops organizational flexibility, reducing the impact of turbulent changes on project implementation. Directions for further research are considered, and conclusions are formulated.

Keywords: project management, dampen-turbulent management, project portfolio, general contracting organization, information technology

Introduction.

The construction industry is one of the most complex and risky areas of economic activity, which combines numerous external and internal factors of influence. General contracting organizations operate in an environment where market, economic, technological and socio-political conditions change at a high speed. Such an environment has been called *turbulent*, which reflects its uncertainty, instability and tendency to sharp fluctuations. In such conditions, traditional approaches to project management, in particular within the framework of a portfolio of construction projects, are not effective enough.

Turbulence affects not only the deadlines and financial indicators, but also the *team dynamics*, the psychological state of the personnel, the level of communications and the effectiveness of management decisions. This makes it necessary to create *dampen-turbulent management models* that are able not only to adapt to changes, but also, if possible, to anticipate them, prepare for a reactive response to them. An important role in this process is played by information technologies that provide monitoring, analytics, risk forecasting and support for decision-making in real time.

The purpose of this study is to develop a *conceptual dampen-turbulent model of team management of a portfolio of construction projects in a general contracting organization using modern IT tools* that ensure the stability of the construction company's functioning and the effective implementation of its project portfolio in conditions of environmental turbulence.

To achieve the goal, the following main tasks have been identified:

- to analyze the nature and factors of turbulence in the activities of general contracting organizations;
- to determine the principles of forming dampen-turbulent models of team management;

- to develop a conceptual approach to IT support for project portfolio management of a construction company;
- to assess the potential impact of dampen-turbulent models on the effectiveness of project portfolio management of a construction company.

Main part.

The concept of “organizational environment turbulence” encompasses a set of factors that cause unpredictable changes in the external and internal environment of the organization. According to H. Kerzner [1], turbulence in project management is due to high interdependence of processes, rapid changes in technologies and insufficient stability of external markets. For construction companies, this manifests itself in fluctuations in the cost of materials, changes in regulatory requirements, shortage of qualified personnel and delays in supply chains.

Project portfolio management in such conditions should be based on the principles of *flexibility*, *systematicity* and *information transparency*. Scientific and applied research and standards [2-5] emphasize that the digitalization of management processes and the integration of IT platforms are the key to increasing the sustainability of organizations. In modern general contracting structures, this means the implementation of integrated portfolio management systems (MS Project, Jira, Oracle Primavera, IT tools of the Office 365 ecosystem, etc.), which ensure coordination between teams, resource control and risk forecasting.

The study uses a *systems approach* that considers the project portfolio as a single adaptive system. The methodological basis is the principles of *Agile* and *Scrum*, which allow for flexible response to environmental changes and increase team autonomy. For analytical data processing, business analytics tools (Power BI, Qlik Sense) are recommended, which provide visualization of turbulence and efficiency indicators.

It is proposed to develop a Dampen-turbulent model for managing a project portfolio of a construction company with IT support based on the following proposed principles:

- Maximum possible flexibility in management, in particular through the use of proactive models and flexible Agile methodology;
- Proactive response to turbulence should be based not on pre-developed detailed response plans, but on scenario-based framework strategies in response to classes of predicted turbulent impacts;
- To respond to unpredictable turbulent impacts, the Dampen-turbulent model should use four artifacts, such as: 1) knowledge base on project and portfolio management, 2) team intuition, 3) team emotional intelligence, 4) artificial intelligence models;
- Use of modern flexible, customizable IT tools with elements of artificial intelligence.

The Dampen-turbulent model is proposed as one that involves the design, implementation, and use of three interconnected subsystems:

1. ***Turbulence monitoring subsystem.*** Implemented using IT tools for collecting data from the market, project systems and HR analytics. It allows you to

identify trends in changes, signal risks and assess their impact on the portfolio project management teams, the portfolio as a whole and the construction company. It should contain a set of metrics and indicators to visualize the state of the portfolio environment.

2. ***Team adaptation subsystem.*** Provides models and algorithms for redistributing resources, defining flexible roles for construction project portfolio management team members, and forming backup action scenarios. The principles of team coaching, motivation and psychological resilience of personnel play a key role in this subsystem.

3. ***Decision support subsystem.*** It integrates analytical tools with portfolio management systems, giving the manager the opportunity to see the situation in real time. This allows you to quickly change project priorities and reduce the impact of negative factors.

It is important that these subsystems are formalized in the form of business processes, debugged and implemented in the IT system of the construction company. Implementation of business process models in the environment of executed business processes, integration of such an environment into a single IT system (IT platform) of the construction company will allow to implement adaptive, IT-supported dampen-turbulent management of the project portfolio of the construction company.

The proposed dampen-turbulent model of project portfolio management, using an IT system to support and ensure such management, will provide the construction company with the following capabilities:

- *Ensure the stability of management processes.* The model allows you to quickly identify signs of instability and prevent crisis scenarios through early response.
- *Increase the efficiency of communications.* The use of integrated digital platforms ensures transparency of information flows between project portfolio participants.
- *Balance the load on teams.* IT support algorithms allow you to automatically optimize the distribution of tasks, resources and deadlines depending on the level of turbulence.
- *Develop organizational flexibility.* The system contributes to the formation of a culture of rapid adaptation, which increases the competitiveness of the general contracting organization when implementing project portfolios.

The following risks of using the dampen-turbulent model of construction project portfolio management can be identified:

- Excessive complexity of the management system using such models, as a result of which the efficiency of the system will decrease;
- Incompetent use, insufficient professionalism of the construction project portfolio management team members;
- A large number of significant turbulent changes in the environment, simultaneously affecting the construction project portfolio, can unbalance the project portfolio management system;

- Artificial intelligence errors and the lack of verification of the results provided by such a tool can affect the quality of management decisions regarding the construction project portfolio.

The model can be implemented in practice in the form of an *interactive information dashboard* that displays key indicators of project portfolio stability: risk level, team workload, financial deviations, time delays, environmental turbulence index, other indicators and metrics.

Comparison of the proposed model with traditional management approaches showed that the dampen-turbulent model not only reduces the level of uncertainty, but also creates the prerequisites for strategic learning of the organization. In particular, data analysis from previous crisis situations allows improving algorithms for responding to turbulent impacts.

Among the limitations of the model, it is worth noting the dependence on the quality of the input data and the readiness of personnel for digital transformation. The success of the implementation largely depends on the management culture, management support and access to modern IT resources. Therefore, it is necessary to combine technological and humanitarian approaches to team development, ensuring a balance between automation and the human factor.

Conclusions

1. Modern general contracting organizations operate in an extremely turbulent environment, where traditional management methods lose their effectiveness. Therefore, an urgent scientific task is to develop new management models that would take into account turbulence and contain patterns of response to such turbulence, or ensure effective coexistence with it.

2. Dampen-turbulence models allow to ensure the stability of the project portfolio team through a combination of monitoring, adaptation and decision-making support.

3. IT support of such models creates conditions for digital transparency, reducing reaction time and increasing the flexibility of management processes.

4. A promising direction for further research is the development of *mathematical models for predicting turbulence* and the use of machine learning algorithms to build a system for dynamic stabilization of the project portfolio in real time.

References:

1. Kerzner H. Project Management: A Systems Approach to Planning, Scheduling, and Controlling. – Wiley, 2022.
2. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition. USA, PMI, 2021, 250 p.
3. ISO 21504:2015. Project, programme and portfolio management — Guidance on portfolio management.
4. Ahern T., Leavy B., Byrne P.J. Complex project management as complex problem solving: A distributed knowledge management perspective. – International Journal of Project Management, 2014, 32(8), 1371-1381.
5. ISO 56002:2019. Innovation Management System – Guidance.

¹ Oleg Zhulkovskyi

Candidate of Technical Sciences, Associate Professor, Acting Head of Department of Software Systems

² Inna Zhulkovska

Candidate of Technical Sciences, Associate Professor of Department of Computer Science and Software Engineering

³ Anastasiia Shevchenko

Student of Department of Cybersecurity and Information Technologies

¹ *Dniprovsky State Technical University, Ukraine*

^{2,3} *University of Customs and Finance, Ukraine*

OPTIMIZING NUMERICAL THERMAL SIMULATIONS THROUGH DATA-LEVEL PARALLELISM

Mathematical modeling of thermal processes plays a crucial role in designing energy systems and developing materials with specific thermophysical properties. Computational modeling reduces experimental costs, enables visualization of complex processes, and supports informed engineering decisions. The growing complexity and nonlinearity of simulated phenomena demand higher accuracy and computational efficiency. As hardware scaling approaches its limits, further performance improvement depends on parallel computing methods. Among these approaches, Single Instruction Multiple Data (SIMD) technology offers a practical means of accelerating heat conduction simulations by executing identical operations on multiple data elements concurrently [1]. Data-level parallelism has therefore become an essential tool for efficient modeling of heat and mass transfer in modern computational systems.

According to Flynn's taxonomy, SIMD is a form of parallel computing where a single instruction operates on multiple data elements simultaneously. The integration of SIMD extensions such as SSE, AVX, and AVX-512 in x86 processors, and NEON in ARM architectures, has significantly improved computational throughput for vector operations on data blocks up to 512 bits while preserving energy efficiency [2]. Numerous studies have demonstrated substantial performance gains from SIMD-based optimization in scientific and engineering computations, including dense and sparse matrix operations, FFT, Gaussian filtering, large-integer arithmetic, advanced numerical solvers, and data compression algorithms. Reported speedups ranged from 1.5× to over 10× depending on data precision and algorithm characteristics [3].

This study presents a vectorized algorithm for solving the unsteady heat conduction equation using AVX instructions. The problem considers a one-dimensional steel rod with a thermally insulated lateral surface and Robin (convective-type) boundary conditions at both ends. Material properties depend on temperature, and the initial temperature distribution is known. The heat conduction equation is discretized using an explicit finite-difference scheme derived based on the thermal balance method [4] on a uniform grid.

Algorithmic optimization is achieved through the explicit vectorization of the

computational loop that updates the internal nodes, while scalar processing handles the remaining array elements. SIMD AVX2 intrinsics enable simultaneous processing of four double-precision or eight single-precision values per processor cycle, using 256-bit YMM registers and intrinsic functions.

Numerical experiments were conducted on an Intel Core i7-12700H CPU with 32 GB RAM. The simulations considered a steel rod 0.1 m long with an initial temperature of 293 K, the boundary temperatures 573 K (left) and 293 K (right), and the heat transfer coefficients of 500 and 1000 W·m⁻²·K⁻¹. Spatial grid steps varied from 10⁻² to 4×10⁻⁵ m, with time steps determined by the CFL stability criterion. Model adequacy was validated against reference data.

Three algorithmic variants were tested: scalar (No-AVX), vectorized double-precision (AVX-double), and vectorized single-precision (AVX-float). Vectorization efficiency was observed to increase with grid refinement, achieving speedups of 1.43× for AVX-double and 1.9× for AVX-float. The single-precision implementation provided up to 1.35× additional acceleration due to processing twice as many values per SIMD register. Actual speedups remained below the theoretical maximum because of memory bandwidth limitations, loop tails, and partial vectorization. For coarse grids, vectorization benefits were negligible.

These results confirm the potential of SIMD vectorization for computationally intensive thermophysical simulations, particularly at high spatial resolutions. Future research may explore adaptive vectorization schemes tailored to specific problem characteristics and processor architectures, as well as hybrid parallelization approaches combining SIMD with multithreading or asynchronous computations [5], to further accelerate multidimensional heat and mass transfer simulations.

References:

1. R. Zheng, and S. Pai, Efficient execution of graph algorithms on CPU with SIMD extensions, in 2021 IEEE/ACM Int. Symp. Code Gener. Optim. (CGO). Seoul, (2021), pp. 262–276. doi: 10.1109/cgo51591.2021.9370326
2. J. Wang, L. Yu, W. Zhuang, X. Yang, S. Zhang, and Z. Qin, Research on vector extension of instruction set architecture, in 2024 3rd Int. Conf. Cloud Comput., Big Data Application Softw. Eng. (CBASE), Hangzhou, (2024), pp. 378–385. doi: 10.1109/cbase64041.2024.10824427
3. G. Sansone, and M. Cococcioni, Experiments on speeding up the recursive fast fourier transform by using AVX-512 SIMD instructions, Lecture Notes in Electrical Engineering, Cham: Springer, 1036, (2023), pp. 255–263. doi: 10.1007/978-3-031-30333-3_34
4. O. Zhulkovskiy, I. Savchenko, I. Zhulkovska, I. Petrenko, O. Davitaia, and A. Titov, Features of mathematical simulation of the processes of combined heat transfer in waveguides, in 2022 IEEE 4th Int. Conf. Modern Elect. Energy System (MEES), (2022), pp. 1–5, doi: 10.1109/mees58014.2022.10005676
5. O. Zhulkovskiy, I. Zhulkovska, P. Kurliak, O. Sadovoi, Y. Ulianova, and H. Vokhmianin, Using asynchronous programming to improve computer simulation performance in energy systems, *Energetika*, 71(1), (2025), pp. 23–33. doi: 10.6001/energetika.2025.71.1.2

Vadym Ziuziun

PhD, Associate Professor of the Department of Technologies Management

Tetiana Danilina

Master's degree student of the Department of Technologies Management

Taras Shevchenko National University of Kyiv

FEATURES OF DATABASE STRUCTURE DEVELOPMENT FOR THE "PETHEALTH" MOBILE APPLICATION

Abstract. The work is devoted to designing a database for the "PetHealth" mobile application. A three-stage approach (conceptual, logical, and physical modeling) was used to solve the problem of fragmentation of veterinary data. Key entities and their relationships were identified. The selected PostgreSQL DBMS provides a reliable and scalable foundation for data management.

Keywords: Database design, PostgreSQL, conceptual model, logical model, relational database, mobile application.

The veterinary sector in Ukraine lags behind the general digitization of medical services. Medical records for pets are often kept in separate local systems or on paper. This fragmentation makes it difficult to track an animal's medical history and plan visits. Researchers in the field of mobile health (mHealth) note that, despite its advantages, the sector faces challenges in standardization, data security, and usability, making reliable data design critically important [1].

The "PetHealth" mobile application project, the theoretical background for which was explored previously [2], aims to bridge this gap by creating a unified digital platform for owners and veterinary clinics. A central component is the database, which must efficiently manage profiles, pet data, appointment schedules, and medical histories.

The database design process followed a standard three-step approach to ensure a reliable and efficient structure:

1. Conceptual modeling. Identification of key entities in the subject area and high-level relationships between them.
2. Logical modeling. Detailing the data structure, attributes, and primary and foreign keys to implement relationships.
3. Physical modeling. Transformation of the logical model into a specific implementation for the selected DBMS.

This approach, which separates models into levels, allows for the systematic development of the project's information structure, which is key to successful implementation.

At the conceptual stage, the fundamental entities were defined: Owner, Pet, Clinic, Appointment, and Service. Key one-to-many relationships were also identified (for example, one Owner can have several Pets).

These entities and relationships were transformed into a formal logical model (relational schema). The generalized database structure showing the key tables and their relationships is shown in Fig. 1.

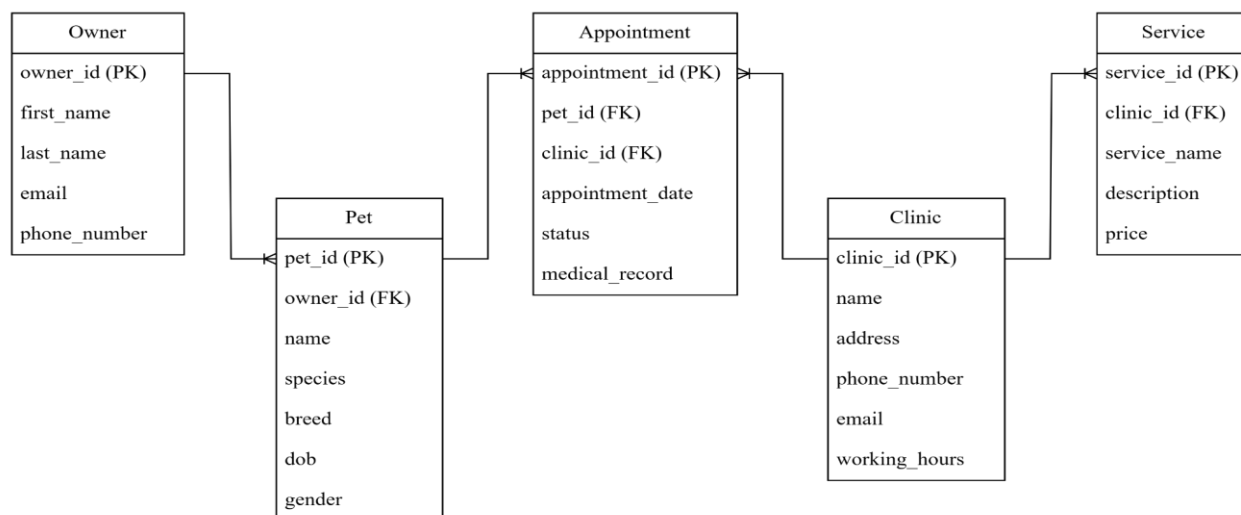


Fig. 1. Logical model of the "PetHealth"

The physical model defines the details of the implementation of the logical structure (Fig. 1) in the PostgreSQL DBMS environment.

PostgreSQL was chosen for physical implementation. This DBMS is known for its reliability, compliance with ACID (Atomicity, Consistency, Isolation, Durability) standards, scalability, and powerful feature sets, which are crucial for processing sensitive medical data and ensuring its integrity.

Key decisions in physical design included:

Data types: Appropriate types were selected for each attribute, such as SERIAL for auto-increment primary keys, VARCHAR for text fields, DATETIME for date and time of admission, and DECIMAL for prices. *Constraints:* PRIMARY KEY, FOREIGN KEY, NOT NULL, and UNIQUE constraints were implemented to ensure data integrity. *Relationships:* Relationships with foreign keys were defined with the ON DELETE CASCADE option. This ensures that related records are automatically deleted.

The systematic design process has created a solid foundation for the "PetHealth" system. The relational database created on PostgreSQL effectively organizes information, ensures data integrity, and provides scalability. The visualized structure (Fig. 1) confirms the readiness of the data architecture to implement the stated functionality and solve the problem of veterinary data fragmentation.

References:

1. Designing and implementing mHealth technology: the challenge of meeting the needs of diverse communities - PMC. PMC Home. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10314586/>.
2. Даніліна Т., Зюсюн В. Теоретичне обґрунтування розробки мобільного додатку "PETHEALTH: ветеринарний помічник". 2st international scientific and practical conference «Information Systems and Technology: Results and Prospects» (IST 2025), Kyiv, Ukraine, 2025, 198-200.

Vadym Ziuziun

PhD, Associate Professor of the Department of Technologies Management

Oleksii Korytnyi

Master's degree student of the Department of Technologies Management

Taras Shevchenko National University of Kyiv

RISK CONTROLLING AND MONITORING SOFTWARES THAT CONSIDER HUMAN BIAS AS A MAJOR RISK

Abstract. This research argues that a specialized IS offers a structured, quantitative, and proactive framework to not only identify and assess typical technical and financial risks but also to meticulously manage these complex human factors. The system must integrate advanced features like behavioral analytics and lessons-learned repositories to move beyond simple checklists. Furthermore, this paper provides a concise overview of existing risk management programs and platforms currently available in the market. By critically analyzing the capabilities and limitations of commercial-off-the-shelf (COTS) solutions – and noting their common deficiencies in handling the nuance of human risk – this work solidifies the requirement for a tailored IS that can effectively bridge the gap between technical risk assessment and human-centric project realities.

Keywords: information system, risk management, IT projects, human-induced risks, cognitive biases, risk management software.

Previous works have established the foundational importance of addressing human-induced risks in technology management, moving beyond generic risk models [2-3] and how can we assess them in IT projects [1]. This need for proactive management requires a critical comparison with existing solutions; thus, we must now analyse current commercial and proprietary software programs available for tracking IT project risks.

SHERPA human factor reliability [4] provides its software for pharmaceutical, oil & gas, defence and healthcare sectors. The SHERPA human factors software package integrates several different modules to help you stay on top of all the most critical tasks and improve human reliability:

- **Risk ranking:** generates risk matrices to indicate great concern areas and set priorities in risk management. The module allows to evaluate the most common zone for a human error, output risk index and likelihood and suggests tasks prioritization.
- **Hierarchical task analysis and human factors risk analysis:** conducts human factor risk assessments, including COMAH Safety Cases required by the Health and Safety Executive in the UK; Hierarchical Task Analysis (HTA) to systematically document tasks, PHEA etc.

Apart from this SHERPA also provides Failure Mode / Performance Influencing Factors (PIF) Analysis and Swimlane or Sequentially Timed Event Plot (STEP) Diagramming.

The NIST AI Risk Management Framework (AI RMF) [5] is a voluntary framework developed by the U.S. National Institute of Standards and Technology (NIST). Its primary purpose is to help organizations across both the public and private sectors manage the specific risks associated with designing, developing,

deploying, and using Artificial Intelligence (AI) products and systems. The AI RMF is distinct from traditional IT risk frameworks because it focuses on trustworthiness in addition to security. It aims to improve the ability to incorporate considerations such as:

- Fairness and Bias.
- Transparency and Explainability.
- Validity and Reliability.

Although the system does not directly specialise on human bias risks and their indication/evaluation, AI RMF outlines their 4 key pillars (Govern, Map, Measure, Manage) as to prevent extensive human risks with AI regulatory data governance. In order to start the “machine” the Institution suggests the next steps taken:

- 1) Establish a Strong Governance Structure.
- 2) Conduct a Comprehensive Risk Assessment.
- 3) Implement Risk Management Strategies.
- 4) Monitor and improve.
- 5) Knowledge sharing.

Summarizing, various worldwide practices incorporate standard time-checked risk management models and approaches to define and manage risks in projects. However, with AI slightly taking over the edge of new insights, people also adopt the machine learning to optimise and control the common project management to a new level, which predicts the outcome. Thus, it supports our previous statements [2-3] about the necessity to create an information system for risk management of it projects and make an artificial intelligence its core module of analysis and prediction

References:

1. Ziuziun V. Analysis of possible risks in human resources management in IT companies. 5th International Scientific and Practical Internet Conference «Integration of Education, Science and Business in Modern Environment: Summer Debates», August 3-4, 2023. P.79-81. URL: <http://www.wayscience.com/konferentsiya-5-3-4-serpnya-2023/>. Access date: 23.10.2024
2. Korytnyi O., Ziuziun V. Digital Ways of Controlling Risk Management and Assessment in Projects. 2st International scientific and practical conference «Information Systems and Technology: Results and Prospects» (IST 2025), Kyiv, Ukraine, 2025, p. 167-171.
3. Korytnyi O., Ziuziun V. Justification of the need to create an information system for risk management of IT projects. Information Technology and Implementation (Satellite): Conference Proceedings, November 21, 2024, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publishing House «Caravela» – P. 236-237.
4. SHERPA Human Reliability 2025. Official website. <https://www.humanreliability.com/human-factors-software/#demo>.
5. Leia Manchanda, CISO Perspectives: A Practical Guide to Implementing the NIST AI Risk Management Framework (AI RMF). September 30, 2025. Field CISO.

Vadym Ziuziun

PhD, Associate Professor of the Department of Technology Management

Yuliia Koziuk

Master's degree student of the Department of Technology Management

Taras Shevchenko National University of Kyiv

KEY CHALLENGES AND RISK MITIGATION STRATEGIES FOR UKRAINIAN IT STARTUPS DURING THE WAR

Abstract. This paper analyses the main challenges and risks faced by Ukrainian IT startups under wartime conditions. It examines the impact of military instability, financial and market constraints, personnel losses, and technical vulnerabilities. The study proposes approaches to risk management to overcome the challenges.

Keywords: startups, IT sector, risks, financing, marketing, competition, war, Ukraine.

The unstable military and political situation have become an uncontrollable risk factor for Ukrainian businesses. Startups face infrastructure destruction, resource supply disruptions, team evacuations, and staff mobilization [1]. Only about 1% of IT professionals have received deferment from conscription, which complicates company operations [2]. The war and geopolitical instability also discourage foreign clients and investors [2]. Consequently, some projects have closed, and the number of orders and market opportunities has declined. At the same time, surveys indicate that Ukrainian startups remain more optimistic about the future than traditional businesses [1].

Financial instability is another significant risk for startups. Young companies often struggle to attract external funding, leaving them short of resources to implement their ideas. According to The Gradus Survey, 42% of startup founders mentioned difficulties in securing investment and seed capital [1]. During the war, venture investors have become more cautious, leading to a temporary decline in interest in Ukrainian startups [2].

Marketing and market-related challenges remain among the key concerns for Ukrainian startups. For 46% of respondents, finding clients and new markets has become the main challenge [1]. The domestic market is constrained, and consumers' purchasing power has decreased, making it difficult for many projects to reach sufficient audiences – 47% reported a lack of orders due to declining demand [1]. Internal analysis has also revealed the issue of ineffective marketing strategies, which can prevent startups from properly communicating their product value to clients.

Startups often operate in highly competitive niches. Despite the difficulties, new projects continue to emerge in Ukraine, increasing competition and forcing startups to invest substantial resources in promotion. The absence of a unique value proposition or a clear competitive advantage can be fatal to a project. According to CB Insights, 42% of startups fail because they offer a product for which there is no real market need [3]. Thus, the creation of unique value and flexibility in responding to changing demand is essential for survival. Market risks also include rapid shifts in consumer preferences and the emergence of new technological trends, which require

startups to adapt their strategies continuously. Startups operate under limited resources, which increases their vulnerability to internal disruptions and errors. Operational risks include technical failures and human factors such as equipment breakdowns, software malfunctions, coding errors, and system architecture flaws. The analysis also revealed threats such as security vulnerabilities and software instability, which undermine user trust and may lead to financial losses.

Personnel risks are also important. Small startup teams are particularly sensitive to the loss of key employees. The war has exacerbated the talent shortage due to migration and mobilization, leading to deficits of engineers and managers. This results in development delays and increased training costs for new staff. Global talent competition further drives up hiring costs and complicates project scaling.

For effective risk management, it is important to distinguish risks by their level of controllability. This allows prioritization of response measures and optimization of managerial decisions. The table below presents a classification of key startup risks according to their controllability and main characteristics.

Table 1 – Startup risk classification by controllability

Level of controllability	Risk category	Main manifestations / challenges
Uncontrollable	Military and political	War, infrastructure destruction, occupation threats, power outages.
Partially controllable	Human resources	Mobilization of key specialists, talent shortages.
Controllable	Market / marketing	Ineffective marketing strategies, lack of unique value proposition, client and market acquisition challenges.
	Operational / technical	Coding errors, software malfunctions, security vulnerabilities, human factor.

The analysis shows that the degree of risk controllability defines the appropriate mitigation strategy. Uncontrollable risks require adaptive measures, partially controllable ones call for managerial action, and controllable risks demand continuous process improvement. Following the Russian full-scale invasion, Ukrainian startups have faced multiple challenges. Effective response requires systematic risk management – early threat identification, contingency planning, financial diversification, and investment in product reliability and security. Despite the difficulties, Ukrainian startups remain resilient, proving that even in crisis, proper preparation enables growth opportunities.

References:

1. Gradus Research. Official report. URL: <https://gradus.app/uk/open-reports/ukrainian-startups-feel-more-optimistic-compared-traditional-businesses/>. Access date: 25.10.2025.
 2. BizMag. IT-галузь в Україні: поточний стан та перспективи. URL: <https://bizmag.com.ua/it-galuz-v-ukrayini/#:~:text=>. Access date: 25.10.2025.
- EnlightenBio. Why most startups fail: the brutal truth and how to beat the odds. News & Features. URL: <https://enlightenbio.com/news-and-features/2025/08/11/why-most-startups-fail-the-brutal-truth-and-how-to-beat-the-odds/#:~:text=1>. Access date: 25.10.2025.

Vadym Ziuziun

PhD, Associate Professor of the Department of Technologies Management

Daria Shcherbak

Master's degree student of the Department of Technologies Management

Taras Shevchenko National University of Kyiv

MANAGEMENT OF QUALITY IN AN IT PROJECT DURING THE DEVELOPMENT OF THE ECOLOGICAL WEB APPLICATION “СМАЧНОГО”

Abstract. The article explores the application of quality management practices in an ecological IT project aimed at reducing food waste. It presents the development of the “СмачноGo” web application, designed with a user-centric approach and aligned with ISO 9001 and ISO/IEC 25010 standards. Key quality characteristics, metrics, and agile practices are outlined to demonstrate a systematic and sustainable development process.

Keywords: Quality management, IT project, software quality model, ISO 9001, ISO/IEC 25010, ecological web application, food waste.

In today’s world, the problem of food waste is reaching a global scale. According to the Food and Agriculture Organization of the United Nations (FAO), around 1.3 billion tons of food are lost worldwide each year – almost one-third of all food produced [1]. In Ukraine, estimates by the Ministry of Environmental Protection and Natural Resources show that over nine million tons of food waste are generated annually. At the same time, the lack of clear legislative regulation aimed at preventing food losses complicates efforts in this field [2].

One of the potential solutions is the web application “СмачноGo”, the Ukrainian counterpart to the Too Good To Go platform, which helps reduce food waste by offering surplus meals and products from restaurants, cafés, bakeries, and grocery stores at a discounted price.

The aim of the project is to develop a web application that contributes to the reduction of food waste by creating a digital platform enabling interaction between dining establishments, grocery stores and consumers, with the purpose of ecological and economical use of food products.

Quality management policies:

1. Ensuring a positive user experience through speed, stability and secure operations of the web application.
2. Promoting the reduction of food waste and improving the efficiency of cooperation with suppliers.

The project team includes a project manager, frontend & backend developers, UI/UX designer, QA tester and a security specialist. Quality management is implemented according to the principles of ISO 9001 [3] and ISO/IEC 25010 [4].

The quality planning stage included defining functional requirements (user registration, authorization, order creation and viewing, profile management, payment, notification system) and non-functional requirements (performance, security, compatibility, usability, scalability).

For the “СмачноGo” web application, the most important quality characteristics according to ISO/IEC 25010 were selected – those that most influence user efficiency and system stability (Table 1).

Table 1. Metrics for assessing quality characteristics

Characteristic	Metric	Data collection method	Acceptance criteria
Functional suitability	Percentage of implemented functions	Functional testing	$\geq 95\%$ of functions implemented
Performance efficiency	Average server response time (ms)	System monitoring	≤ 1 second
Usability	User satisfaction rate (%)	Surveys, usability testing	$\geq 80\%$ positive responses
Reliability	Mean time between failures (hours)	Automatic log monitoring	$\geq 99\%$ uptime
Security	Number of critical vulnerabilities	Penetration testing, security audit	0 critical vulnerabilities

The innovation of the “СмачноGo” project lies in the combination of a technological solution with an ecological mission. The application not only helps reduce food waste but also fosters a culture of conscious consumption among users. In the future, implementation is planned for a bonus system, ratings of ecological responsibility of establishments, and a charitable module for the transfer of food products to social organizations.

Thus, quality management of the IT project “СмачноGo” is a key factor for its successful implementation. It ensures technical stability, user satisfaction and sustainable ecosystem development of the product. Adherence to international standards ISO 9001 and ISO/IEC 25010 enhances the web application’s competitiveness and its positive environmental impact.

References:

1. Food and Agriculture Organization of the United Nations. (2022). Cutting food waste while improving food security and environment in Ukraine. URL: <https://www.fao.org/europe/news/detail/Cutting-food-waste-while-improving-food-security-and-environment-in-Ukraine/en>.
2. ТОВ «SEC Ecology». (2016). Національна стратегія поводження з відходами для України (попередній варіант). Київ.
3. ДСТУ ISO 9001:2015 (ISO 9001:2015, IDT). (2016). Системи управління якістю. Вимоги. Київ: ДП «УкрНДНЦ». URL: <https://ontu.edu.ua/download/pubinfo/dcc/standard-ISO-9001-2015-ua.pdf>.
4. ДСТУ ISO/IEC 25010:2016 Інженерія систем і програмних засобів. Вимоги до якості систем і програмних засобів та її оцінювання (SQuaRE). Моделі якості системи та програмних засобів (ISO/IEC 25010:2011, IDT).



Syngenta is a leading science-based agtech company. They help millions of farmers around the world to grow safe and nutritious food, while taking care of the planet.

Syngenta is a global company with headquarters in Switzerland. 30,000 employees, in more than 90 countries are working to transform how crops are grown and protected.

Syngenta's portfolio includes more than 80 plant protection products, as well as many hybrids and seed varieties of field and vegetable crops. More than 300 specialists work in the Ukrainian branch of the company, including many candidates in agricultural sciences. The goal of the Syngenta company in Ukraine is to offer comprehensive solutions for Ukrainian farmers and a partnership preferred by Ukrainian and international companies. The Syngenta company is distinguished by its high-quality products, an integrated approach to solving tasks and using the advantages of Ukrainian agriculture.

Наукове видання

Інформаційні технології та взаємодії (сателітна): Матеріали
конференції, 21 листопада 2025 р., Київ, Україна

Information Technology and Implementation (Satellite): Conference
Proceedings, November 21, 2025, Kyiv, Ukraine

В авторській редакції

Редактор: Снитюк В.Є.

Верстка: Гнатієнко Г.М.

Обкладинка: Дубницька А.С.

Керівник видавничих проєктів Ю.В. Піча

Підписано до друку 21.11.2025 р. Формат 60x84 1/16.
Папір офсетний. Друк офс. Гарнітура Times New Roman
Умовн. друк. арк. 5,5. Обл.-вид. арк. 5,8
Зам. № 01/21-11-2025

Видавництво «Каравела»,
просп. М. Рокоссовського, 8а, м. Київ, 04201, Україна.
Тел. (068) 364-05-77. E-mail: caravela@ukr.net
www.caravela.com.ua

Свідоцтво
про внесення суб'єкта видавничої справи до
Державного реєстру видавців,
виготівників і розповсюджувачів
видавничої продукції:
ДК №2035 від 16.12.2004 р.