ISSN 2788-5518

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

№ 1 (09), 2025

НАУКОВИЙ ЖУРНАЛ

Видавничий дім
«Гельветика»
2025

ISSN 2788-5518

INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 1 (09), 2025

SCIENTIFIC JOURNAL

Publishing House
“Helvetica”
2025

Засновник, видавець та виробник журналу: «Відкритий міжнародний університет розвитку людини «Україна».

Реєстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 2317 від 11.07.2024 року

Ідентифікатор медіа: R30-05387.

Суб'єкт у сфері друкованих медіа – Заклад вищої освіти «Відкритий міжнародний університет розвитку людини «Україна» (вул. Львівська, буд. 23, м. Київ, 03115, viddilorgrobotu@ukr.net, тел. +380 (67) 503-64-52).

Журнал включений до переліку наукових фахових видань України категорія «Б», додаток 4 до наказу МОН України № 735 від 29.06.2021 р. та додаток 2 до наказу МОН України № 320 від 07.04.2022 р.

Журнал відповідає вимогам наказу МОН України № 32 від 15.01.2018 р.

Має міжнародні коди: ISSN 2788-5518 та DOI 10.36994/2788-5518.

Журнал індексується у міжнародних базах та каталогах Google Scholar, Index Copernscus, Vernadsky National Library of Ukraine.

Мова видання: українська, англійська

Рекомендовано до друку вченою радою Відкритого міжнародного університету розвитку людини «Україна» (протокол № 4 від 01.07.2025 р.)

URL-адреса видання: visn-icct.uu.edu.ua

Тематика журналу:

- Теоретичні основи інфокомунікацій та комп'ютерних наук
- Інфокомунікаційні системи
- Безпроводові технології
- Технічна електродинаміка та антени
- Волоконно-оптичні системи
- Радіорелейні та супутникові системи
- Телевізійні системи
- Системи відеоконференцзв'язку та відеоспостереження
- Комп'ютерні системи та мережі
- Комп'ютерна і програмна інженерія
- Системи SCADA
- Захист інформації та кібербезпека
- Прикладна математика та моделювання
- Метрологія та інформаційно-вимірювальні системи
- Медична та екологічна електроніка
- Автоматизація управління технологічними процесами

В журналі публікуються статті для захисту дисертацій зі спеціальностей: F2 Інженерія програмного забезпечення, F3 Комп'ютерні науки, F7 Комп'ютерна інженерія, F5 Кібербезпека та захист інформації, G5 Електроніка, електронні комунікації, приладобудування та радіотехніка.

Founder, publisher and producer of the journal: "Open International University of Human Development" Ukraine".

Registration of Print media entity: Decision of the National Council of Television and Radio Broadcasting of Ukraine No. 2317 as of 11.07.2024

Media ID: R30-05387.

Media entity – Higher Education Institution 'Open International UNIVERSITY of Human Development 'UKRAINE' (Lvivska str., 23, Kyiv, 03115, viddilorgrobotu@ukr.net, tel. +380 (67) 503-64-52).

The journal is included in the list of scientific professional publications of Ukraine – category "B", Annex 4 to the order of the Ministry of Education and Science of Ukraine № 735 from 29.06.2021 and Annex 2 to the order of the Ministry of Education and Science of Ukraine № 320 from 07.04.2022

The journal meets the requirements of the order of the Ministry of Education and Science of Ukraine № 32 from 15.01.2018.

It has international codes: ISSN 2788-5518 and DOI 10.36994 / 2788-5518.

The journal is indexed in international databases and directories of Google Scholar, Index Copernicus, Vernadsky National Library of Ukraine

Language of publication: Ukrainian, English

Recommended for publication by the Academic Council of the Open International University of Human Development "Ukraine" (Protocol № 4 dated 01.07.2025)

Publication URL: visn-icct.uu.edu.ua

Subject journal:

- Theoretical foundations of infocommunications and computer science
- Infocommunication systems
- Wireless technology
- Technical electrodynamics and antennas
- Fiber optic systems
- Radio relay and satellite systems
- Television systems
- Video conferencing and video surveillance systems
- Computer systems and networks
- Computer and software engineering
- SCADA systems
- Information security and cybersecurity
- Applied mathematics and modeling
- Metrology and information-measuring systems
- Medical and environmental electronics
- Automation of process control

The journal publishes articles for the defense of dissertations in the following specialties: F2 Software engineering, F3 Computer science, F7 Computer engineering, F5 Cybersecurity, G5 Electronics, Electronic Communications, Instrument Engineering and Radio Engineering

ЗМІСТ

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

CONTENTS

INFOCOMMUNICATION TECHNOLOGIES

COMPUTER TECHNOLOGIES

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.6:621.39
DOI https://doi.org/10.36994/2788-5518-2025-01-09-01

МЕТОДИКА ФОРМУВАННЯ СТРУКТУРИ ЦЕНТРУ ЗБИРАННЯ ТА ОБРОБЛЕННЯ ДАНИХ ПІД ЧАС МОНІТОРИНГУ СТАНУ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Гнатюк В. О., *к.т.н., доцент, Державний університет «Київський авіаційний інститут», Київ, Україна; ДержНДІ технологій кібербезпеки, Київ, Україна. https://orcid.org/0000-0002-4916-7149, viktor.hnatiuk@npp.kai.edu.ua*

Зандер К. Ю., *PhD аспірант, Державний університет «Київський авіаційний інститут», Київ, Україна. https://orcid.org/0009-0006-4944-9249, 8390983@stud.kai.edu.ua*

Анотація. *У статті подано комплексну методику проєктування архітектури центру збирання та обробки даних, призначену для моніторингу стану об'єктів критичної інфраструктури. Дослідження включає розроблення структурної та функціональної схеми, а також логічної моделі, що ілюструє взаємодію між компонентами системи, що беруть участь у збиранні та обробці даних у контексті моніторингу критичної інфраструктури. Запропонований підхід ураховує сучасні вимоги до надійності, відмовостійкості, масштабованості, кібербезпеки та продуктивності в режимі реального часу. Особлива увага приділяється типам датчиків, що використовуються в різних сферах критичної інфраструктури, включаючи енергетику, транспорт, водопостачання, телекомунікації й охорону здоров'я. Методика також досліджує інтеграцію сенсорних мереж з технологіями зв'язку п'ятого покоління (5G) та інтернетом речей (IoT), що дає змогу створювати більш чутливі й гнучкі системи моніторингу. Крім того, дослідження підкреслює багаторівневий підхід до обробки даних за допомогою парадигм периферійних, туманних і хмарних обчислень. Це дає можливість зменшити затримку, збалансувати навантаження між обчислювальними ресурсами й покращити реагування на критичні інциденти. Уведено математичну модель, основану на теорії множин, для визначення складу датчиків, необхідного для ефективного моніторингу інфраструктури. Методика розроблена як основа для побудови інтелектуальних систем моніторингу, здатних забезпечити стабільну та безпечну роботу основних послуг. Вона також підтримує інтеграцію інструментів штучного інтелекту для прогнозування стану системи й раннього виявлення аномалій. Це може значно покращити можливості прогнозного обслуговування та безпеку експлуатації. Результати цього дослідження є особливо цінними для фахівців, які працюють у сфері телекомунікацій, кібербезпеки, автоматизації та реагування на надзвичайні ситуації. Крім того, вони надають практичну інформацію урядовим і регуляторним органам, відповідальним за захист національних інфраструктурних активів.*

Ключові слова: *критична інфраструктура, сенсори, IoT, 5G, обробка даних у реальному часі.*

METHODOLOGY FOR DESIGNING THE ARCHITECTURE OF A DATA COLLECTION AND PROCESSING CENTER FOR CRITICAL INFRASTRUCTURE MONITORING

Viktor Gnatyuk, *Ph.D., Ass. Prof., State University "Kyiv Aviation Institute"; ICTIP, Kyiv, Ukraine. https://orcid.org/0000-0002-4916-7149, viktor.hnatiuk@npp.kai.edu.ua*

Kostiantyn Zander, *State University "Kyiv Aviation Institute", Kyiv, Ukraine. https://orcid.org/0009-0006-4944-9249, 8390983@stud.kai.edu.ua*

Abstract. *The paper presents a comprehensive methodology for designing the architecture of a data collection and processing center intended for monitoring the condition of critical infrastructure facilities. The study includes the development of a structural and functional scheme, along with a logical model illustrating the interaction between system components involved in data acquisition and processing within the context of critical infrastructure monitoring.*

The proposed approach considers modern requirements for reliability, fault tolerance, scalability, cybersecurity, and real-time performance. Special attention is given to the types of sensors used across various critical infrastructure domains, including energy, transportation, water supply, telecommunications,

and healthcare. The methodology also explores the integration of sensor networks with fifth-generation (5G) communication technologies and the Internet of Things (IoT), enabling more responsive and flexible monitoring systems.

Furthermore, the research emphasizes the layered approach to data processing through edge, fog, and cloud computing paradigms. This allows for reduced latency, balanced load distribution across computing resources, and improved responsiveness to critical incidents. A mathematical model based on set theory is introduced to define the sensor composition needed for effective infrastructure monitoring.

The methodology is designed to serve as a foundation for building intelligent monitoring systems capable of ensuring the stable and secure operation of essential services. It also supports the integration of artificial intelligence tools for system condition forecasting and early anomaly detection. This can significantly enhance predictive maintenance capabilities and operational safety.

The results of this study are particularly valuable to professionals working in telecommunications, cybersecurity, automation, and emergency response. Moreover, they provide practical insights for governmental and regulatory bodies responsible for protecting national infrastructure assets.

Key words: *critical infrastructure; sensors; IoT; 5G; real-time data processing.*

Вступ

Критична інфраструктура (далі – КІ) є основою функціонування держави й суспільства, охоплюючи енергетику, транспорт, водопостачання, телекомунікації та інші життєво важливі галузі. В умовах сучасних викликів, таких як кіберзагрози, техногенні аварії, природні катастрофи й військові ризики, необхідність забезпечення надійного моніторингу стану об'єктів КІ набуває стратегічного значення.

Одним із ключових аспектів ефективного моніторингу є створення центру збирання й оброблення даних, який дає змогу в режимі реального часу отримувати, аналізувати та інтерпретувати інформацію з розподілених сенсорних систем. Розробка методики формування структури такого центру є актуальною через низку причин: усе більша складність і масштаб КІ (сучасні об'єкти КІ генерують великі обсяги даних, що потребують ефективної обробки й аналізу; використання технологій Big Data, машинного навчання та інтернету речей (далі – IoT) дає можливість підвищити точність прогнозування аварійних ситуацій); необхідність забезпечення безперервності роботи критичних систем (відмова будь-якого елемента КІ може призвести до каскадних збоїв, значних економічних збитків і загрози для життя людей; централізовані й розподілені системи збору даних допомагають швидко реагувати на аномалії та запобігати критичним відмовам); високі вимоги до безпеки та стійкості (дані, що передаються й обробляються в центрі моніторингу, мають бути захищеними відповідно до міжнародних стандартів (ISO/IEC 27001, IEC 62443); використання 5G, Edge Computing і квантово-стійких алгоритмів шифрування підвищує стійкість системи до атак; інтеграція сучасних технологій зв'язку (5G, SDN (Software-Defined Networking), MEC (Multi-Access Edge Computing) дає змогу зменшити затримки в передачі даних і масштабувати інфраструктуру відповідно до вимог часу; використання блокчейн-технологій у центрі моніторингу забезпечує прозорість і захист даних); міжнародні й національні вимоги до моніторингу КІ (європейська Директива NIS2 (Network and Information Security) і національні законодавчі акти вимагають підвищеної стійкості КІ до кіберзагроз; запровадження єдиних стандартів для центрів збирання даних підвищить ефективність міжгалузевої взаємодії).

Таким чином, розробка методики формування структури центру збирання та оброблення даних є важливим кроком до підвищення безпеки й надійності критичної інфраструктури, що сприятиме своєчасному реагуванню на загрози та мінімізації ризиків для суспільства.

Розробка ефективної методики формування структури центру збирання та оброблення даних (далі – ЦЗОД) для моніторингу об'єктів КІ потребує комплексного підходу, що базується на сучасних дослідженнях у сфері IoT, обробки великих даних, 5G-комунікацій, edge computing і кібербезпеки.

У роботі [1] запропоновано модель побудови графа центрів оброблення даних на основі часових рядів сенсорних вимірювань, що дає змогу виявляти аномалії в поведінці КІ. Такий підхід є перспективним для динамічного моніторингу в режимі реального часу. Подібну архітектуру збору даних від рівня обладнання до рівня застосунків розглянуто в праці [2], де розроблено систему DCDB для модульного та масштабованого моніторингу.

З огляду на потребу в надійній обробці потокових IoT-даних, у роботі [3] представлено систему з використанням механізмів fault-tolerance, що забезпечує надійність передачі й оброблення інформації в умовах міських інфраструктур.

Поглиблений огляд типів даних і їх використання в системах оцінювання ризиків для КІ наведено в публікації [4]. Автори акцентують на необхідності стандартизації форматів даних і їх доступності для державних і комерційних структур.

На особливу увагу заслуговує дослідження [5], де обґрунтовано переваги застосування fog- і cloud-комп'ютингу в розподіленій обробці IoT-даних, що є ключовим при проєктуванні ЦЗОД у системах критичного моніторингу.

У публікації [6] представлено методику оцінювання загроз і ризиків для об'єктів критичної інфраструктури, яка ґрунтується на сценарному підході до аналізу розвитку надзвичайних ситуацій. Автори

обґрунтовують доцільність використання сценаріїв для ідентифікації потенційних загроз, оцінки їхніх наслідків і формування превентивних заходів із захисту інфраструктурних об'єктів. Запропонована методика дає змогу здійснювати комплексний аналіз уразливостей і ризиків з урахуванням специфіки об'єкта, типів загроз і можливих сценаріїв їх реалізації, що є актуальним для розробки ефективних систем моніторингу й реагування.

Проаналізовані джерела дають змогу виокремити основні напрями розвитку: використання розподілених обчислень (Edge/Fog Computing), упровадження 5G та IoT в архітектуру збору даних, застосування AI/ML для прогнозування відмов, а також розробку стандартизованих моделей взаємодії між елементами інфраструктури. На цьому тлі необхідною є розробка цілісної методики побудови структури ЦЗОД з урахуванням особливостей саме критичних об'єктів.

Мета роботи полягає в розробці методики формування структури ЦЗОД, орієнтованої на ефективний моніторинг стану об'єктів критичної інфраструктури з урахуванням сучасних інформаційно-комунікаційних технологій, типів сенсорних даних, вимог до надійності, масштабованості, безпеки та реального часу.

Виконання досліджень

Здійснимо опис структурно-функціональної схеми роботи ЦЗОД під час моніторингу стану об'єктів КІ. Запропонована модель відображає логічну структуру взаємодії компонентів системи збирання та оброблення даних у контексті моніторингу стану об'єктів критичної інфраструктури (далі – ОКІ). Вона базується на таких функціональних рівнях (рис. 1):

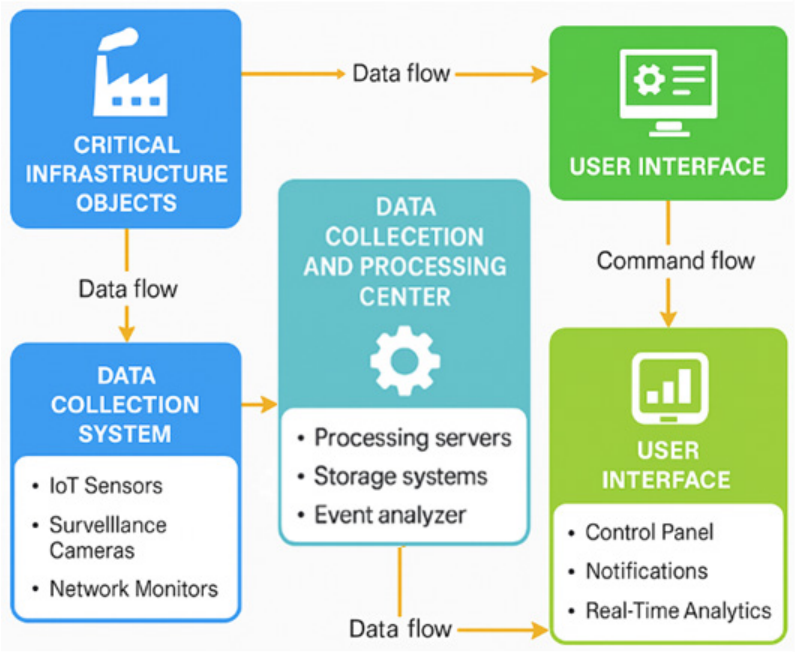


Рис. 1. Схема роботи ЦЗОД під час моніторингу стану об'єктів критичної інфраструктури

1. Об'єкти критичної інфраструктури. До цієї категорії належать інфраструктурні вузли, що мають високу стратегічну важливість: енергетичні системи, транспортні мережі, об'єкти водопостачання, аеропорти тощо. Моніторинг таких об'єктів здійснюється з метою виявлення аномалій, запобігання аваріям і забезпечення безперервності функціонування.
2. Система збору даних складається з набору сенсорів і пристроїв (датчики температури, тиску, вібрації, камери відеоспостереження, GPS-трекери тощо), які розміщуються на ОКІ та виконують безперервне зчитування параметрів стану. Дані формуються в реальному часі й передаються до шлюзу.
3. Комунікаційний шлюз (Communication Gateway) слугує проміжним етапом між сенсорами та центром обробки. Реалізує захищені канали зв'язку (VPN, TLS, IPsec) для передачі даних у режимі реального часу через мережі 5G, LTE, LPWAN, Wi-Fi або дротові інтерфейси.
4. ЦЗОД – основний обчислювальний осередок, де здійснюється агрегація та фільтрація даних; зберігання в централізованій базі (Data Lake або Data Warehouse); попередній і глибокий аналіз із використанням AI/ML-модулів (класифікація, кластеризація, виявлення аномалій); формування інцидентів і передача їх до системи прийняття рішень.
5. Користувацький інтерфейс/центр управління. Інтерфейс для операторів або автоматизованих систем, де здійснюється візуалізація показників; оперативне реагування; автоматичне або ручне управління інфраструктурними об'єктами на основі аналізу даних.

Представлена схема (рис. 1) дає змогу забезпечити масштабовану, адаптивну й безпечну структуру системи моніторингу, яка здатна ефективно функціонувати в реальному часі. Її реалізація забезпечує підвищення якості обслуговування ОКІ та зниження ризиків техногенних або кіберінцидентів.

Підходи, методи, технології та стандарти до побудови ЦЗОД:

1. Підходи до побудови ЦЗОД.

Централізований підхід передбачає використання єдиного центру обробки даних (Data Processing Center, DPC), куди стікається вся інформація. Переваги: спрощене адміністрування, єдина точка контролю безпеки. Недоліки: висока затримка при передачі даних, вразливість до відмови.

Розподілений підхід базується на кількох вузлах обробки даних, які розташовані ближче до джерел даних. Переваги: підвищена відмовостійкість, зменшення затримки. Недоліки: складніше управління, вища вартість інфраструктури.

1.2. Хмарні, гібридні й локальні рішення. Хмарні рішення (AWS, Microsoft Azure, Google Cloud) забезпечують масштабованість і доступність. Гібридні рішення комбінують локальну інфраструктуру з хмарними сервісами, що забезпечує баланс між швидкістю й безпекою. Локальні рішення (on-premise) використовуються в ОКІ, де є жорсткі вимоги до безпеки.

1.3. Використання Edge Computing дає змогу обробляти дані безпосередньо на пристроях або проміжних вузлах перед передачею до основного центру. Це зменшує затримки й навантаження на мережу.

2. Методи оброблення й аналізу даних.

2.1. Машинне навчання для виявлення аномалій використовується для аналізу показників ОКІ та виявлення потенційних відмов. Основні методи: кластеризація (k-Means, DBSCAN), нейронні мережі (Autoencoder, LSTM), байєсівські моделі (Bayesian Networks).

2.2. Фільтрація й агрегування даних: алгоритм Калмана – для прогнозування стану систем, фільтрація за пороговими значеннями – спрощує аналіз даних, методи зменшення шуму – наприклад, Wavelet-трансформація.

2.3. Використання Big Data: Hadoop, Apache Spark для обробки великих обсягів даних, Cassandra, InfluxDB – для зберігання часових рядів.

3. Технології.

3.1. IoT-сенсори, датчики температури, вологості, тиску, вібрації, електромагнітного випромінювання, які збирають дані з ОКІ.

3.2. Бази даних: SQL (PostgreSQL, MySQL) – для структурованих даних, NoSQL (MongoDB, Cassandra) – для масштабованих сховищ, Time-Series DB (InfluxDB, TimescaleDB) – для потокових даних.

3.3. Протоколи передачі даних: MQTT – легкий протокол для IoT, CoAP – протокол для пристроїв з низьким енергоспоживанням, OPC UA – для промислових систем.

3.4. Контейнеризація й оркестрація: Docker – контейнеризація сервісів, Kubernetes – управління контейнерами для масштабованих систем.

4. Стандарти й нормативні вимоги.

4.1. Безпека інформації: ISO/IEC 27001 – забезпечення безпеки інформації, IEC 62443 – кібербезпека для промислових систем.

4.2. Моніторинг і надійність: ISO 22301 – управління безперервністю бізнесу, ITU-T Y.3053 – архітектури для моніторингу КІ.

Перелік сенсорів для моніторингу ОКІ.

Об’єкти критичної інфраструктури охоплюють різні галузі: енергетику, водопостачання, транспорт, зв’язок, оборону тощо. Для кожної сфери використовуються спеціалізовані сенсори для моніторингу стану об’єктів (рис. 2).

1. Сенсори фізичних параметрів.

1.1. Температурні сенсори: термопари (K, J, T, E типи) – для вимірювання високих і низьких температур, RTD (Platinum Resistance Temperature Detectors, PT100/PT1000) – точні вимірювання, ІЧ-пірометри – безконтактний моніторинг температури.

1.2. Вологоміри та гігрометри: ємнісні сенсори вологості – для середовища (вода, повітря), резистивні датчики – контроль рівня вологості.

1.3. Барометри (датчики тиску): абсолютні датчики тиску – атмосферний тиск, диференціальні датчики – моніторинг різниці тисків у системах, п’єзоелектричні датчики – високоточний контроль.

1.4. Сенсори вібрації та механічних коливань: акселерометри MEMS – контроль структурних коливань, геофони – сейсмічний моніторинг, п’єзоелектричні вібраційні сенсори – контроль механічного зносу.

2. Сенсори для моніторингу навколишнього середовища.

2.1. Сенсори газового складу: детектори чадного газу (CO), датчики метану (CH_4), пропану (C_3H_8), бутану (C_4H_{10}), аналізатори кисню (O_2), озону (O_3), азоту, VOC-сенсори (леткі органічні сполуки).

2.2. Радіаційні сенсори: Гейгер-Мюллерові лічильники – гамма, бета-випромінювання, напівпровідникові детектори – альфа, нейтронне випромінювання, сцинтиляційні датчики – спектральний аналіз радіації.

2.3 Хімічні аналізатори: іон-селективні електроди (ISE) – аналіз складу води, оптичні спектрометри – контроль забруднень.

3. Сенсори для моніторингу енергетичних об'єктів.

3.1. Електричні сенсори: напругоміри (Voltage Sensors) – контроль високої та низької напруги, амперметри (Current Sensors) – вимірювання струму, трансформатори струму (CT Sensors) – контроль змінного струму.

3.2. Сенсори потужності й енергії: датчики активної та реактивної потужності, лічильники електроенергії (Smart Meters).

3.3. Теплові сенсори: тепловізори – контроль нагріву електрообладнання, оптичні пірометри – безконтактний контроль перегріву.

4. Сенсори для моніторингу транспортної інфраструктури.

4.1. Датчики навантаження й напруги: тензодатчики (Strain Gauges) – контроль мостів, доріг, п'єзоелектричні сенсори – вимірювання тиску на поверхню.

4.2. GPS і GNSS-сенсори: приймачі GPS/GLONASS/Galileo – відстеження руху, IMU (Inertial Measurement Unit) – тривимірний моніторинг руху,

4.3. Контроль залізничної інфраструктури: лазерні сенсори деформацій, сенсори температури рейок.

5. Сенсори для кібербезпеки й телекомунікацій.

5.1. Сенсори мережевого трафіку: DPI (Deep Packet Inspection) – аналіз трафіку, Intrusion Detection Systems (IDS) – виявлення атак.

5.2. Фізичні сенсори захисту: детектори руху (PIR, ультразвукові), камери відеоспостереження (LiDAR, тепловізійні).

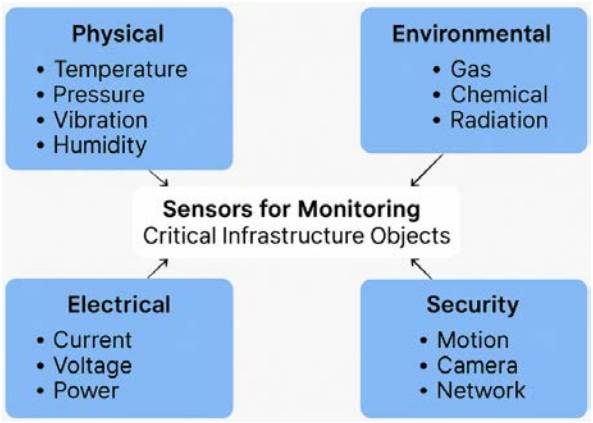


Рис. 2. Сенсори для моніторингу ОКІ

Передача показників з датчиків за допомогою 5G.

Передача показників з датчиків за допомогою 5G є одним із ключових сценаріїв використання технології. 5G забезпечує низьку затримку (1–10 мс) – важливо для моніторингу в реальному часі; високу пропускну здатність – дає змогу передавати великі масиви даних, наприклад, відео з камер або тепловізорів; масштабованість – підтримка до 1 мільйона пристроїв на 1 км² (mMTC); якість обслуговування (QoS) – можливість пріоритезації критичних даних.

Основні 5G-режими передачі даних: eMBB (Enhanced Mobile Broadband) – для камер відеоспостереження, тепловізорів, радарів; mMTC (Massive Machine-Type Communications) – для IoT-сенсорів (температури, вологості, вібрації, газів); URLLC (Ultra-Reliable Low-Latency Communications) – для критично важливих даних (енергетична інфраструктура, залізниця, кібербезпека).

Протоколи для передачі даних по 5G: MQTT (Message Queuing Telemetry Transport) – легкий протокол для IoT; CoAP (Constrained Application Protocol) – ефективний для низькоенергетичних пристроїв; OPC UA (Open Platform Communications Unified Architecture) – стандарт для промислових сенсорних мереж; HTTP/HTTPS + REST API – для інтеграції з вебсервісами.

Приклади використання 5G для сенсорів: енергетика (передача показників трансформаторів, генераторів, розподільчих щитів у режимі реального часу); транспорт (моніторинг залізничних колій, мостів через 5G-з'єднані тензодатчики); промисловість (контроль стану обладнання на заводах через підключені по 5G сенсори вібрації та температури); безпека (передача відео з камер і тепловізорів у центри обробки даних без затримок). 5G забезпечує ефективну та швидку передачу даних від сенсорів, що дає можливість будувати розумні системи моніторингу для КІ.

Математичний опис переліку сенсорів для моніторингу ОКІ на основі теорії множин.

Розглянемо множину S усіх сенсорів, що використовуються для моніторингу ОКІ. Цю множину можна представити як об'єднання підмножин, що відповідають різним сферам моніторингу:

$$S = S_{фіз} \cup S_{екол} \cup S_{енерг} \cup S_{трансп} \cup S_{безп} \ , \quad (1)$$

де $S_{фіз}$ – сенсори фізичних параметрів (температура, тиск, вібрація тощо); $S_{екол}$ – сенсори екологічного моніторингу (газові, хімічні, радіаційні); $S_{енерг}$ – сенсори енергетичної інфраструктури (струм, напруга, потужність); $S_{трансп}$ – сенсори для транспорту (навантаження, GPS, дорожній стан); $S_{безп}$ – сенсори для безпеки та кібербезпеки (камери, мережеві IDS).

Сенсори фізичних параметрів:

$$S_{фіз} = \{S_T, S_P, S_V, S_A, S_H\} , \quad (2)$$

де S_T – температурні сенсори, S_P – сенсори тиску, S_V – вібраційні сенсори, S_A – акселерометри, S_H – сенсори вологості.

Сенсори екологічного моніторингу:

$$S_{екол} = \{S_{CO}, S_{CH4}, S_{O2}, S_{VOC}, S_{рад}\} , \quad (3)$$

де S_{CO} – датчики чадного газу, S_{CH4} – детектори метану, S_{O2} – аналізатори кисню, S_{VOC} – сенсори летких органічних сполук, $S_{рад}$ – радіаційні сенсори.

Сенсори енергетичної інфраструктури:

$$S_{енерг} = \{S_V, S_I, S_P, S_Q\} , \quad (4)$$

де S_V – сенсори напруги, S_I – сенсори струму, S_P – датчики активної потужності, S_Q – датчики реактивної потужності.

Сенсори транспортної інфраструктури:

$$S_{трансп} = \{S_{GPS}, S_{IMU}, S_{strain}\} , \quad (5)$$

де S_{GPS} – GPS-приймачі, S_{IMU} – інерційні сенсори, S_{strain} – датчики напружень.

Сенсори безпеки та кібербезпеки:

$$S_{безп} = \{S_{PIR}, S_{CAM}, S_{IDS}\} , \quad (6)$$

де S_{PIR} – датчики руху, S_{CAM} – відеокамери, S_{IDS} – мережеві сенсори.

Множина сенсорів, які можуть передавати дані через 5G, визначається так:

$$S_{5G} = \{S_{високошвидк}, S_{масові IoT}, S_{критичні}\} , \quad (7)$$

де $S_{високошвидк}$ – сенсори, що потребують високої пропускної здатності (наприклад, камери, тепловізори), $S_{масові IoT}$ – численні сенсори для моніторингу (температури, вологості, газу), $S_{критичні}$ – сенсори для управління аварійними ситуаціями з мінімальною затримкою.

$$S_{5G} = \{S_{CAM}, S_{GPS}, S_{IMU}, S_{VOC}, S_{CO}, S_{PIR}, S_{IDS}, S_{strain}, S_{рад}\} , \quad (8)$$

Отже, множина сенсорів, що використовують 5G, є підмножиною загальної множини S, тобто $S_{5G} \subseteq S$.

Математичний опис процесу збору, передачі й обробки сенсорних даних із використанням технології 5G на прикладі ОКІ енергетичної підстанції:

1. Модель об'єкта як графа. Розглянемо підстанцію як граф:

$G = (V, E)$, де $V=\{v_1, v_2, \ldots, v_n\}$ – вузли (елементи: трансформатори, вимикачі, захисти, акумулятори), E – зв'язки (електричні або інформаційні).

Кожен вузол обладнаний сенсорами: $S=\{s_1, s_2, \ldots, s_m\}$ — множина сенсорів: s_1 – температура трансформатора; s_2 – рівень напруги; s_3 – струм витоку; s_4 – вібрація корпуса; s_5 – наявність диму/газу.

2. Процес збирання та передачі даних через 5G.

2.1. Формалізація даних.

Кожен сенсор s_i створює вимір: $d_i(t) \in R,\ t \in T,$ де T – дискретний часовий простір.

Узагальнено: $D(t)=[d_1(t), d_2(t), \ldots, d_m(t)]^T$.

2.2. Передача даних.

Передача сенсорних даних здійснюється через 5G з такими параметрами:

затримка: $\tau \approx 1-5$ мс;

пропускна здатність: $B \geq 10$ Гбіт/с;

надійність: $R \geq 99.999\%$;

кількість пристроїв: $N \leq 10^6$/км^2.

Модель каналу можна подати так: $y(t)=h(t)\cdot x(t)+n(t)$, де $x(t)$ – сигнал від сенсора; $h(t)$ – коефіцієнт каналу 5G (може бути стохастичним); $n(t)$ – шум (наприклад, гаусівський).

3. Edge-Fog-Cloud обробка.

3.1. Рівні обробки.

Edge (периферійні шлюзи): $f_{edge}(D(t)) = D'(t) \Rightarrow$ попередня обробка: фільтрація, нормалізація, тригер.

Fog: $f_{fog}(D')$ = агрегація даних, локальна аналітика, кластеризація.
Cloud: $f_{cloud}(D')$ = глобальна аналітика, історичний аналіз, машинне навчання.
4. Модель затримки передачі та обробки.
Загальний час обробки сенсорного запиту: $T_{total}=T_{sense}+T_{tx}+T_{edge}+T_{fog}+T_{cloud}$, де T_{sense} – час збору; $T_{tx}\approx\tau$ – час передачі через 5G; T_{edge} – обробка на шлюзі; T_{fog} – обробка на ближчому обчислювальному вузлі; T_{cloud} – аналітика в хмарі.
Можна оптимізувати T_{total} з урахуванням критичності подій: $min\ T_{total}$ за умовами $D'(t) >$ порогу.
5. Виявлення аномалій (напр., машинне навчання).
Формалізуємо вхідний вектор сенсорних даних: $\vec{X} = [d_1(t), d_2(t), ..., d_m(t)]$.
Класифікація стану: $y = f_{ML}(\vec{X})$, $y \in$ *{«норма», «відхилення», «аварія»}*.
Можна використати, наприклад, алгоритм: SVM; Random Forest; Autoencoder для аномалій.
6. Формалізація подій.
Множина подій: $\Omega=\{\omega_1,\omega_2,...,\omega_k\}$.
Наприклад: ω_1 – перевищення температури; ω_2 – утрата живлення; ω_3 – вторгнення в приміщення.
Тоді можна ввести матрицю реакцій: $A \in \{0,1\}^{k\times r}$, де $a_{ij}=1$, якщо подія ω_i вимагає реакції r_j.

Програмна реалізація на Java (фрагмент)

```
class Sensor {
    String id;
    double readValue() {
        return Math.random() * 100; // симуляція
    }
}

class Node {
    String name;
    List<Sensor> sensors;

    public double[] readAllSensors() {
        return sensors.stream().mapToDouble(Sensor::readValue).toArray();
    }
}

class Transmission5G {
    public double[] transmit(double[] data) {
        // моделювання затримки та шуму
        return Arrays.stream(data)
                     .map(d -> d + new Random().nextGaussian() * 0.1)
                     .toArray();
    }
}

class EdgeProcessor {
    public double[] preprocess(double[] data) {
        return Arrays.stream(data)
                     .filter(d -> d > 5.0) // умовна фільтрація
                     .toArray();
    }
}

class FogNode {
    public double aggregate(double[] data) {
        return Arrays.stream(data).average().orElse(0);
    }
}

class CloudAnalytics {
    public String classify(double value) {
        if (value > 90) return "Аварія";
        if (value > 60) return "Відхилення";
        return "Норма";
    }
}
```

Таблиця 1

Характеристики передачі та обробки даних на етапах системи моніторингу

Показник	Sensor → Edge	Edge → Fog	Fog → Cloud	**Коментарі**
Затримка передачі (мс)	1,0	1,2	0,8	Модельні значення з урахуванням 5G URLLC
Надійність зв'язку (%)	99,9	99,99	99,95	Локальний зв'язок забезпечує високу надійність; іноді не враховується окремо
Об'єм переданих даних (МБ/год)	120	100	250	Зниження обсягу на проміжному рівні за рахунок фільтрації
Частота подій (разів/год)	6	N/A	N/A	Визначається лише на рівні сенсорів
Попереднє оброблення	✓	✓	–	Edge і Fog здійснюють фільтрацію, нормалізацію, оцінку подій
Шифрування	—	✓	✓	Рекомендовано впровадити легке шифрування на сенсорах
Тип обробки	Локальна (Signal)	Агрегація	Аналітика/ML	Зростає складність та обчислювальне навантаження з кожним рівнем

Примітки:

– **N/A** – *Not Applicable, не застосовується для даного рівня;*

– *символ ✓ – функціональність присутня;*

– *символ «–» означає відсутність або неактуальність на відповідному етапі;*

– *усі показники є умовно типовими для енергетичних об'єктів із підтримкою 5G-зв'язку.*

У таблиці 1 наведено узагальнені характеристики роботи математичної моделі збору та обробки даних на прикладі енергетичної підстанції. Розглянуто основні показники для трьох ключових етапів інформаційного ланцюга: від сенсорного рівня (Sensor → Edge), проміжного рівня обробки (Edge → Fog) і передачі до хмарної інфраструктури (Fog → Cloud).

Затримка передачі є критичним параметром для своєчасного реагування на події. Завдяки використанню технологій 5G (режим URLLC), забезпечується низький рівень затримки, особливо на початкових етапах.

Надійність зв'язку зростає в напрямку до хмари, оскільки на рівні Fog і Cloud використовуються більш стабільні канали (оптоволоконні або приватні 5G-мережі). Значення можуть змінюватися залежно від конкретної реалізації.

Обсяг переданих даних зменшується на етапі Edge → Fog за рахунок попередньої обробки (агрегації, фільтрації шумів, усунення надлишкових даних), але збільшується на Fog → Cloud через накопичення історичних даних і їх реплікацію для аналітичної обробки.

Частота подій ураховується лише на рівні сенсорів, оскільки саме там генеруються сповіщення про аномалії або зміни у фізичних параметрах. На вищих рівнях здійснюється лише агрегація або повторне оцінювання їхньої важливості.

Попереднє оброблення реалізується на рівнях Edge та Fog і передбачає виявлення нетипових змін, нормалізацію даних і їх маркування. У хмарній інфраструктурі переважно виконується аналітична обробка, зокрема із застосуванням моделей машинного навчання (ML).

Шифрування впроваджено на етапах Fog → Cloud та Edge → Fog, що забезпечує конфіденційність даних на критичних ділянках. На рівні Sensor → Edge воно відсутнє через обмеження обчислювальних ресурсів сенсорних пристроїв, однак доцільно впровадити легкі криптографічні алгоритми.

Тип обробки на кожному етапі відрізняється: на Sensor та Edge — сигналізація й локальна обробка, на Fog – попередній аналіз та агрегування, на Cloud – глибокий аналіз, виявлення закономірностей, трендів, а також прийняття рішень у рамках систем підтримки.

Висновок

У статті розглянуто комплексний підхід до формування структури ЦЗОД, призначеного для забезпечення ефективного та надійного моніторингу стану ОКІ. Аналіз наявних підходів, технологій і стандартів дав змогу систематизувати перелік сенсорів, які забезпечують багаторівневий моніторинг об'єктів КІ – від фізичних параметрів (температура, тиск, вібрація) до кіберподій (спроби вторгнень, аномальні дії в мережах); оцінити можливості використання технологій 5G для забезпечення високошвидкісного, надійного та масштабованого з'єднання сенсорів і вузлів обробки даних у системах моніторингу; запропонувати математичну модель, що базується на теорії множин, для класифікації

сенсорних потоків відповідно до типу об'єктів КІ та їхніх функціональних характеристик; обґрунтувати доцільність використання гібридної архітектури, яка поєднує edge-, fog- і cloud-технології для обробки даних у реальному часі, а також довгострокового зберігання й аналітики; окреслити структуру ЦЗОД, що враховує вимоги до надійності, відмовостійкості, кібербезпеки та масштабованості, відповідно до особливостей ОКІ; сформулювати наукову новизну, яка полягає у створенні методики, що дає змогу оптимізувати структуру ЦЗОД з урахуванням сучасних технологічних можливостей, характеристик сенсорних потоків і специфіки об'єктів КІ.

Отримані результати можуть бути використані як основа для практичної реалізації центрів моніторингу ОКІ в критично важливих галузях, таких як енергетика, транспорт, водопостачання, зв'язок та оборонна промисловість. Подальші дослідження будуть присвячені моделюванню навантаження на ЦЗОД, а також упровадженню штучного інтелекту для автоматизованого аналізу зібраних даних.

Література

1. Zhang H., Li Z., Ren Z. Data-driven construction of data center graph of things for anomaly detection. *arXiv preprint arXiv:2004.12540*. 2020. https://arxiv.org/abs/2004.12540.

2. Netti A., Mueller M., Auweter A., Guillen C., Ott M., Tafani D., Schulz M. From facility to application sensor data: Modular, continuous and holistic monitoring with DCDB. *arXiv preprint arXiv:1906.07509*. 2019. https://arxiv.org/abs/1906.07509.

3. Morgan K. Geldenhuys, Jonathan Will, Benjamin J. J. Pfister, Martin Haug, Alexander Scharmann, Lauritz Thamsen. Dependable IoT data stream processing for monitoring and control of urban infrastructures. *arXiv preprint arXiv:2108.10721*. 2021. https://arxiv.org/abs/2108.10721.

4. Aron Larsson. Data use and data needs in critical infrastructure risk analysis. *Journal of Risk Research.* 2023. № 26(2). P. 190–210. https://doi.org/10.1080/13669877.2023.2181858.

5. Yassine A., Singh S., Hossain M. S., Muhammad G. (2019). IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems.* 2019. № 91. P. 563–573. https://doi.org/10.1016/j.future.2018.08.040.

6. Мурасов Р., Нікітін А., Мещеряков І., Підгородецький М., Поплавець С. Методика оцінювання загроз і ризиків для об'єктів критичної інфраструктури за сценаріями розвитку надзвичайних ситуацій. *Сучасні інформаційні технології у сфері безпеки та оборони*. Київ, Україна, 2024. № 48(3). С. 35–43. https://doi.org/10.33099/2311-7249/2023-48-3-35-43.

References

1. Zhang, H., Li, Z., & Ren, Z. (2020). Data-driven construction of data center graph of things for anomaly detection. *arXiv preprint arXiv:2004.12540*. https://arxiv.org/abs/2004.12540.

2. Netti, A., Mueller, M., Auweter, A., Guillen C., Ott M., Tafani D., Schulz M. (2019). From facility to application sensor data: Modular, continuous and holistic monitoring with DCDB. *arXiv preprint arXiv:1906.07509*. https://arxiv.org/abs/1906.07509.

3. Morgan K. Geldenhuys, Jonathan Will, Benjamin J. J. Pfister, Martin Haug, Alexander Scharmann, Lauritz Thamsen (2021). Dependable IoT data stream processing for monitoring and control of urban infrastructures. *arXiv preprint arXiv:2108.10721*. https://arxiv.org/abs/2108.10721.

4. Aron Larsson (2023). Data use and data needs in critical infrastructure risk analysis. *Journal of Risk Research, 26*(2), 190–210. https://doi.org/10.1080/13669877.2023.2181858.

5. Yassine, A., Singh, S., Hossain, M. S., & Muhammad, G. (2019). IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems, 91*, 563–573. https://doi.org/10.1016/j.future.2018.08.040.

6. Murasov R., Nikitin A., Meshcheriakov I., Pidhorodetskyi M., Poplavets S. (2024) "Methodology for assessing threats and risks for critical infrastructure objects under emergency development scenarios", Modern Information Technologies in the Sphere of Security and Defence. Kyiv, Ukraine, 48(3), pp. 35–43. https://doi.org/10.33099/2311-7249/2023-48-3-35-43.

Дата першого надходження рукопису до видання: 12.05.2025
Дата прийнятого до друку рукопису після рецензування: 16.06.2025
Дата публікації: 25.07.2025

УДК 792.01: 778.534.4
DOI https://doi.org/10.36994/2788-5518-2025-01-09-02

ДОСЛІДЖЕННЯ ІНДИВІДУАЛЬНИХ ВЛАСТИВОСТЕЙ ЗВУКОРЕЖИСЕРА ЩОДО СПРИЙНЯТТЯ ЗВУКОВИХ СИГНАЛІВ ОРГАНОМ СЛУХУ

Гребінь О. П., *к.т.н., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0003-4180-1720, alexgstudio.2016@gmail.com*

Левенець Н. Ф., *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0003-0984-7622, n.levenets.ua@gmail.com*

Продеус А. М., *д.т.н., проф., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0001-7640-085,: pam51335-ames@lll.kpi.ua*

Швайченко В. Б., *к.т.н., доц., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0001-9736-0800 vbs2011@ukr.net*

Анотація. *Метою роботи є розроблення досить простого способу обстеження слухових вад для працівників звукової індустрії, зокрема звукорежисерів. Дослідження має на меті не точне визначення вад слуху, а визначення необхідної корекції частотної характеристики звуковідтворення для якісного сприйняття звукового контенту. Методика дослідження ґрунтується на вимірюванні рівня гучності на порозі чутності й частотної характеристики сприйняття звукових сигналів на порозі чутності. Результати дослідження слухового сприйняття звукових сигналів одним із авторів і групи слухачів з 35 осіб для визначення вад слуху свідчать, що серед реальних 28 результатів відхилення від теоретичної кривої порогу чутності в межах 10…20 дБ у звуковому діапазоні, що вважають відсутністю вад слуху за медичними показниками, мали 20 осіб, більше ніж 20 дБ на окремих частотах звукового діапазону мали 7 осіб, а більше ніж 20 дБ у широких смугах частот – 1 особа, ці особливості обов'язково варто враховувати при відтворенні фонограм для якісного сприйняття. Дослідження слухового сприйняття одного з авторів запропонованим методом доводить, що форма частотної характеристики сприйняття звукових сигналів збігається за формою з теоретичною кривою однакової гучності на порозі чутності, однак на частотах максимальної чутливості слуху (2–4 кГц), чутливість слуху цієї особи значно погіршена, також сприйняття високих частот, починаючи вже з частоти 1 кГц, теж має вади. Наукова новизна полягає в удосконаленні методу визначення вад слуху осіб, яких долучено до визначення якості сприйняття аудіоконтенту. Практична цінність полягає в тому, що враховано звичну для звукорежисерів навколишню обстановку з усіма акустичними перевагами й недоліками, попереднє встановлення оптимальної гучності для сигналу та контенту, за яким «навчався» слухач, забезпечено можливість проведення досліджень у всьому частотному діапазоні сприйняття звуку людиною й у разі виявлення недоліків (вад) слуху більш детально визначено частоти, на яких проводиться вимірювання, причому завжди можна прослухати сигнал з достатньою гучністю й очікувати саме цей сигнал на порозі чутності. Також визначено вимоги до обладнання, які обов'язково варто враховувати під час відтворення фонограм для якісного сприйняття.*

Ключові слова: *аудіограма; вади слуху; звуковий контент; звукорежисер; мінімальний рівень; поріг чутності.*

RESEARCH OF INDIVIDUAL PROPERTIES OF SOUND DESIGNERS IN RELATION TO THE PERCEPTION OF SOUND SIGNALS BY THE HEARING

Oleksandr Grebin, *Ph.D., Department of Acoustic and Multimedia Electronic Systems of the Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine. https://orcid.org/0000-0003-4180-1720, alexgstudio.2016@gmail.com*

Ninel Levenets, *Department of Acoustic and Multimedia Electronic Systems of the Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine. https://orcid.org/0000-0003-0984-7622, n.levenets.ua@gmail.com*

Arkadii Prodeus, *D.Sc., Prof., Department of Acoustic and Multimedia Electronic Systems of the Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine. https://orcid.org/0000-0001-7640-0850, pam51335-ames@lll.kpi.ua*

Volodymyr Shvaichenko, *Department of Acoustic and Multimedia Electronic Systems of the Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine. https://orcid.org/0000-0001-9736-0800, vbs2011@ukr.net*

Abstract. *The aim of the research is to develop a fairly simple method for examining hearing impairments for workers in the audio industry, in particular audio engineers. The aim of the research is not to accurately determine hearing impairments, but to determine the necessary correction of the frequency response of*

sound reproduction for high-quality perception of sound content. The research is based on measuring the loudness level at the threshold of hearing and the frequency response of the perception of sound signals at the threshold of hearing. A study of auditory perception of sound signals by one of the authors and a group of listeners of 35 people to determine hearing impairments shows that among the real 28 results, 20 people had deviations from the theoretical threshold curve of audibility within 10…20 dB in the sound range, which is considered the absence of hearing impairments according to medical indicators, 7 people had more than 20 dB at individual frequencies of the sound range, and 1 person had more than 20 dB in wide frequency bands. These features must be taken into account when reproducing phonograms for high-quality perception. The study of auditory perception by one of the authors using the proposed method proves that the shape of the frequency response of the perception of sound signals coincides in shape with the theoretical curve of the same loudness at the threshold of audibility, however, at the frequencies of maximum hearing sensitivity (2–4 kHz), the hearing sensitivity of this person is significantly impaired, and the perception of high frequencies, starting from a frequency of 1 kHz, also has defects. The method for determining hearing impairments of individuals involved in determining the quality of perception of audio content has been improved. The environment familiar to sound engineers with all its acoustic advantages and disadvantages has been taken into account, the optimal volume has been previously set, and for the signal and content by which the listener was "trained", the possibility of conducting research in the entire frequency range of human sound perception has been ensured, and when hearing deficiencies (defects) are detected, the frequencies at which measurements are performed have been determined in more detail, and it is always possible to listen to a signal with sufficient volume and expect this signal at the threshold of audibility. Equipment requirements have also been determined, which must be taken into account when reproducing phonograms for high-quality perception.

Key words: *audiogram; hearing impairment; sound content; sound engineer; minimum level; threshold of hearing.*

Вступ

Робота звукорежисера в будь-яких своїх проявах – це робота зі звуком, формуванням звукового контенту, звукових програм і завжди пов'язана з прослуховування звукового контенту. Якщо переглянути визначення звукорежисера – це працівник звукової індустрії, який здійснює виробництво звукового контенту як незалежного продукту, так і звукового супроводу для аудіовізуального контенту – телевізійних програм, кінофільмів тощо [1; 2].

Звукорежисер – один із провідних учасників творчого процесу, пов'язаного з формуванням звуку у відповідних видах і жанрах музичного мистецтва, є повноправним учасником творчого процесу формування не лише технічного, а й художнього складника звукового контенту. Звукорежисер визначає умови створення та кінцевий результат звукового образу, зафіксованого на носії звукозапису [3].

Звукорежисер керує всім процесом створення звукового контенту, створенням звукової картини загалом, відповідає за кінцевий результат створеного звукового матеріалу. Звукорежисер у процесі виробництва звукового контенту виконує такі операції зі звуком: установлює необхідне підсилення сигналу до оптимального рівня, забезпечує обробку сигналу за частотою, зокрема формує тембральне забарвлення, формує різноманітні ефекти у звучанні, наприклад, за рахунок затримок сигналу, зсув фази сигналу тощо. Усі ці ефекти та зміни в сигналі звукорежисер повинен чути.

Із поширенням якісного телебачення, радіомовлення, звукозапису, ураховуючи розмаїття музичних жанрів, радіофікації різноманітних майданчиків – від концертних залів до торговельних центрів тощо, суттєво розширено сферу діяльності звукорежисера, що робить його працю масовою й зумовлює подальшу диференціацію. Виникає необхідність у концертних, студійних, театральних, постпродакшн, власних звукорежисерів музичних груп, звукорежисерів торговельно-розважальних закладів тощо [4].

Звичайна робота звукоінженера або навіть звукотехніка в сучасному світі потребує й деяких здібностей звукорежисера, тобто творчого складника в роботі зі звуком. Багато вищих навчальних закладів стали готувати звукорежисерів, адже робота звукорежисера вимагає достатнього рівня знань технічної точки зору, знання звуку й основ його створення, знання акустики, знання технологічного обладнання тощо, бути музично освіченим, а головне, глибоко розуміти сутність процесу виробництва звукового контенту.

Основна особливість звукорежисера, що пов'язана з опрацюванням звукового контенту, є його спроможність слухати й чути звук у будь-якому прояві. Звукорежисер має вміти аналізувати звуковий контент, виділяти з нього корисний контент і завади, визначати спотворення звукового сигналу, неточності у відтворенні, відчувати простір звукового образу, зосереджуватися на окремих звуках, створюваних інструментами й іншими джерелами звуку, відчувати музичний баланс тощо – усе це формує професійний складник звукорежисера [5].

Щоб чути звук і коректно його сприймати, зрозуміло, звукорежисер має чути сигнали припустимо мінімальної гучності й у частотному діапазоні, визначеному природою, тобто бути повністю здоровим щодо сприйняття звуку слуховим органом. Звукорежисер з вадами слуху, зрозуміло, не зможе здійснювати якісне виробництво контенту, адже він буде налаштовувати звучання відповідно до свого сприйняття, що для здорового слухача може призводити до відчуття наявності дефектів у звучанні.

Робота звукорежисера переважно пов'язана з контролем звукового сигналу в акустичному просторі за допомогою гучномовців – звукових моніторів, причому рівень гучності, за яким відбувається прослуховування, становить 90–95 дБ (94 дБ – відповідає звуковому тиску в 1 Па). Часто для більш детального контролю звукового контенту застосовують навушники. Тому для визначення слухових властивостей звукорежисера важливо розуміти власне сприйняття звукового контенту в різних умовах.

Деякі звукорежисери, зокрема молодь, уважають, що їм не потрібне дослідження вад слуху, що їхнє сприйняття звукового контенту є досить досконалим. Разом із тим, слухаючи весь вільний час доби звуковий контент у навушниках з невизначеною гучністю, при цьому вважаючи свої навушники найякіснішими серед будь-яких інших, не завжди розуміють, що і як повинно звучати, головне, не враховують, що можуть утрачати слух. Але це не так. Слух – найважливіший інструмент звукорежисера, як музичний інструмент для музиканта – може з часом погіршувати свої властивості [6].

З часом слухові властивості людини, звукорежисера в тому числі, змінюються в бік погіршення. Діапазон частот, що сприймає людина, з віком звужується, високі частоти сприймаються гірше, нижні частоти не дають щільності, дещо «розмазуються». На жаль, відновити слухові відчуття неможливо. Але можливо знати ці вади й ураховувати їх при формуванні звукового контенту.

Для дослідження слухових властивостей людини в медичній практиці використовують метод аудіометрії, який дає змогу визначати спроможність людини сприймати звуки різного походження, зокрема тональні звуки у визначеній частині звукового частотного діапазону, і за результатами дослідження зазначати вади слуху. Після дослідження (обстеження) будують аудіограму як графічне відображення сприйняття звуків людиною [7]. Вимірювання аудіограми проводить у медичних закладах, що спеціалізуються на дослідженні слуху, лікар-отоларинголог, аудіолог. Важливо зазначити, що аудіометрія передусім передбачає визначення вад слуху людини, а не якість сприйняття, що необхідно для людей, професійною діяльністю яких є, наприклад, виробництво звукового контенту, тобто звукорежисерів. Саме в цьому полягає один із недоліків цього методу.

Основою для визначення гостроти слуху, тобто того, з чим порівнюють спроможність людини щодо сприйняття тональних звуків різної частоти на порозі чутності, є крива однакової гучності на порозі чутності [8].

Гостроту слуху досліджують методами повітряної та кісткової провідності зазвичай на частотах від 125 Гц до 8000 Гц та рівнем звуку до 120 дБ (повітряна провідність) і до 70 дБ (кісткова провідність).

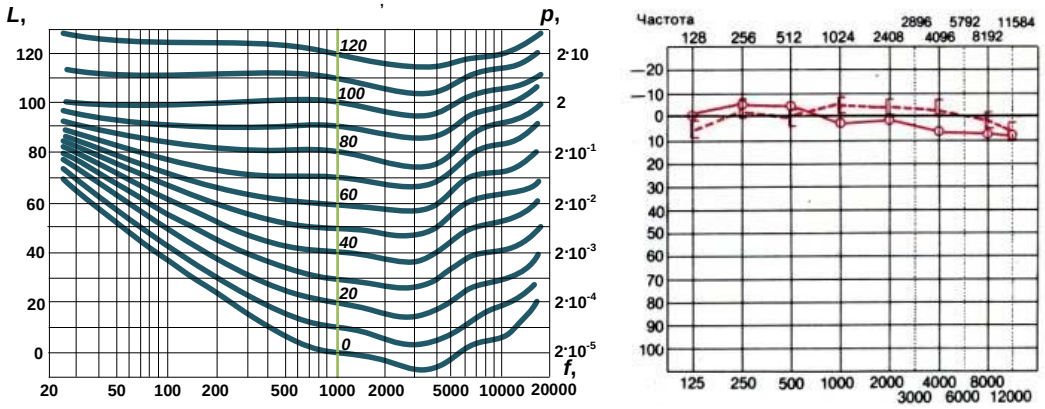
Методика дослідження слуху, що проводить аудіолог, передбачає такі дії. Особу розміщують у заглушеній акустичній камері, а на слуховий орган (вуха) надягають закриті навушники. З генератора тестових сигналів, що входить до складу аудіометра, у навушники подають сигнали відповідних частот у діапазоні від 125 Гц до 8 кГц зі збільшенням гучності. У момент, коли піддослідний починає чути звук відповідної частоти, він натискає кнопку, а аудіолог відмічає значення рівня сигналу відповідної частоти. Потім аудіолог змінює частоту й знову спрямовує сигнал у навушники, змінюючи гучність. Таке дослідження проводять на всіх необхідних частотах. Після закінчення обстеження будується аудіограма для кожного вуха й обох вух одночасно, тобто проводять три «підходи» вимірювання.

Нормальною щодо сприйняття звуків різних частот на порозі чутності для людини вважають таку частотну характеристику, коли її форма відповідає приблизно кривій однакової гучності на порозі чутності з нерівномірністю в діапазоні 125…8000 Гц 25-30 дБ, з максимальною чутливістю на частотах 2–4 кГц [9].

Для здорової людини відмінностей не повинно бути або припускають відмінності до 25 дБ, що вважається нормальним слухом. Відмінність у 26–40 дБ вважають як ступінь із легкою втратою слуху, 41–55 дБ – проміжний ступінь утрати слуху, 56–70 дБ – важкий ступінь [10].

Переваги зазначеного методу аудіометрії – це стандартизована методика вимірювань, наявність спеціалізованої медичної техніки (приладів) з установленими оптимальними параметрами для проведення дослідження, проведення досліджень через повітряну й кісткову провідності для уточнення результатів і більш точного виявлення вад слуху. Недоліком, зокрема, для звукорежисера, окрім необхідності проведення цього обстеження в спеціалізованому медичному центрі з використанням відповідного медичного обладнання й акустичної камери з роз'ясненнями спеціаліста-отоларинголога [11], є обмежений частотний діапазон, зазвичай визначений діапазоном від 125 Гц до 8 кГц, притаманний мові, хоча людина сприймає від 20 Гц до 20 кГц, вхідний тональний сигнал не завжди зрозумілий для слухача тобто він не завжди знає, що має чути і яка приблизно гучність повинна бути порівняно з оптимальною. Для порівняння на рис. 1 наведено вигляд кривих однакової гучності (рис. 1, а) та аудіограми за повітряною провідністю (рис. 1, б) – суцільна крива.

Метою роботи є вдосконалення методу визначення вад слуху, який дасть змогу з мінімальними витратами самостійно визначати спроможність і якість сприйняття звукових сигналів у звуковому частотному діапазоні, причому у звичних для слухача умовах і при оптимальній (зручній) для прослуховування гучності. У разі виявлення вад сприйняття окремих частотних складників дати можливість корегувати гучність і частотну характеристику (далі – ЧХ) звуковідтворення у відповідних частотних смугах таким чином, щоб звуковий образ найкраще сприймався особами – слухачами режисованого звукового контенту, без вад слуху.



а) б)

Рис. 1. Частотні характеристики: а) криві однакової гучності; б) аудіограма за повітряною провідністю (суцільна лінія) та за кістковою провідністю (штрихова лінія)

Виконання досліджень

Сутність пропонованого методу полягає в такому:

1) слухач, готуючись до дослідження, визначає тестові сигнали різного змісту, а саме шумові, музичні, тональні;

2) створюють комфортні умови дослідження щодо розташування відносно гучномовців і встановлюють доцільну гучність;

3) відтворюють тестові сигнали, гучність яких збільшують від нуля до номінального рівня за визначений проміжок часу, за допомогою вбудованого у спеціалізоване програмне забезпечення (далі – ПЗ) вимірювача рівня визначають рівень сигналу, при якому з'являється відчуття звукового контенту;

4) за результатами дослідження будують графік частотної характеристики сприйняття тональних звуків і порівнюють із теоретичною кривою однакової гучності на порозі чутності. Відхилення між кривими будуть сигналізувати щодо погіршення якості сприйняття звуків і наявності вад слуху.

Процедура визначення якості сприйняття звукового контенту пропонованим методом потребує наявності такого апаратного й програмного забезпечення:

- персональний комп'ютер (далі – ПК) зі спеціалізованим ПЗ для роботи зі звуком (необов'язково професійним), наприклад, Audacity, Sound Forge, Adobe Audition тощо, у якому є можливість синтезувати тональні звуки різних частот звукового діапазону й убудовано досить точний вимірювач рівня, наприклад, PPM типу;

- гучномовців, бажано з високими технічними параметрами, наприклад, звукові монітори або гучномовці Hi-Fi класу, а також навушники.

Як лабораторна робота для студентів у праці [12] описано метод дослідження слуху у звичайних умовах житлової кімнати або звукорежисерської апаратної.

Сам процес передбачає визначення рівня гучності сигналу, що відповідає порогу чутності в умовах відтворення, і ЧХ сприйняття звуку. Значення рівня сигналу, за яким визначено поріг чутності, у дослідженні встановлюють за значенням рівня електричного сигналу, що відображено на вбудованих у ПЗ індикаторах рівня, зокрема індикаторі PPM. Потім за значеннями електричних рівнів будують криву ЧХ сприйняття, форму якої порівнюють із формою кривих однакової гучності.

Структурну схема підключення обладнання й умови проведення обстеження наведено на рис. 2.

Для дослідження застосовано ПК з ПЗ для роботи зі звуком (Sound Forge [13]) і звичні для слухача пристрої звуковідтворення. Як джерело тестових сигналів застосовано генератор синусоїдних сигналів, убудований у ПЗ.

Перед початком обстеження на гучномовцях, якщо вони активні й мають регулятори тембру, положення регуляторів тембру встановлюють у «0», що відповідає відсутності корекції ЧХ. На підсилювачах потужності в разі застосування пасивних гучномовців регулятори тембру встановлюють у «0». У програмних регуляторах тембру ПК регулятори тембру теж виводять у «0».

Після цього на головний екран монітора ПК завантажують Sound Forge й досить якісний і зрозумілий для піддослідного музичний фрагмент, бажано з насиченим інструментальним супроводом (рис. 3). У режимі відтворення фрагменту встановлюють оптимальну гучність для слухача, що знаходиться на певній (зручній для комфортного прослуховування) відстані від гучномовців. Гучність установлюється регулятором гучності на гучномовцях або підсилювачі, а також віртуальним регулятором ПК.

Рис. 2. Структура обладнання й умови проведення дослідження

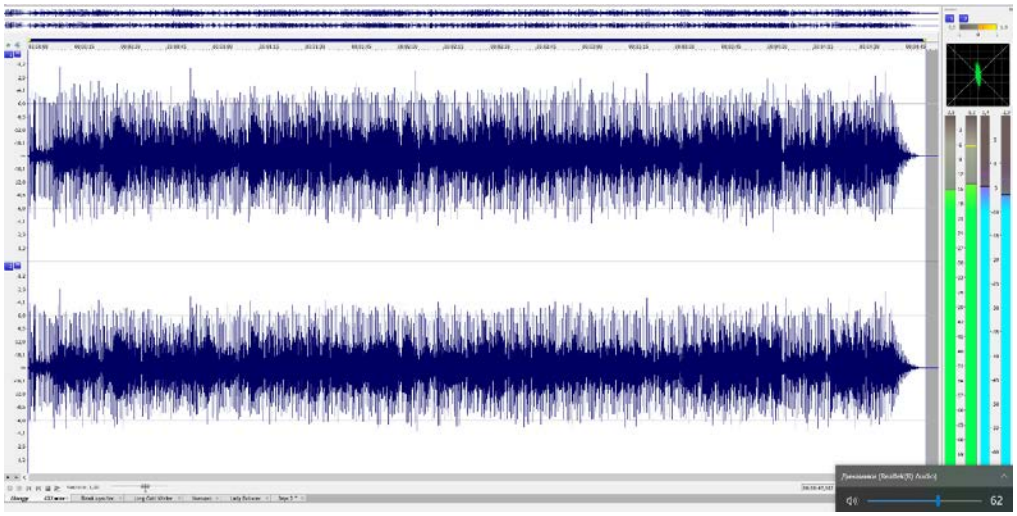


Рис. 3. Головне вікно ПЗ Sound Forge із завантаженим музичним фрагментом

Критерієм оптимальної гучності в контексті дослідження можна вважати, наприклад, гучність, яку встановлює пересічний слухач при комфортному прослуховуванні звукових програм у побутових умовах (з досвіду авторів, це приблизно 75–80 дБ), або гучність, що відповідає рівню, за якого звукорежисер контролює звукові програми в студії, тобто 90–95 дБ.

Зрозуміло, що в побутових умовах визначати числове значення рівня не завжди можливо, адже пересічний слухач не володіє вимірювальним обладнанням для вимірювання рівня акустичного сигналу. Навпаки, звукорежисер зазвичай має такі прилади або має завантажений у смартфон додаток, за допомогою якого можна вимірювати рівень звукового акустичного сигналу.

Одночасно з прослуховуванням фонограм в акустичному просторі приміщення можна за допомогою завантаженого в смартфон вимірювача рівня акустичного сигналу, наприклад, ПЗ Sound Meter [14], вимірювати оптимальний рівень прослуховування в місці прослуховування (обстеження), а також спектр відтворюваного звукового музичного сигналу за допомогою додатка в смартфоні, наприклад, ПЗ RTA Analyzer [15]. У разі вимкнення відтворення фонограм вимірюють рівень сторонніх шумів у приміщенні, у якому здійснюють дослідження.

Після встановлення оптимального рівня гучності для музичної програми завантажують у програмний засіб сигнал білого шуму як сигнал із широким частотним спектром, з рівнем -20 дБ, тривалістю 600 с (10 хв. – максимально програмований час тестових сигналів у Sound Forge) і, увімкнувши відтворення, прослуховують звучання білого шуму в гучномовцях, усвідомлюючи (запам'ятовуючи) його звучання.

Ця тривалість фрагменту необхідна для створення більш плавного збільшення гучності сигналу й, відповідно, більш точного встановлення значення рівня сигналу на порозі чутності. Можна брати

тривалість фрагмента 1 хв., але максимальний рівень установлювати менше, наприклад, -40 дБ. Якщо меж рівня буде недостатньо для визначення рівня порогу чутності або, навпаки, важко буде визначити рівень сприйняття, нормалізацією можна змінювати (збільшувати або зменшувати) максимальний рівень.

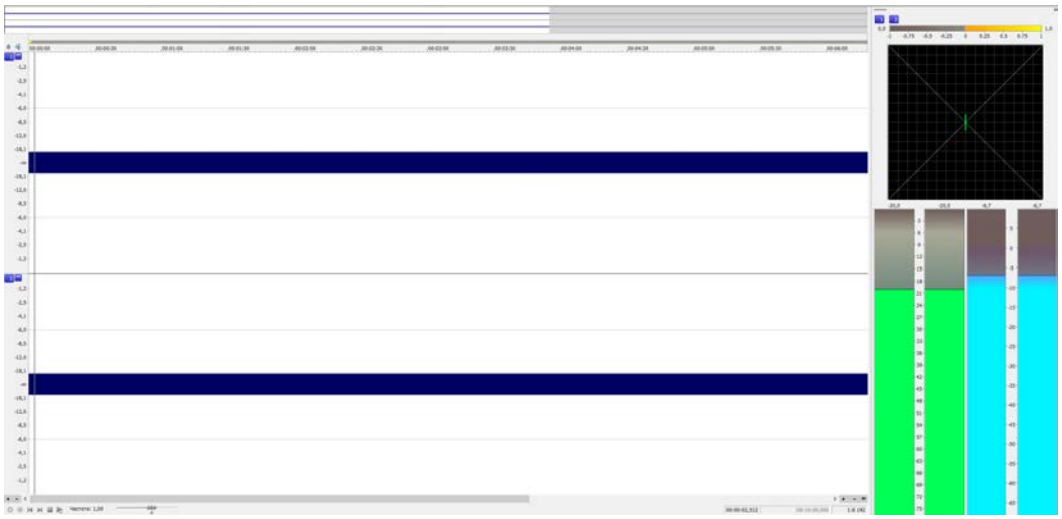
Послідовність операцій:

1. Файл – Створити – Нове вікно (частота дискретизації – 44100 Гц, квантування – 16 біт, 2 канали стерео).

2. Insert (вставити) – Synthesis (синтезувати) – FM – білий шум (тривалість – 600 с, рівень – -20 дБ) (рис. 4, а).

За допомогою обробки сигналу, що передбачає плавне наростання гучності, установлюють наростання сигналу вздовж усього фрагменту.

Послідовність операцій: 1) Обробка – Згасання – Зростання (рис. 4, б).



а)

б)

Рис. 4. Головне вікно ПЗ Sound Forge: а) із завантаженим білим шумом без зміни рівня; б) зі зміною рівня й зупиненим курсором у місці відчуття сигналу на порозі чутності

Після цього необхідно ввімкнути відтворення з початку фонограми, вслуховуючись у звучання в гучномовцях. Як тільки сигнал буде достатньо чітко почутий, натиснути «пауза», у місці встановлення курсору ПЗ визначають рівень сигналу, який і буде відповідати рівень порога чутності тестового сигналу, у цьому випадку білого шуму (рис. 4, б) для слухача.

Для спрощення визначення рівня сигналу при прослуховуванні й формування звіту дослідження можна на клавіатурі комп'ютера натискати клавішу PrtSc, роблячи тим самим скриншот (знімок) екрану. Потім цей знімок уставити в текстовий документ як підтвердження дослідження.

Після визначення рівня сприйняття білого шуму в ПЗ завантажують тестовий сигнал частотою 63 Гц, тривалістю 600 с, рівнем -20 дБ (або з меншою тривалістю й, відповідно, меншим рівнем, залежно від очікуваного результату відчутого рівня).

Послідовність операцій: 1) Файл – Створити – Нове вікно (частота дискретизації – 44100 Гц, квантування – 16 біт, 2 канали стерео).

2) Insert (вставити) – Synthesis (синтезувати) – Звичайна синусоїда (тривалість – 600 с, частота – 63 Гц, рівень – -20 дБ) (рис. 5, а).

Знову, використовуючи обробку зростання рівня сигналу (послідовність операцій: 1) Обробка – Згасання – Зростання (рис. 5, б), формують сигнал для випробування. Після цього потрібно увімкнути відтворення сигналу й очікувати чітке відчуття його.

Таке дослідження проводять на всіх центральних частотах октавних смуг звукового діапазону (63, 125, 250, 500, 1000, 2000, 4000, 8000, 16000 Гц). Як варіант, можливо обирати проміжні частоти, наприклад, 80, 100, 400, 3000, 6000, 10000, 12000, 14000, 18000 Гц тощо.

Для кожної частоти вимірюють рівень сприйняття, складають таблицю рівня чутних сигналів і будують графік, який порівнюють із тестовою кривою однакової гучності. При виявленні вад у сприйнятті окремих частотних смуг звукорежисер має враховувати ці вади при встановленні параметрів формування звуку, особливо щодо тембрального забарвлення звучання. Дослідження необхідно проводити спочатку для одного вуха, закришив інше вушною вкладкою (берушею), потім для іншого й для двох відкритих вух.

а)

б)

Рис. 5. Головне вікно ПЗ Sound Forge: а) із завантаженим синусоїдним сигналом частотою 63 Гц без зміни рівня; б) зі зміною рівня й зупиненим курсором у місці відчуття сигналу на порозі чутності

Для кожної частоти вимірюють рівень сприйняття, складають таблицю рівня чутних сигналів і будують графік, який порівнюють із тестовою кривою однакової гучності. При виявленні вад у сприйнятті окремих частотних смуг звукорежисер має враховувати ці вади при встановленні параметрів

формування звуку, особливо щодо тембрального забарвлення звучання. Дослідження необхідно проводити спочатку для одного вуха, закривши інше вушною вкладкою (берушею), потім для іншого й для двох відкритих вух.

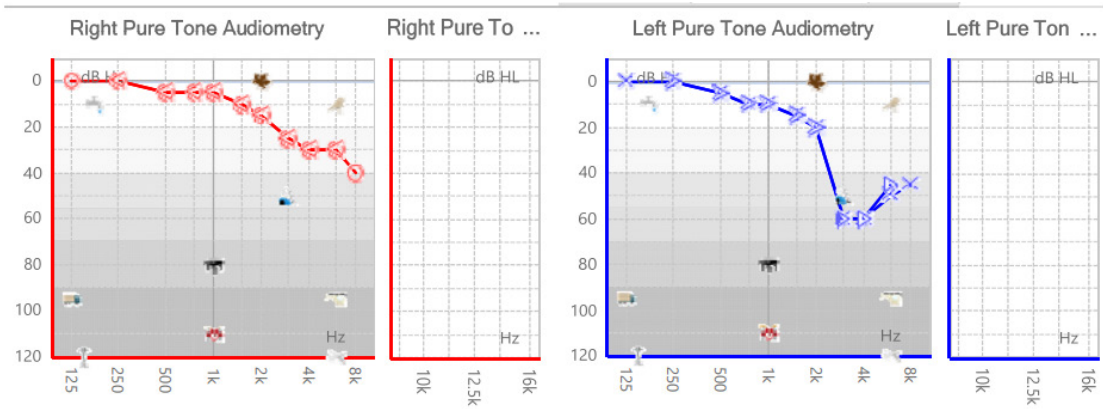
При застосуванні для вимірювання навушників початкові рівні сигналів у програмному засобі необхідно встановлювати в межах -30…-40 дБ. Для спостереження за формою сигналограми необхідно виконувати масштабування не тільки в горизонтальній площині (праворуч + та -), а й у вертикальній (ліворуч + та -).

Порядок проведення передбачає відтворення тестових тональних сигналів з різною гучністю в реальних умовах прослуховування фонограм і визначення рівня гучності сигналу, при якому піддослідний цей сигнал сприймає, тобто починає чути. Значні відхилення від графіка будуть сигналізувати про вади слуху. Хоча окремі відхилення можуть бути визначені й недоліками відтворення в приміщенні, недосконалим низькоякісним обладнанням, наявністю сторонніх звуків тощо. Тому доцільно проводити оцінювання якості відтворюваного аудіоконтенту, зокрема як акустичного сигналу [16].

Натурний експеримент. Верифікацію зазначеного методу проведено порівнянням з аудіограмою на прикладі аналізу вад слуху 1 особи. Як приклад дослідження слухового сприйняття людиною звукових сигналів різних частот наведено дослідження, виконане одним із авторів у медичному центрі Беттертон у м. Києві [17]. Результати аудіограми подано на рис. 6.

Відповідно до висновку, слухач має порушення функції апарату звукосприйняття за сенсоневральним типом 1 ступеня. Нормою хорошого слуху є нерівномірність ЧХ сприйняття звукових сигналів тональних частот менше ніж 20 дБ. 1 ступінь порушення слуху говорить про те, що ЧХ звукосприйняття може бути від 20 дБ до 50 дБ. Праве вухо в слухача – на нижніх частотах, до 2 кГц у нормі, а на частотах вище за 2 кГц є порушення. Ці порушення можливо компенсувати підвищенням гучності сигналу, щоб він був почутий, наприклад, на частоті 8 кГц на 40 дБ, порівняно із сигналом частотою 250 Гц. Для лівого вуха існує ще більший «провал» у сприйнятті звукових тональних сигналів на частотах 3–4 кГц до 60 дБ. Тобто, щоб слухач почув ці сигнали, необхідно збільшити їх гучність на 60 дБ, порівняно з частотою 250 Гц.

а)



б)

Рис. 6. Графіки аудіограми для правого та лівого вух: а) головне вікно вимірювального приладу, б) скриншот паперового висновку

Дослідження слухового сприйняття звукових сигналів методом, запропонованим у статті, проведено в житловій кімнаті об'ємом приблизно 50 м^3 з установленими елементами облаштування – м'якими та корпусними меблями, шторами на вікнах, телевізором, ПК тощо. Прослуховування тестових сигналів здійснено через стереофонічну систему мультимедійних гучномовців, установлених на робочому столі під монітором персонального комп'ютера. Відстань між гучномовцями – приблизно 50 см, відстань до слухача – приблизно 80 см (рис. 7, а).

Практичне дослідження проводили для двовушного прослуховування, тобто, на відміну від аудіограми, звуковий контент сприймався одночасно обома вухами. Як гучномовці застосовано мультимедійні гучномовці Edifier R1600T III Multimedia Speaker [18] (рис. 7, б).

Основні параметри гучномовців:

– номінальна електрична потужність кожного гучномовця – 30 Вт;
– відношення сигнал/шум – не менше ніж 85 дБА;
– нелінійні спотворення – не більше ніж 0,5%;
– частотний діапазон – 65…20000 Гц.

Гучномовці активні з можливістю регулювання гучності й тембру на НЧ та ВЧ.

а) б)

Рис. 7. Особливості натурного експерименту: а) місце прослуховування для дослідження; б)загальний вигляд гучномовців

У кімнаті під час дослідження рівень сторонніх акустичних шумів становив приблизно 30 дБА. Вимірювання акустичних сигналів проведено вимірювачем Phonic PAA3, а також додатком до смартфону Sound Meter. Перед дослідженням регулятори тембру на гучномовцях установлено в середнє положення, що забезпечує відсутність налаштування ЧХ.

Налаштування на комфортну гучність здійснено за музичним твором групи Deep Purple «Smoke on the Water». Рівень акустичного сигналу в точці прослуховування становив приблизно 75–80 дБА. Головне вікно програми Sound Forge з відображенням сигналограми музичного твору й рівня електричного сигналу наведено на рис. 8.

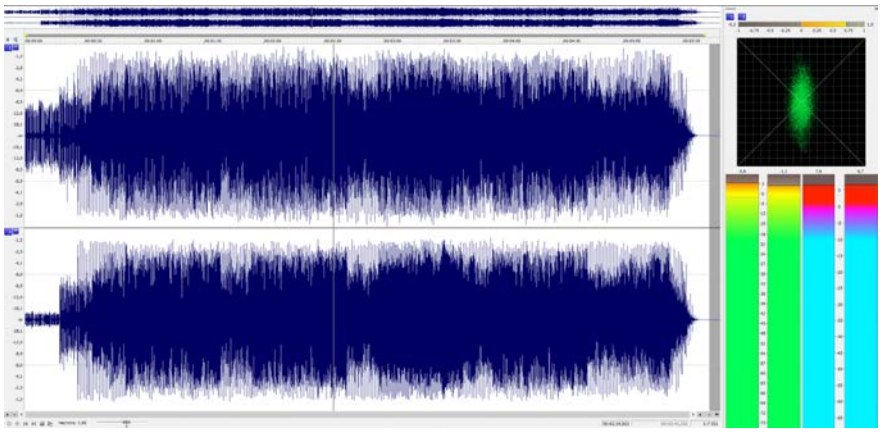


Рис. 8. Головне вікно програми Sound Forge із завантаженим музичним твором у режимі відтворення

Рівень електричного сигналу при встановлених умовах прослуховування – відповідному приміщенні й комфортній гучності – становить приблизно -58 дБ.

Після дослідження сприйняття білого шуму (рис. 9) досліджено сприйняття тональних сигналів різної частоти в діапазоні від 31,5 Гц до 16 кГц.

Рис. 9. Головне вікно програми Sound Forge із завантаженим тестовим сигналом білим шумом і визначеною точкою сприйняття звуку

Попередньо сигнали синтезовано вбудованим віртуальним генератором ПЗ, для них забезпечено необхідний часовий інтервал, збільшення гучності впродовж усієї тривалості сигналу та максимальну гучність, передбачену у фрагменті. Як приклад результатів дослідження наведено скриншоти результатів дослідження на частотах 63 Гц (рис. 10, а), 1000 Гц (рис. 10, б), 14000 Гц (рис. 10, в).

За результатами складено таблицю (див. таблицю 1) і побудовано графік частотної характеристики сприйняття сигналів на порозі чутності, який і порівнюють із кривою однакової гучності на порозі чутності (рис. 11).

Таблиця 1

Результати вимірювань ЧХ звукосприйняття на порозі чутності в умовах домашнього прослуховування

f, Гц	31,5	63	125	250	500	1к	2к	4к	8к	10к	12к	14к	16к
N, дБ	-11	-44	-63	-70	-75	-72	-64	-49	-31	-25	-27	-8	0

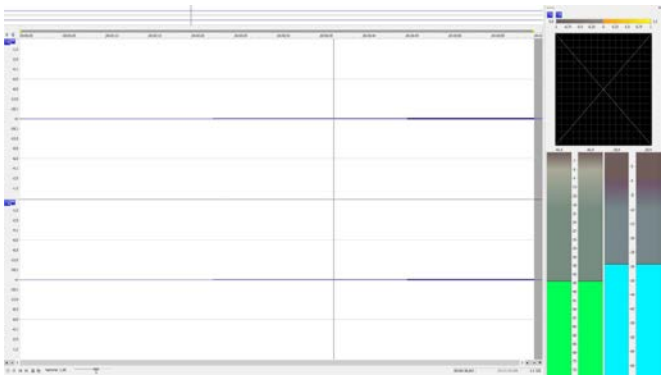
Рис. 11. Форма частотної характеристики сприйняття акустичних сигналів: крива світло-зеленого кольору – особою на порозі чутності, крива червоного кольору – особою за результатами дослідження, крива коричневого кольору – особою з явними вадами слуху, крива синього кольору – показів вимірювального приладу акустичних сигналів у точці прослуховування, крива темно-зеленого кольору – теоретичної кривої порогу чутності

Рівень електричного звукового сигналу в роботі звукорежисера визначають максимальним значенням у 0 дБ. Відповідність рівня акустичного сигналу електричному рівню визначають з урахуванням чутливості мікрофона, який акустичний сигнал перетворює в електричний, тобто електричний рівень щодо акустичного сигналу залежить від чутливості мікрофона. Цю залежність може бути визначено формулою:

$$N_{\text{м}} = 20\lg\frac{E_{\text{ос}}\cdot p_{\text{зв}}}{U_0} = 20\lg\frac{E_{\text{ос}}\cdot p_0 10^{\frac{L_{\text{р}}}{20}}}{0{,}775} = 20\lg\frac{E_{\text{ос}}\cdot 2\cdot 10^{-5}10^{\frac{L_{\text{р}}}{20}}}{0{,}775} = \tag{1}$$

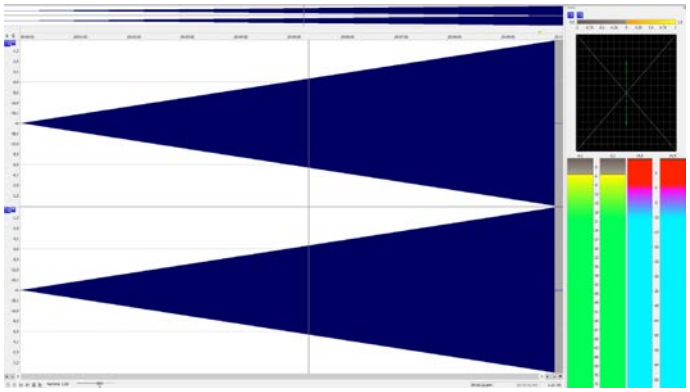
$$= 20\lg(2{,}58\cdot 10^{-5}E_{\text{ос}}\cdot 10^{\frac{L_{\text{р}}}{20}}) \approx -92 + 20\lg(E_{\text{ос}}\cdot 10^{\frac{L_{\text{р}}}{20}}),$$

де $N_{\text{м}}$ – рівень електричного сигналу на виході мікрофона; $E_{\text{ос}}$ – осьова чутливість мікрофона або ЕРС на виході мікрофона при підведенні до нього на робочій осі акустичного сигналу відповідного звукового тиску; $p_{\text{зв}}$ – звуковий тиск; p_0 – звуковий тиск, що відповідає порогу чутності й нормоване значення звукового тиску, щодо якого визначають рівень акустичного сигналу; $L_{\text{р}}$ – рівень акустичного сигналу за звуковим тиском.



а)

б)



в)

Рис. 10. Головне вікно програми Sound Forge із завантаженими тестовими сигналами та визначеними точкою сприйняття звуку на частотах: а) 63 Гц; б) 1000 Гц; в) 14 кГц

Рівень акустичного сигналу $L_р$, для якого рівень електричного сигналу дорівнює 0 дБ, визначають за формулою:

$$0 \approx -92 + 20\lg E_{ос} \cdot 10^{\frac{L_р}{20}},$$

$$L_р = 20\lg \frac{10^{\frac{92}{20}}}{E_{ос}} = 20\lg \frac{10^{4,6}}{E_{ос}}. \quad (2)$$

За цією формулою, при вихідній напрузі (ЕРС) мікрофона E=0,775 В, що відповідає 0 дБ, рівень акустичного сигналу становить $L_р$=94,2 дБ ($p_{зв}$ = 1 Па). З таким приблизно рівнем (від 90 до 95 дБ) зазвичай звукорежисер у студії прослуховує (контролює) звучання міксу звукового сигналу.

Отже, відповідність рівня електричного сигналу рівню акустичного сигналу має становити таке:

$$L_{\text{р}}=94 \text{ дБ} \rightarrow N_{\text{м}}=0 \text{ дБ}.$$

З огляду на те що рівень акустичного сигналу симфонічного оркестру у форте-фортисимо (максимальна гучність) становить приблизно 100 дБ, а також зазвичай на концертах рівень у зоні озвучення й, наприклад, для глядацької зали кінотеатру система озвучення повинна забезпечувати рівень 100 дБ, а максимальний неспотворений рівень електричного сигналу – 0 дБ, у роботі на рисунках графіків ЧХ й прийнята відповідність рівня акустичного сигналу 100 дБ саме рівню електричного сигналу 0 дБ, тобто:

$$L_{\text{р}}=100 \text{ дБ} \rightarrow N_{\text{м}}=0 \text{ дБ}.$$

За результатами оцінювання слухового сприйняття одним запропонованим методом можна зазначити, що форма ЧХ сприйняття звукових сигналів (рис. 11, а, червоного кольору) збігається за формою з теоретичною кривою однакової гучності на порозі чутності, однак на частотах максимальної чутливості слуху (2–4 кГц) чутливість піддослідного значно погіршена, також сприйняття високих частот, починаючи вже з частоти 1 кГц, теж має вади. Наприклад, різниця між сприйняттям сигналів частотою 8 кГц і 1 кГц для «ідеального» слуху за кривою однакової гучності повинна становити приблизно 5 дБ, а в особи, якій досліджували слух за цим методом, ця різниця становить 41 дБ.

Така різниця потребує значної корекції частотної характеристики відтворення сигналів. Якщо звукорежисер при таких вадах буде формувати відповідне тембральне забарвлення звучання фонограми, для здорового слухача в сприйнятті будуть завищені рівні високих частот. Отримані результати приблизно відповідають результатам аудіограми (рис. 6).

Способом, запропонованим у статті, досліджено сприйняття звукових сигналів на порозі чутності серед молоді – студентів майбутніх звукорежисерів (35 студентів, віком від 18 до 25 років). Дослідження особи проводили індивідуально відповідно до протоколу дослідження без залучення оператора або інших сторонніх осіб, тому можливі додаткові похибки результатів дослідження зумовлені суб'єктивним фактором.

Особливість перевірки слухових спроможностей для вибірки з групи молоді полягала в тому, що кожна особа мала змогу перевірити свої слухові властивості у власній житловій кімнаті, тобто в тому місці, де найчастіше здійснює прослуховування звукового контенту й із застосуванням власних засобів звуковідтворення. Побутові умови дослідження слухових властивостей дають можливість більше зосередитися на сигналі, що відтворюють, створює більш спокійну обстановку, хоча при цьому можливі побутові шуми. Але сторонні побутові шуми, навпаки, підвищують цінність дослідження, адже наближають слухові властивості до реальних умов прослуховування.

Вибірка становила групу з 25 чоловіків і 10 жінок. Серед поданих результатів дослідження результати, що відповідають реальним, становлять 28, а 7 досліджень за результатами не відповідають дійсним, серед яких і такі, що здійснювали дослідження з використанням низькоякісних гучномовців, убудованих у ноутбуки.

Реальні або дійсні результати характеризуються передусім відповідністю форми графіка ЧХ сприйняття тональних сигналів кривій однакової гучності на порозі чутності (рис. 11, б), по-друге, реальними умовами дослідження, зокрема вибором засобу прослуховування сигналів – достатньо якісними гучномовцями або навушниками, по-третє, вибором оптимальної гучності музичного сигналу перед початком дослідження при застосування гучномовців, що підтверджено вимірювачем рівня сигналу, убудованого як додаток у смартфоні.

Отримані ЧХ слухового сприйняття мають деякі відхилення на окремих частотах, але загалом повторюють криву однакової гучності. Приклади реальних ЧХ, отриманих у результаті дослідження, подано на рис. 11, б, приклади тих досліджень, що мають на окремих частотах значні відхилення від реальних ЧХ, – на рис. 11, в, результати дослідження, що не відповідають дійсності і є помилковими, що не враховувалися як позитивне дослідження, – на рис. 11, г. Як засіб прослуховування учасники дослідження обирали гучномовці: звукові монітори – 10 осіб; мультимедійні акустичні системи високої якості – 8 осіб; високоякісні повногабаритні навушники – 15 осіб; гучномовці, що вбудовані в ноутбуки, – 2 особи.

За результатами дослідження, серед реальних 28 результатів відхилення від теоретичної кривої порогу чутності в межах 10…20 дБ у звуковому діапазоні, що вважають відсутністю вад слуху за медичними показниками, мали 20 осіб, більше ніж 20 дБ на окремих частотах звукового діапазону – 7 осіб, а більше ніж 20 дБ у широких смугах частот – одна особа. Графіки (форма кривої) частотних характеристик сприйняття тональних звуків практично збігаються з кривою однакової гучності (наприклад, для трьох осіб на рис. 11, б) і знаходяться в межах заштрихованої зони, у якій у більшості для піддослідних були отримані досліджувані криві (рис. 11, б). Відмінність визначено насамперед розташуванням кривої по вертикалі, що залежить від установлення початкової оптимальної гучності, за якої проводили дослідження.

Висновок

Запропоноване вдосконалення методу визначення вад слуху не можна вважати ідеальним і таким, що повністю відповідає вимогам за медичними показниками, але воно достатньо ефективне, адже дає змогу визначати вади слуху в комфортних зручних умовах.

Запропонований метод має на меті не точне визначення вад слуху, а визначення необхідної корекції частотної характеристики звуковідтворення для якісного сприйняття звукового контенту. Недоліки методу полягають у такому: по-перше, для дослідження бажано мати пристрої звуковідтворення з лінійною частотною характеристикою, що не завжди можна досягти, адже вимірювання здійснюють у власних кімнатах або студіях, де можуть бути недоліки акустичних умов приміщення, що призводить до нелінійності ЧХ звуковідтворення; по-друге, саме обладнання не завжди може забезпечувати лінійну частотну характеристику, ті ж гучномовці, адже навіть установлення регулятора тембру в середнє положення, що відповідає лінійності ЧХ, не гарантує лінійну АЧХ звуковідтворення; по-третє, розшифрування отриманої аудіограми потребує деяких знань психоакустики. Крім того, психологічний стан слухача для підвищення своєї самооцінки може призвести до хибних показань, щоб не відчувати себе як людину з вадами слуху. Краще, щоб результати оцінював незалежний аудіолог.

Недоліком у виборі навушників для дослідження є неможливість визначення рівня сигналу, що відтворюють навушники, а перевагою – більш точне визначення сигналу на порозі чутності, адже не заважають зовнішні шумові фактори.

Перевага застосування гучномовців – реальний акустичний сигнал у навколишньому просторі приміщення, що вважають як більш натуральний щодо сприйняття звукового контенту, можливість визначення оптимальної гучності при прослуховуванні фонограм у місці прослуховування.

Переваги запропонованого вдосконалення методу – це звична для звукорежисера навколишня обстановка з усіма акустичними перевагами й недоліками, попереднє встановлення оптимальної гучності причому для сигналу й контенту, до якого звик слухач, можливість проведення досліджень у всьому частотному діапазоні прийняття звуку людиною, у разі виявлення недоліків (вад) слуху більш детально зазначати частоти, на яких проводиться вимірювання, причому завжди можна прослухати сигнал з достатньою гучністю й очікувати саме цей сигнал на порозі чутності. Дослідження проводять як в акустичному просторі приміщення, так і в індивідуальних навушниках.

Запропонований метод не дає результатів високої точності, але дає змогу проводити вимірювання в зручних і звичних для досліджуваного умовах, створює почуття розуміння сприйняття різних звуків, дає можливість порівнювати звучання тональних звуків із широкосмуговими (мовними, музичними), застосовує для вимірювання вимірювальні прилади програмних засобів ПК, є безкоштовним методом, дає змогу проводити дослідження в будь-який зручний для досліджуваного час тощо.

Подальший напрям роботи пов’язаний із визначенням припустимих похибок залежно від особливостей умов застосування запропонованого методу, визначення якості аудіоконтенту, який прослуховують, і застосування алгоритмів штучного інтелекту для узгодження результатів за запропонованим методом з особливостями аудіометрії.

Література

1. Основи звукорежисури : навчальний посібник / Н. Д. Бєлявіна, В. Ф. Бєлявін, Н. Л. Бондарець, В. В. Дьяченко ; під ред. Н. Д. Бєлявіної. Київ : НАКККіМ, 2011. Ч. І. 84 с. URL: https://www.academia.edu/23244513/.

2. Десятник Г. О., Бадіон С. В. Професія: звукорежисер : тексти лекцій. Київ : Інститут журналістики КНУ, 2019. 69 с. URL: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ktm.journ.knu.ua/wp-content/uploads/2020/02/Profesiia-zvukorezhyser-verstka-konvertyrovan.pdf.

3. Українська музична енциклопедія. Київ : Інститут мистецтвознавства фольклористики та етнології ім. М. Т. Рильського, 2008. Т. 2. 663 с. URL: https://shron3.chtyvo.org.ua/Skrypnyk_Hanna/Ukrainska_muzychna_entsyklopediia_Tom_2.pdf?PHPSESSID=q1ri75iflnbcd1qrqnagblnmu0.

4. Pras A., Guastavino C. The role of music producers and sound engineers in the current recording context, as perceived by young professionals. *Musicae Scientiae*. 2011. № 15 (1). P. 73–95. https://doi.org/10.1177/1029864910393407.

5. Neuenfeldt K. Learning to Listen When There Is Too Much to Hear: Music Producing and Audio Engineering as “Engaged Hearing”. *Media International Australia, Incorporating Culture & Policy*. 2007. № 123. P. 150–160. URL: https://search.informit.org/doi/10.3316/informit.898978684851745.

6. Porcello T. Speaking of Sound: Language and the Professionalization of Sound-Recording Engineers. *Social Studies of Science*. 2004. № 34 (5). P. 733–758. https://doi.org/10.1177/0306312704047328.

7. Guide for the Evaluation of Hearing Handicap. *JAMA*. 1979. № 241(19). P. 2055–2059. DOI: 10.1001/jama.1979.03290450053025.

8. Основи звукотехніки : навчальний посібник / КПІ ім. Ігоря Сікорського ; укладачі : О. П. Гребінь, В. Б. Швайченко, Н. Ф. Левенець. URL: https://ela.kpi.ua/handle/123456789/55729.

9. Profile analysis in listeners with normal and elevated audiometric thresholds: Behavioral and modeling results / Daniel R. Guest, David A. Cameron, Douglas M. Schwarz, U-Cheng Leong, Virginia M. Richards, Laurel H. Carney. *J. Acoust. Soc. Am*. 2024. № 156. P. 4303–4325. https://doi.org/10.1121/10.0034635.

10. Аудіометрія: своєчасна і точна діагностика слуху. URL: https://lordoctor.dp.ua/uk/poslugi/diagnostika-slukha/audiometriya.html.

11. Van Eeckhoutte M., Jasper B. S., Kjærbøl E. F., Jordell D. H., Dau T. In-situ Audiometry Compared to Conventional Audiometry for Hearing Aid Fitting. *Trends in Hearing*. 2024. № 28. DOI: 10.1177/23312165241259704.

12. Архітектурна акустика. Лабораторний практикум : навчальний посібник / КПІ ім. Ігоря Сікорського ; укладачі: О. П. Гребінь, С. А. Луньова, Н. Ф. Левенець. URL: https://ela.kpi.ua/items/7813753c-5ce5-4e22-ac4c-ace3eace3e39

13. SOUND FORGE Audio Studio. URL: https://www.sony.com/electronics/support/downloads/00015797.

14. Sound meter Pro. URL: https://apps.microsoft.com/detail/9ntm0x0nwlfw?hl=uk-UA&gl=UA.
15. RTA – Real Time Analyzer. URL: https://studiosixdigital.com/audiotools-modules-2/acoustic-analysis-modules/rta/.
16. Prodeus A., Didkovska M. Fourth Moment and Its Functional Transformations as Measures of Clipping Degree and Quality of Acoustic Signal. *Radioelectronics and Communications Systems.* 2021. Vol. 64. № 5. P. 255–265: https://doi.org/10.3103/S0735272721050046.
17. Аудіометрія слуху. URL: https://bettertone.com.ua/uk/service/audiometriya/.
18. Мультимедійні гучномовці Edifier R1600T III Multimedia Speaker. URL: https://edifier.com.ua/kolonki/stereo/r1600tiii.html.

References

1. Osnovy zvukorezhysury : navch. posib. CH. I [Fundamentals of sound engineering: a manual. Part I] / Byelyavina N. D., Byelyavin V. F., Bondarets' N. L., D'yachenko V. V.; edit. by. N. D. Byelyavina. Kyiv: NAKKKiM, 2011. 84 p. URL: https://www.academia.edu/23244513/ [in Ukrainian].
2. Desyatnyk H. O., Badion S. V. Profesiya: zvukorezhyser: teksty lektsiy. [Profession: sound engineer: lecture texts]. Kyiv: Instytut zhurnalistyky = Institute of Journalism KNU, 2019. 69 p. URL: https://ktm.journ.knu.ua/wp-content/uploads/2020/02/Profesiia-zvukorezhyser-verstka-konvertyrovan.pdf [in Ukrainian].
3. Ukrayins'ka muzychna entsyklopediya. T. 2 [Ukrainian Musical Encyclopedia. Vol. 2.] – Kyiv: Vydavnytstvo Instytutu mystetstvoznavstva fol'klorystyky ta etnolohiyi im. M. T. Ryl's'koho = Publishing House of the M. T. Rylsky Institute of Art History, Folklore and Ethnology, 2008. 663 p. URL: https://shron3.chtyvo.org.ua/Skrypnyk_Hanna/Ukrainska_muzychna_entsyklopediia_Tom_2.pdf?PHPSESSID=q1ri75iflnbcd1qrqnagblnmu0 [in Ukrainian].
4. Pras, A., & Guastavino, C. (2011). The role of music producers and sound engineers in the current recording context, as perceived by young professionals. *Musicae Scientiae*, *15*(1), 73–95. https://doi.org/10.1177/1029864910393407.
5. Neuenfeldt, K. (2007). Learning to Listen When There Is Too Much to Hear: Music Producing and Audio Engineering as "Engaged Hearing". *Media International Australia, Incorporating Culture & Policy*, (123), 150–160. URL: https://search.informit.org/doi/10.3316/informit.898978684851745.
6. Porcello, T. (2004). Speaking of Sound: Language and the Professionalization of Sound-Recording Engineers. *Social Studies of Science*, *34*(5), 733–758. https://doi.org/10.1177/0306312704047328.
7. Guide for the Evaluation of Hearing Handicap. *JAMA*. 1979;241(19):2055–2059. DOI: 10.1001/jama.1979.03290450053025.
8. Osnovy zvukotekhniky [Basics of sound engineering] : teaching aids / Igor Sikorsky Kyiv Polytechnic Institute; compilers: O. P. Grebin, V. B. Shvaichenko, N. F. Levenets. URL: https://ela.kpi.ua/handle/123456789/55729 [in Ukrainian].
9. Profile analysis in listeners with normal and elevated audiometric thresholds: Behavioral and modeling results / Daniel R. Guest, David A. Cameron; Douglas M. Schwarz; U-Cheng Leong; Virginia M. Richards; Laurel H. Carney. *J. Acoust. Soc. Am*. 156, 4303–4325 (2024). https://doi.org/10.1121/10.0034635.
10. Аудіометрія: своєчасна і точна діагностика слуху. URL: https://lordoctor.dp.ua/uk/poslugi/diagnostika-slukha/audiometriya.html.
11. Van Eeckhoutte M., Jasper B. S., Kjærbøl E. F., Jordell D. H., Dau T. In-situ Audiometry Compared to Conventional Audiometry for Hearing Aid Fitting. *Trends in Hearing*. 2024;28. https://doi.org/10.1177/23312165241259704.
12. Arkhitekturna akustyka. Laboratornyy praktykum [Architectural Acoustics. Laboratory Workshop] : teaching aids / Igor Sikorsky Kyiv Polytechnic Institute; compilers: O. P. Grebin, S. A. Lunyova, N. F. Levenets. URL: https://ela.kpi.ua/items/7813753c-5ce5-4e22-ac4c-ace3eace3e39 [in Ukrainian].
13. SOUND FORGE Audio Studio. URL: https://www.sony.com/electronics/support/downloads/00015797.
14. Sound meter Pro. URL: https://apps.microsoft.com/detail/9ntm0x0nwlfw?hl=uk-UA&gl=UA
15. RTA – Real Time Analyzer. URL: https://studiosixdigital.com/audiotools-modules-2/acoustic-analysis-modules/rta/.
16. Prodeus A., Didkovska M. Fourth Moment and Its Functional Transformations as Measures of Clipping Degree and Quality of Acoustic Signal. Radioelectronics and Communications Systems, 2021, Vol. 64, № 5, pp. 255–265. https://doi.org/10.3103/S0735272721050046.
17. Audiometriya slukhu [Hearing audiometry]. URL: https://bettertone.com.ua/uk/service/audiometriya/ [in Ukrainian].
18. Mul'tymediyni huchnomovtsi Edifier R1600T III [Multimedia speakers Edifier R1600T III]. URL: https://edifier.com.ua/kolonki/stereo/r1600tiii.html [in Ukrainian].

Дата першого надходження рукопису до видання: 26.05.2025
Дата прийнятого до друку рукопису після рецензування: 30.06.2025
Дата публікації: 25.07.2025

УДК 004.624
DOI https://doi.org/10.36994/2788-5518-2025-01-09-03

СУЧАСНИЙ СТАН ТЕРАГЕРЦОВИХ ТЕХНОЛОГІЙ ДЛЯ РОЗПОДІЛЕНИХ СУПУТНИКОВИХ СИСТЕМ

Охременко Є. О., *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0009-0000-4907-6477, ohremenko2003@gmail.com*

Наритник Т. М., *к.т.н., академік Української академії наук, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0009-0008-3152-4356, t.natytnyk@ukr.net*

Капштик С. В., *к.т.н., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0002-3509-8015, sergii.kapshtyk@gmail.com.*

Авдєєнко Г. Л., *к.т.н., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0002-4788-7273, django2006@ukr.net.*

Анотація. *Стаття пропонує ґрунтовний огляд апаратних та алгоритмічних рішень для побудови високошвидкісних безпроводових систем у терагерцовому (ТГц) діапазоні, призначених для мереж шостого покоління (6G) і розподілених низькоорбітальних супутникових угруповань. Актуальність тематики зумовлена стрімким зростанням кількості IoT-пристроїв і потребою в глобальному обміні даними із затримкою < 1 мс і швидкістю > 100 Гбіт/с. Аналіз літератури й експериментальних прототипів показує, що поєднання масивних MIMO, гібридного формування променів і каналів 100–300 ГГц забезпечує спектральну ефективність до 20 біт/с/Гц при помірній кількості RF-ланцюгів. Лабораторні радіорелейні лінки 130–134 ГГц уже демонструють швидкість 1 200 Мбіт/с на відстані 1 км при BER ≤ 10^{-6}, що підтверджує життєздатність ТГц-інтерфейсів для міських і космічних сценаріїв. Особливу увагу приділено концепції «розподіленого супутника», де інтегруються туманні обчислення й міжсупутникові лінії зв'язку 5G-NTN і NB-IoT, що дає змогу локально обробляти телеметрію та зменшувати навантаження на наземний сегмент.*

Оглянуті матеріали свідчать, що перехід до монолітних GaN-RFIC з убудованими фазообертачами підвищує енергетичний бюджет на 30%, а композитні графенові тепловідводи знижують пікову температуру мікросхем на 15 °C. Водночас залишаються відкритими питання калібрування антенних масивів у вакуумі, пом'якшення фазового шуму вище за 200 ГГц і гармонізації розподілу спектра між активними й пасивними службами. Запропонована дорожня карта включає орбітальні експерименти з групою CubeSat, розроблення самонавчальних алгоритмів керування променями та створення адаптивних FEC-кодів, стійких до доплерівських зсувів. Таким чином, робота формує цілісне бачення переходу від лабораторних стендів до повномасштабних ТГц-мереж, здатних стати ядром 6G-екосистеми й глобального інтернету речей.

Ключові слова: *терагерцовий діапазон; 6G; розподілений супутник; гібридне формування променів; MIMO; міжсупутникові лінії зв'язку.*

PROSPECTS OF TERAHERTZ TECHNOLOGIES FOR DISTRIBUTED SATELLITE SYSTEMS

Yehor Okhremenko, *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0009-0000-4907-6477, ohremenko2003@gmail.com.*

Teodor Narytnyk, *PhD (Tech.), Academician of the Ukrainian Academy of Sciences, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0009-0008-3152-4356, t.natytnyk@ukr.net.*

Serhii Kapshtryk, *PhD (Tech.), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0000-0002-3509-8015, sergii.kapshtyk@gmail.com.*

Hlib Avdiienko, *PhD (Tech.), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0000-0002-4788-7273, django2006@ukr.net.*

Abstract. *The paper presents a comprehensive survey of hardware and algorithmic solutions for building high-speed wireless systems in the terahertz (THz) band that target sixth-generation (6G) networks and distributed low-Earth-orbit satellite constellations. The topic is driven by the rapid growth of IoT devices*

and the need for global data exchange with latency < 1 ms and throughput > 100 Gb/s. A review of the literature and experimental prototypes shows that the combination of massive-MIMO, hybrid beamforming and 100–300 GHz channels can yield spectral efficiencies up to 20 bit/s/Hz while keeping the number of RF chains moderate. Laboratory radio-relay links at 130–134 GHz already deliver 1.2 Gb/s over 1 km with BER ≤ 10^{-6}, proving the viability of THz interfaces for both urban and space scenarios. Special emphasis is placed on the "distributed satellite" concept, which integrates fog computing and inter-satellite links based on 5G-NTN and NB-IoT, allowing local telemetry processing and easing the load on the ground segment.

The surveyed materials indicate that the transition to monolithic GaN RFICs with built-in phase-shifters can raise the energy budget by 30%, while composite graphene heat spreaders reduce peak chip temperatures by 15 °C. Outstanding challenges remain, including vacuum calibration of antenna arrays, mitigation of phase noise above 200 GHz, and spectrum-sharing coordination between active and passive services. The proposed roadmap covers orbital experiments with a CubeSat cluster, the development of self-learning beam-control algorithms, and the creation of adaptive FEC codes resilient to Doppler shifts. Thus, the study offers a holistic vision for moving from laboratory testbeds to full-scale THz networks capable of becoming the backbone of the 6G ecosystem and the global Internet of Things.

Key words: *terahertz band; 6G; distributed satellite; hybrid beamforming; MIMO; inter-satellite links.*

Вступ

Нестримне збільшення медіаресурсів створює нові виклики для технологій передачі даних. Створення й упровадження нових технологій, а також збільшення інформації, що потрібно передати, призводить до безпрецедентного зростання кількості взаємозалежних пристроїв, що, у свою чергу, потребує збільшення швидкості передачі даних і нових технологій для задоволення все більших потреб у зв'язку, обміні даними, навігації та інших сферах. У відповідь на цю проблему розглядається можливість використання інших частотних діапазонів у радіочастотному спектрі, що передбачає не лише покращення наявних технологій, а й розширення діапазонів частот, які раніше не використовувалися.

Виконання досліджень

Одним із таких є ТГЧ-діапазон (терагерцовий діапазон частот). Згідно з рекомендаціями Міжнародного союзу електрозв'язку (Регламент радіозв'язку 5 стаття), ТГЧ-діапазон займає частоти від 300 ГГц до 3 ТГц (рис. 1). Проте в найбільш широкій інтерпретації цей діапазон займає ділянку частот від 100 ГГц до 3 ТГц [1]. Терагерцовий діапазон частот дає унікальні можливості для безпроводових систем зв'язку (відносно мала дальність, низька завантаженість зв'язку) завдяки широкій смузі пропускання, що забезпечує передачу даних зі швидкістю, недосяжною для нижчих частотних діапазонів [2].

Створено цифрові радіорелейні системи з пропускною здатністю до 1200 Мбіт/с у діапазоні частот 130–134 ГГц. Розроблені лабораторні зразки включають малогабаритні елементи, такі як смугові фільтри й рупорні антени з лінзовими концентраторами [2].

Проведені випробування підтвердили ефективність зв'язку в ТГц діапазоні на відстанях до 1 км за умови низького рівня помилок (BER ≤ 10^{-6}). Дослідження показали стабільність передачі сигналів при використанні сучасних схем модуляції, таких як КАМ-64. Ці результати свідчать про перспективність застосування ТГц технологій у складних середовищах, таких як міські або промислові зони [2].

Програма 6Genesis Flagship, яка займається розробкою технологій для мобільних мереж шостого покоління (6G), передбачає значний прорив у частотних діапазонах, швидкості передачі даних та ефективності використання ресурсів мережі. Використання терагерцового діапазону (ТГц) розглядається як один із ключових напрямів 6G, що дасть можливість забезпечити надвисоку пропускну здатність мережі радіодоступу, перехід до мікро- та пікасот, малу затримку й високу надійність для різноманітних сценаріїв надання послуг, включаючи розширений мобільний широкосмуговий зв'язок (eMBB), масовий зв'язок машинного типу (mMTC) і високонадійні комунікації з малою затримкою (uRLLC). Упровадження цих технологій створить фундамент для інтеграції інтелектуальних систем, що функціонують у різних середовищах, таких як повітряний простір, космос, наземні мережі й навіть підводні комунікації [3].

Технологія MIMO (Multiple Input Multiple Output) є основою для 5G і 6G мереж, підвищуючи пропускну здатність і надійність завдяки одночасній передачі кількох потоків даних. У субтерагерцовому діапазоні (100 ГГц-1 ТГц) MIMO дає змогу досягати швидкостей понад 100 Гбіт/с. Наприклад, 2×2 MIMO-система на частотах 142 ГГц і 285 ГГц з модуляцією 16-QAM забезпечила пропускну здатність 126 Гбіт/с на канал шириною 12,5 ГГц [12].

Реалізація MIMO:

5G – у сантиметровому (3–30 ГГц) і міліметровому (30–300 ГГц) діапазонах масивні антенні решітки (massive MIMO) компенсують утрати сигналу. Наприклад, базові станції 5G у міських районах використовують 64×64 MIMO для одночасного обслуговування десятків користувачів.

6G – очікується інтеграція ТГц-частот із сотнями антенних елементів для просторового мультиплексування, що підтримує сценарії типу голографічного зв'язку чи масового IoT.

Особливе значення в розробці 6G надається системам передачі даних у терагерцовому діапазоні, які вимагають оптимізації апаратних засобів і створення нових матеріалів. Учасники програми

6Genesis працюють над створенням ефективних антен із великим діапазоном частот, використовуючи наноматеріали та тривимірний друк. Це дасть змогу розробляти енергоефективні системи з високою продуктивністю, які будуть здатні працювати в умовах високої щільності пристроїв і передачі великих обсягів даних у космічних і наземних системах [3].

Основними перевагами й перспективами розвитку терагерцового діапазону частот для систем зв'язку є перехід до цього діапазону, який здатен суттєво підвищити пропускну здатність мереж, що є критично важливим для державних, військових і спеціалізованих систем зв'язку. Однак реалізація цього потенціалу вимагає створення високоефективних засобів генерації, оброблення та випромінювання сигналів у ТГц діапазоні частот.

Використання субтерагерцевих частот у системах MIMO є важливим складником технології 5G, що дає змогу підвищити пропускну здатність та енергоефективність завдяки використанню великої кількості незалежних керованих променів для одночасної передачі кількох потоків даних до різних абонентів. У системі стандартів 5G передбачається освоєння нових, більш високих частотних діапазонів, для збільшення пропускної здатності мережі й покращення послуг, що надаються. У цьому напрямі перспективним уважається освоєння субтерагерцових частот (від 100 ГГц до 1 ТГц) у технологіях 5G, 6G технологіях, для досягання швидкостей передачі даних більше ніж 100 Гбіт/с. На цей час уже реалізовано 2×2 MIMO-систему, що працює на частотах 142 ГГц і 285 ГГц. Створена система показала спроможність забезпечити пропускну здатність для 126 Гбіт/с для 16-QAM модуляції при використанні 12,5 ГГц смуги пропускання для кожного каналу [12].

Важливим елементом технології MIMO є метод формування незалежних спрямованих променів і їх орієнтації по визначеному напрямку. Для цього можуть застосовуватися цифрові й аналогові технології фазованих антенних решіток. Аналогова технологія передбачає використання фазообертачів або ліній затримок зі змінною затримкою для формування приймального та передавального променів у визначеному напрямку. Цифрова технологія здійснює операції формування променів цифровим методом, наприклад, використовуючи алгоритм оберненого двомірного цифрового перетворення Фур'є.

Технологія MIMO (Multiple Input Multiple Output) є ключовою для систем 5G, особливо в сантиметровому (cmWave, 3-30 ГГц) та міліметровому (mmWave, 30-300 ГГц) діапазонах, де великомасштабні антенні решітки дають змогу компенсувати втрати на шляху поширення сигналу [13]. Для реалізації MIMO у 5G використовуються різні архітектури, кожна з яких має свої переваги й обмеження.

Архітектура базової смуги: кожен антенний порт керується окремим трансивером, що забезпечує високу гнучкість у формуванні променів, але потребує значних енергетичних витрат і складного апаратного забезпечення.

RF-орієнтована архітектура: формування променів відбувається на радіочастотному рівні, зменшуючи кількість трансиверів та енергоспоживання, але з обмеженою гнучкістю.

Гібридна архітектура: комбінує RF і базову смугу, оптимізуючи ресурси й забезпечуючи високу пропускну здатність.

Методи формування променів включають сітку променів (Grid-of-Beams), нульове примушення (Zero-Forcing), методи на основі кодових книг і формування власного променя (Eigen-Beamforming), кожен із яких адаптований до конкретних умов експлуатації [13].

Гібридне формування променів (Hybrid Beamforming, HBF) оптимізує використання антенних масивів, комбінуючи аналогові та цифрові методи. У mmWave massive MIMO HBF зменшує кількість RF-ланок, використовуючи фазообертачі й перемикачі. Дослідження в Університеті Ліможа показали, що архітектура PFC (Partially/Fully Connected) із дробовим програмуванням досягає спектральної ефективності до 20 біт/с/Гц [15]. Наприклад, HBF застосовується в 5G-станціях для спрямованого зв'язку з рухомими об'єктами, такими як автомобілі чи дрони, зберігаючи енергоефективність.

Проблеми включають складність калібрування антен і потребу в адаптації до динамічних умов. У ТГц-діапазоні гібридна технологія долає обмеження аналогових методів, розбиваючи антени на групи з апаратно заданими фазовими зсувами, що спрощує управління променями [15].

Терагерцовий діапазон частот охоплює частоти від 100 ГГц і вище. Довжина хвилі в цьому діапазоні становить 3мм і менше. Такі показники суттєво ускладнюють використання як аналогових методів на базі керованих фазообертачів, так і цифрових методів. Перспективним напрямом для цього діапазону вважається використання гібридної технології формування променів. У цій технології формування променів здійснюється у 2 етапи. Опромінювачі фазованої антенної решітки розбиті на групи, для яких апаратно визначені фазові зсуви. Цифрова частина формує розподіл фаз сигналів, що передаються/приймаються на ці групи, з метою управління сформованим променем у просторі.

Спрямування сигналів з антен у потрібному напрямку використовують технології формування променів. Виділимо кілька основних: вузькосмугове (Narrowband Beamforming) застосовується в системах із малим частотним діапазоном, але неефективне для високошвидкісних мереж. Широкосмугове (Wideband Beamforming), у свою чергу, – для мереж із надширокосмуговими системами з дуже великим пропусканням, що є дуже корисним у нашому випадку. Фіксоване формування використовується в ситуаціях, де середовище є стабільним, проте його ефективність значно погіршується в змінних умовах [9].

Використання просторово-часового кодування OSTBC (Orthogonal Space-Time Block Coding) дасть змогу ефективно передавати дані через кілька антен із забезпеченням стійкості до перешкод і шумів. Для динамічних умов, у нашому випадку для забезпечення стійкості сигналу в умовах руху, краще використовувати адаптивне кодування в каналу [11].

Дуже велике значення приділяється компонентам, які використовуються в обладнанні. Одними з таких є підсилювачів потужності на основі GaN, зокрема QPB3810, що включає інтегрований контролер біасу для оптимальних точок налаштування, забезпечує середню потужність 8 Вт у діапазоні 3,4–3,8 ГГц. Це дає змогу значно зменшити розміри порівняно з традиційними рішеннями на окремих компонентах QPB9380. Інтегрований модуль для малих клітин BTS, що підтримує потужність до 20 Вт, використовується для TDD і mMIMO систем. Він об'єднує два етапи підсилення й вимикач для підтримки високої потужності в компактному розмірі. Завдяки GaN підсилювачам можна досягати високого ККД, отримуючи перевагу ще й у розмірах порівняно з традиційними рішеннями, що особливо важливо для застосувань у 5G та супутникових системах, де розмір і вага є критично важливими показниками [10].

У цьому напрямі особливе місце посідають програми Агентства перспективних оборонних дослідницьких проектів (DARPA). Наприклад, програма SWIFT (Sub-millimeter Wave Imaging Focal Plane Technology) спрямована на розвиток технологій субміліметрової візуалізації для фокальних площин, а програма TFAST (Technology for Frequency Agile Digitally Synthesized Transmitters) зосереджена на створенні частотно-агільних цифрових синтезованих передавачів. Не менш важливі розробки в рамках програми Terahertz Electronics, яка займається створенням високошвидкісних малошумливих транзисторів. Варто також згадати програму NEXT, що працює над створенням транзисторів на основі напівпровідникових нітридів. Ці інновації є ключовими для впровадження ТГц технологій у системи зв'язку майбутнього [4].

ТГц технології відкривають нові горизонти для транспортних мереж мобільного зв'язку. Радіорелейні системи в цьому діапазоні забезпечують передачу великих обсягів даних, таких як відеопотоки чи графічні дані, зі швидкістю понад 1 Гбіт/с. Такі системи використовуються для створення компактних і завадозахищених мереж, що дає змогу суттєво підвищити ефективність інфраструктури мобільного зв'язку нового покоління [2].

Інновації в галузі супутникових технологій сприятимуть подальшому розвитку тенденцій у передаванні й обробці інформації. Якщо супутники матимуть розширений функціонал і можливість працювати на вищих частотах, це забезпечить покращену якість передачі даних на великій швидкості та стане важливим кроком до розв'язання сучасних потреб у високошвидкісному інтернеті й розширенні можливостей глобальної мережі зв'язку. Корисне навантаження супутників перспективних систем є багатофункціональним. Його робота потребує передавання та прийому великих обсягів інформації, що, у свою чергу, передбачає наявність широкосмугових високошвидкісних каналів.

Для вирішення завдань швидкої передачі трафіку й зменшення часу реакції пропонується концепція «розподіленого супутника», що адаптована під вимоги інтернету речей. Вона передбачає використання кількох супутників малого класу, які спільно виконують функції одного великого супутника. Приклади проєктів:

Starlink – тисячі низькоорбітальних супутників для широкосмугового інтернету зі швидкістю до 300 Мбіт/с.

OneWeb – мережа для зв'язку у віддалених регіонах із затримкою 20–30 мс.

IoT-застосування – ретрансляція даних від сенсорів у реальному часі, наприклад, для моніторингу клімату чи логістики.

Низькоорбітальні системи (500–1100 км) забезпечують затримки 1,67–5,2 мс, а інтеграція туманних обчислень дає змогу обробляти дані на орбіті, зменшуючи навантаження на наземні канали [14]. Координація між супутниками здійснюється через GPS і міжсупутникові зв'язки (ISL) з алгоритмами типу «Leader-Follower». Сенсори SAW забезпечують компактність та енергоефективність, а протоколи LoRaWAN і NB-IoT адаптуються до космічних умов.

Такий підхід дає змогу зменшити масу й енергоспоживання супутників, а також скоротити затримки передачі даних. У такій системі група супутників може виконувати функціональні завдання, а також забезпечувати локальні обчислення, що відповідає концепції туманних обчислень, наближаючи обчислювальні потужності до джерел інформації в космосі.

Використання мікро- й наносупутників (зокрема кубсатів) для створення розподіленої архітектури є інноваційним підходом у сучасній космонавтиці. Розподіл функціональних блоків корисного навантаження між кількома супутниками дає можливість оптимізувати їхні ресурси, знизити витрати на виробництво та запуск, а також забезпечити гнучкість у виконанні завдань. Така структура сприяє створенню більш стійких та адаптивних систем [5].

Архітектура розподіленого супутника базується на принципі розподілу функціональних завдань між кількома малими космічними апаратами, такими як мікросупутники й кубсати, що працюють як єдина система. Основна ідея полягає в тому, щоб використовувати легші та менш дорогі апарати замість одного великого супутника. Ця концепція дає змогу збільшити гнучкість системи й забезпечити

виконання складних завдань, таких як зондування Землі, забезпечення телекомунікацій і підтримка інтернету речей (IoT).

Низькоорбітальні супутникові системи для IoT пропонують глобальне покриття завдяки групі наноспутників, які виконують функції ретрансляції, обчислень і маршрутизації. Концепція «розподіленого супутника» інтегрує «туманні» та «граничні» обчислення, даючи змогу обробляти дані IoT у реальному часі безпосередньо на орбіті, зменшуючи затримки й навантаження на канали зв'язку. Затримки передачі даних для орбіт 500–1100 км становлять 1,67–5,2 мс [14].

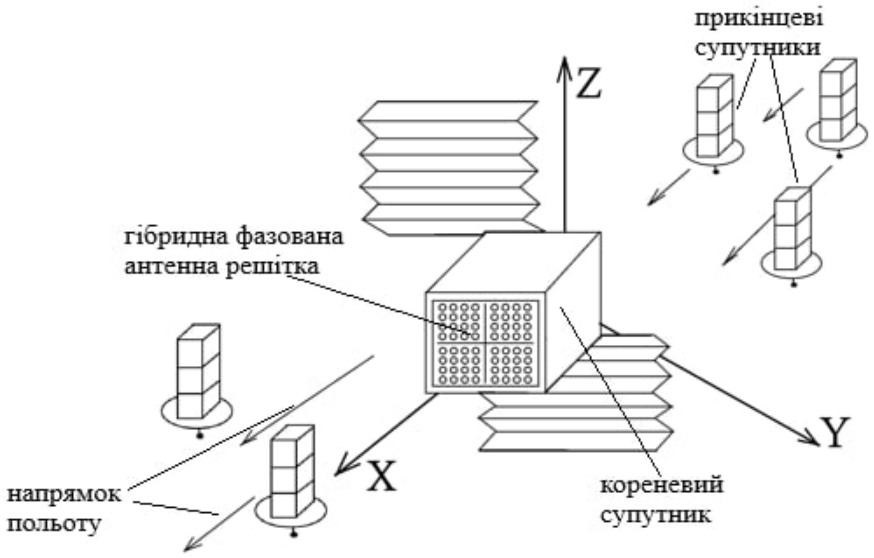


Рис. 1. Архітектура розподіленого супутника

Використовується архітектура двох типів: централізована й децентралізована. Коли один кореневий супутник виконує функції управління групою, забезпечуючи контроль польоту й координацію між іншими супутниками, і коли управління розподілене між усіма супутниками, що дає змогу уникати центральної точки відмови відповідно [8].

У розподілених супутникових системах використовуються адаптивні протоколи, які дають можливість ураховувати особливості космічного середовища, такі як високі швидкості, шум і вплив ефекту Доплера. Основні протоколи, які використовуються: 3GPP (4G, 5G), LoRaWAN та NB-IoT, а також DVB-S2 [8].

Досягнуто значний прогрес у використанні стандартних одноплатних комп'ютерів (COTS – Commercial Off-The-Shelf) для створення обчислювальних модулів супутників. Це дало змогу значно знизити вартість розробки й упровадження супутникових систем, оскільки використовуються вже протестовані компоненти. Крім того, застосування таких технологій забезпечує високу гнучкість і можливість швидкого оновлення апаратного забезпечення відповідно до змін у вимогах IoT [6].

Документ підкреслює важливість застосування комерційно доступних компонентів (COTS) у супутникових системах. Зокрема, використання Wi-Fi-приймачів і GPS-приймачів дає змогу забезпечити ефективну комунікацію між супутниками, а також точне визначення їхнього відносного положення. Сенсори на основі поверхневих акустичних хвиль (SAW) вирізняються компактними розмірами й енергоефективністю, що робить їх ідеальними для інтеграції в малі супутники, такі як CubeSat. Для зв'язку між супутниками використовуються стандарти IEEE 802.11 та IEEE 802.15.4, які, хоч і розроблені для наземних систем, демонструють значний потенціал у космічному середовищі завдяки своїй надійності й гнучкості. Це рішення дає змогу створювати бездротові мережі як для комунікації між супутниками, так і для передачі даних усередині одного супутника [7].

Що стосується координації супутників, варто зазначити такі особливості: GPS-приймачі на кожному супутнику використовують сигнали від глобальної GPS-супутникової системи для визначення своїх координат у тривимірному просторі. Це здійснюється шляхом обчислення затримки сигналу, що надходить від декількох GPS-супутників одночасно. Отримані GPS-координати використовуються для синхронізації даних між супутниками. Супутники у формації обмінюються інформацією через міжсупутникові зв'язки (ISL), наприклад, Wi-Fi, що дає змогу коригувати їхні відносні траєкторії в режимі реального часу [7].

Для підтримання формації супутники використовують алгоритми управління відносною траєкторією. Використовується Leader-Follower конфігурації, де один супутник є ведучим, а інші коригують свої

траєкторії стосовно нього. Алгоритми враховують динамічні збурення, такі як аеродинамічний опір і сонячне випромінювання, що викликають зміщення супутників. Для підвищення точності використовуються фільтри Калмана, які обробляють дані GPS і коригують відносні положення з урахуванням шумів вимірювань і моделей орбітальної динаміки. Це дає змогу досягати точності позиціонування в межах кількох метрів навіть за умови високих швидкостей руху супутників [7].

Таким чином, концепція «розподіленого супутника» відкриває нові можливості для забезпечення широкосмугового зв'язку й ефективної обробки даних у рамках інтернету речей. Це дасть змогу вирішити проблеми передачі великих обсягів трафіку, покращити час відгуку та сприяти подальшому розвитку космічних технологій.

Перспективи подальших досліджень

– Інтелектуальне керування променями. Створити алгоритми машинного навчання для прогнозування тіньових зон у міжсупутникових лінках і динамічного налаштування гібридної ФАР у реальному часі.

– Ультращільне кодування протоколів. Розширити протоколи 5G-NTN адаптивними схемами FEC, стійкими до фазових шумів і доплерівських зсувів на ТГц-частотах.

– Теплова стабілізація антен. Дослідити композитні матеріали з високою теплопровідністю й малим коефіцієнтом теплового розширення для масивів, що працюють у вакуумі.

– Орбітальний тест-стенд. Провести місію-демонстратор із групою CubeSat для валідації ISL на 60–140 ГГц і збору статистики каналу в різних сонячних циклах.

Висновок

Огляд узагальнив ключові наукові й інженерні досягнення у сфері ТГц-зв'язку для розподілених супутникових систем і показав, що комбінація масивних MIMO, гібридного формування променів і GaN-електроніки здатна забезпечити кратне зростання пропускної здатності без непомірного збільшення енергоспоживання. Демонстраційні експерименти довели реальність лінків 1200 Мбіт/с на діапазоні 130–134 ГГц, а модель «розподіленого супутника» підтвердила потенціал скорочення затримок до кількох мілісекунд завдяки туманним обчисленням і локальній маршрутизації трафіку. Разом із тим робота виявила вузькі місця: складність теплового менеджменту, потребу в адаптивних протоколах та обмеження чинної регуляторної бази. Запропоновані напрями майбутніх досліджень мають на меті закрити ці прогалини, уможлививши поступ від лабораторних прототипів до повномасштабних орбітальних сервісів. У підсумку можна стверджувати, що системи ТГц-зв'язку є одним із ключових каталізаторів еволюції 6G-екосистеми та глобальної інфраструктури інтернету речей, а реалізація запропонованих досліджень наблизить створення стійких, високошвидкісних та енергоефективних супутникових мереж наступного покоління.

Література

1. Ilchenko M. Ye., Kuzmin S. Ye., Narytnik T. N., Belous O. I., Radzikhovsky V. N. Transceiver for 130–134 GHz band digital radio relay system *Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika).* 2013. № 72(17). P. 1623–1638.
2. Saiko V., Toliupa S., Brailovskyi M., Narytnyk T., Nakonechnyi V., Shtanenko S. Mathematical Simulation of FMCW Radar Operation: Simulation of the Normalized Signal at the Receiver Input. *Proceedings of the 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)* / Lviv, Ukraine, 21–25 November 2023. IEEE, 2023. Electronic ISBN 979-8-3503-7257-1; Print on Demand ISBN 979-8-3503-7258-8. DOI: 10.1109/AICT61584.2023.10452695.
3. 6Genesis Flagship Program: Building the Bridges towards 6G-enabled Wireless Smart Society and Ecosystem. http://jultika.oulu.fi/files/nbnfi-fe2019081624413.pdf.
4. Майборода І. М., Стороженко І. П., Бабенко В. П., Кайдаш М. В. Огляд досягнень в терагерцових комунікаційних системах. *Збірник наукових праць Національної академії Національної гвардії України.* 2016. Вип. 1. С. 45–46.
5. Ільченко М. Е., Наритник Т. Н., Рассомакін Б. М., Присяжний В. І., Капштик С. В. Створення архітектури «Розподіленого супутника» для низькоорбітальних інформаційно-телекомунікаційних систем на основі групи мікро- і наносупутників. *Інформаційні технології управління підприємствами, програмами і проектами.* С. 33-34.
6. Ilchenko M. Y., Narytnik T. N., Fisun A. I., Belous O. I. Terahertz range telecommunication systemsIlchenko. *Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika).* 2011. № 70(16). P. 1477–1487.
7. Ferrario A., Furlan B. CubeSat formation flying mission as Wi-Fi data transmission and GPS based relative navigation technology demonstrator. POLITECNICO DI MILANO. Academic Year, 2010–2011.
8. Ilchenko M. Ye., Narytnik T. N., Didkovsky R. M. Clifford algebra in multiple-access noise-signal communication systemsIlchenko. *Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika).* 2013. № 72(18). P. 1651–1663.
9. Ali E., Ismail M., Nordin R., Abdullah N. Beamforming techniques for massive MIMO systems in 5G: overview, classification, and trends for future research. *Frontiers of Information.* 2017. 1 June. https://link.springer.com/content/pdf/10.1631/FITEE.1601817.pdf.
10. Figueiredo F. A. P. de. Performance and Integration for 5G Massive MIMO. 2022. 14 April. C. 931–940.
11. Gershman A. B., Sidiropoulos N. D. Space-Time Processing for MIMO Communications. 2005. 22 April. ISBN: 9780470010020.
12. Jue G. A Sub-Terahertz MIMO Testbed for 6G Research. Keysight Technologies, Santa Rosa, Calif. 13 February 2024. https://www.microwavejournal.com/articles/41448-a-sub-terahertz-mimo-testbed-for-6g-research.

13. Ніколаєв С. Ю. та ін. Дослідження методів MIMO-інтеграції з технологіями п'ятого покоління. *Наукові записки ДУТ.* 2022. № 1–2(2). С. 46–53.

14. Ilchenco M. Ye., Kalinin V. I., Narytnik T. N., Cherepenin V. A. Wireless UWB ecologically friendly communications at 70 nanowatt radiation power *CriMiCo 2011 – 21st International Crimean Conference on Microwave and Telecommunication Technology, Conference Proceedings.* 2011. P. 355–356. IEEE. DOI: 10.1109/CRMICO.2011.6068964.

15. Beiranvand J. Hybrid beamforming using massive antenna arrays for fixed sensor networks. Université de Limoges, 2023.

References

1. Ilchenko, M. Ye., Kuzmin, S. Ye., Narytnik, T. N., Belous, O. I., & Radzikhovsky, V. N. (2013). Transceiver for 130–134 GHz band digital radio relay system. Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika), 72(17), 1623–1638.

2. Saiko, V., Toliupa, S., Brailovskyi, M., Narytnyk, T., Nakonechnyi, V., & Shtanenko, S. (2023). Mathematical simulation of FMCW radar operation: Simulation of the normalized signal at the receiver input. In Proceedings of the 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT) (pp. 1–6). IEEE. https://doi.org/10.1109/AICT61584.2023.10452695.

3. 6Genesis Flagship Program. (2019). Building the bridges towards 6G-enabled wireless smart society and ecosystem. Retrieved from: http://jultika.oulu.fi/files/nbnfi-fe2019081624413.pdf.

4. Maiboroda, I. M., Storozhenko, I. P., Babenko, V. P., & Kaidash, M. V. (2016). Review of achievements in terahertz communication systems. Collected Scientific Works of the National Academy of the National Guard of Ukraine, (1), 45–46 [in Ukrainian]

5. Ilchenko, M. E., Narytnyk, T. N., Rassomakin, B. M., Prysiazhnyi, V. I., & Kapshtyk, S. V. (n.d.). Development of the architecture of a "distributed satellite" for low-orbit information and telecommunication systems based on a group of micro- and nanosatellites. Information Technologies of Enterprise Management, Programs and Projects, 33–34 [in Ukrainian]

6. Ilchenko, M. Y., Narytnik, T. N., Fisun, A. I., & Belous, O. I. (2011). Terahertz range telecommunication systems. Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika), 70(16), 1477–1487.

7. Ferrario, A., & Furlan, B. (2011). CubeSat formation flying mission as Wi-Fi data transmission and GPS-based relative navigation technology demonstrator. Politecnico di Milano [in English].

8. Ilchenko, M. Ye., Narytnik, T. N., & Didkovsky, R. M. (2013). Clifford algebra in multiple-access noise-signal communication systems. Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika), 72(18), 1651–1663.

9. Ali, E., Ismail, M., Nordin, R., & Abdullah, N. (2017). Beamforming techniques for massive MIMO systems in 5G: Overview, classification, and trends for future research. Frontiers of Information Technology & Electronic Engineering, 18(6), 753–767. https://link.springer.com/content/pdf/10.1631/FITEE.1601817.pdf.

10. Figueiredo, F. A. P. de. (2022). Performance and integration for 5G massive MIMO. Conference Proceedings, April, 931–940.

11. Gershman, A. B., & Sidiropoulos, N. D. (2005). Space-time processing for MIMO communications. Wiley. ISBN: 9780470010020.

12. Jue, G. (2024). A sub-terahertz MIMO testbed for 6G research. Keysight Technologies. Retrieved from: https://www.microwavejournal.com/articles/41448-a-sub-terahertz-mimo-testbed-for-6g-research.

13. Nikolaiev, S. Yu., et al. (2022). Research on MIMO integration methods with fifth-generation technologies. Scientific Notes of the State University of Telecommunications, 1–2(2), 46–53 [in Ukrainian].

14. Ilchenko, M. Ye., Kalinin, V. I., Narytnik, T. N., & Cherepenin, V. A. (2011). Wireless UWB ecologically friendly communications at 70 nanowatt radiation power. In CriMiCo 2011 – 21st International Crimean Conference on Microwave and Telecommunication Technology, Conference Proceedings (pp. 355–356). IEEE. https://doi.org/10.1109/CRMICO.2011.6068964.

15. Beiranvand, J. (2023). Hybrid beamforming using massive antenna arrays for fixed sensor networks. Université de Limoges [in English].

Дата першого надходження рукопису до видання: 21.05.2025
Дата прийнятого до друку рукопису після рецензування: 17.06.2025
Дата публікації: 25.07.2025

UDC 004.8:004.4
DOI https://doi.org/10.36994/2788-5518-2025-01-09-04

COMPARATIVE EVALUATION OF AI-POWERED IDES: CURSOR AI AND WINDSURF IN SOFTWARE DEVELOPMENT WORKFLOWS

Petrenko S. V., *Ph.D., Ass. Prof., Rivne State University of the Humanities, Rivne, Ukraine. https://orcid.org/0000-0002-5311-0743, serhii.petrenko@rshu.edu.ua*

Abstract. *The emergence of AI-powered integrated development environments (IDEs) marks a transformative phase in software engineering, where intelligent agents actively assist in tasks such as code generation, navigation, refactoring, and deployment. This study presents a comparative evaluation of two prominent GenAI-driven IDEs – Cursor AI and Windsurf – with the objective of exploring their effectiveness, usability, and limitations in real-world programming scenarios. Drawing on a case-based methodology, both tools were deployed in the development of a UI frontend for a backend service, allowing for empirical observation of how these environments interpret context, respond to user prompts, and manage iterative development cycles.*

The analysis confirms that both Cursor AI and Windsurf function as agentic collaborators rather than passive utilities, capable of modifying workflows, managing dependencies, interpreting instructions, and adapting to runtime environments. Cursor AI offers notable flexibility by supporting its agent mode within the free plan, thus lowering the entry barrier for developers. In contrast, Windsurf emphasizes streamlined deployment workflows, providing Netlify integration and live previewing capabilities that enhance its utility for rapid prototyping.

Despite their strengths, several challenges emerged. Both tools exhibited issues such as hallucinated suggestions, redundant change cycles, and unanticipated modifications. These findings underscore the critical role of human oversight, especially in complex tasks requiring domain-specific accuracy. Additionally, the assistants' behaviors reflect current limitations in aligning AI-generated logic with developer mental models – particularly when reasoning about incomplete or ambiguous input.

This research contributes to the broader discourse on generative AI in software development by offering grounded, comparative insights into how intelligent IDEs function under real-world constraints. The results highlight the ongoing need for improved transparency, context retention, and user control in AI-agent design. By treating these environments not just as tools but as evolving teammates, the study offers a perspective for future enhancements, where productivity gains can be achieved without compromising development quality or workflow clarity.

Key words: *AI-powered development; IDE, Cursor AI; Windsurf IDE; code generation; software engineering tools; developer productivity; Generative AI; autonomous programming assistants.*

ПОРІВНЯЛЬНА ОЦІНКА СЕРЕДОВИЩ РОЗРОБКИ НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ: CURSOR AI ТА WINDSURF У РОБОЧИХ ПРОЦЕСАХ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Сергій Петренко, *к.п.н., доцент, Рівненський державний гуманітарний університет, Рівне, Україна. https://orcid.org/0000-0002-5311-0743, serhii.petrenko@rshu.edu.ua*

Анотація. *Поява інтегрованих середовищ розробки (IDE) зі штучним інтелектом знаменує собою трансформаційний етап в інженерії програмного забезпечення, де інтелектуальні агенти активно допомагають у виконанні таких завдань, як генерація коду, навігація, рефакторинг і розгортання. У дослідженні подано порівняльну оцінку двох відомих IDE на основі GenAI – Cursor AI й Windsurf – з метою вивчення їх ефективності, зручності використання й обмежень у реальних сценаріях програмування. Спираючись на кейс-методологію, обидва інструменти застосували для розробки інтерфейсу користувача для бекенд-сервісу, що дало змогу емпірично спостерігати за тим, як ці середовища інтерпретують контекст, реагують на підказки користувача та керують ітеративними циклами розробки.*

Аналіз підтверджує, що й Cursor AI, і Windsurf функціонують як агентні колаборатори, а не пасивні утиліти, здатні змінювати робочі процеси, керувати залежностями, інтерпретувати інструкції й адаптуватися до середовищ виконання. Cursor AI пропонує значну гнучкість, підтримуючи агентський режим у рамках безкоштовного тарифного плану, що знижує вхідний бар'єр для розробників. Windsurf, навпаки, робить акцент на спрощених робочих процесах розгортання, забезпечуючи інтеграцію з Netlify та можливості попереднього перегляду в реальному часі, що підвищує його корисність для швидкого створення прототипів.

Попри наявність низки сильних сторін, виявлені окремі проблемні аспекти. Обидва інструменти продемонстрували такі проблеми, як галюцинації, надлишкові цикли змін і непередбачувані модифікації. Ці висновки підкреслюють критичну роль людського нагляду, особливо в складних завданнях, що вимагають специфічної точності. Крім того, поведінка асистентів відображає наявні обмеження в узгодженні логіки ШІ з ментальними моделями розробників, особливо коли йдеться про неповні або неоднозначні вхідні дані.

Дослідження робить внесок у ширший дискурс про генеративний ШІ в розробці програмного забезпечення, пропонуючи обґрунтоване порівняльне розуміння того, як інтелектуальні IDE функціонують в умовах реальних обмежень. Отримані результати акцентують на сталій потребі забезпечення прозорості, збереження контексту й користувацького контролю в процесі розробки AI-агентів. Розгляд зазначених систем не лише як інструментів, а як еволюціонуючих партнерів у командній взаємодії відкриває перспективні напрями подальшого вдосконалення, у межах яких підвищення продуктивності може бути досягнуте без утрати якості розробки й чіткості організації робочого процесу.

Ключові слова: *розробка із застосуванням ШІ; IDE; Cursor AI; Windsurf IDE; генерація коду; інструменти інженерії програмного забезпечення; продуктивність розробника; генеративний ШІ; автономні помічники програміста.*

Introduction

In the rapidly evolving world of Generative AI (Gen AI), each day brings something new in artificial intelligence (AI). This “new” can be ordinary or unpredictable and curious – sometimes even a potential game changer. The current moment is a remarkable time for AI agents. OpenAI CEO Sam Altman has declared 2025 the “year of AI agents,” envisioning a future where intelligent tools not only automate millions of jobs but also operate entire companies independently of human oversight.

IBM defines an AI agent as a system or program capable of autonomously performing tasks on behalf of a user or another system by designing its own workflow and utilizing available tools [5]. Based on this, an AI-powered integrated development environment (IDE) is indeed an AI agent.

Cursor AI and Windsurf exemplify this by autonomously assisting developers in their everyday tasks – generating code snippets, offering intelligent refactoring suggestions, performing contextual searches, and dynamically interacting with developers through chat-based interfaces. Their ability to independently determine workflows, apply context-aware logic, and provide proactive assistance firmly positions them as modern AI agents in the software development lifecycle.

The emergence of such coding assistants is fundamentally reshaping software engineering. These tools extend traditional IDEs into intelligent agents capable of generating, testing, and refactoring code across large codebases. The evolution from conventional environments to “Intelligent Development Environments” signals a shift where developers orchestrate workflows with the help of AI rather than writing all code manually [6].

Empirical research has shown both potential and limitations of these tools. For example, AI coding assistants can improve task completion and self-perceived productivity, particularly in structured or educational contexts. However, they often produce results that require human repair, especially when lacking sufficient project context [4]. Furthermore, a recent survey of over 400 practitioners reveals that while developers embrace AI for testing and documentation, they remain cautious about its use in debugging or architectural design [1].

Another key concern is how well these tools align with developer mental models. Misaligned suggestions or unpredictable changes may hinder adoption. Researchers propose that personalized, context-aware interfaces and interaction patterns are essential to build trust and usability [4].

An empirical evaluation of IBM’s internal AI assistant, watsonx Code Assistant (WCA), was conducted among 669 enterprise users, complemented by usability testing with 15 participants. The study revealed nuanced effects: while the AI assistant often yielded net productivity gains, the level of benefit varied significantly across users. Developers reported a shift in responsibility regarding ownership of generated code, and motivations and expectations regarding speed and quality influenced usage patterns [8]. These findings underscore that productivity improvements hinge not only on functional capabilities but also on factors such as user trust, perceived responsibility, and organizational context.

Cursor AI and Windsurf are prominent examples of these new AI-augmented IDEs. Both offer similar foundational capabilities, including multi-file editing, integrated terminal commands, and LLM-powered agents. However, they differ in implementation details, interface behavior, and available functionality, such as memory handling or deployment tools [1; 4; 6; 8].

This study is dedicated to the analysis of AI-assisted software development environments, with a specific focus on Cursor AI and Windsurf. These tools represent a new class of intelligent agents that support developers in tasks such as code generation, navigation, refactoring, and contextual understanding. The article provides a comprehensive comparison of their functionalities, interaction paradigms, and user experience design, highlighting the ways in which these systems influence the software development process.

To ground the analysis in practical relevance, a case-based methodology is employed using a representative software project. This enables an in-depth exploration of how each tool integrates into the coding workflow, handles user input, and supports iterative development tasks.

Many developers already rely on familiar IDEs with customized hotkeys and optimized workflows, the increasing capabilities of AI-augmented environments raise the question of whether adopting such agents can enhance productivity without disrupting existing habits. Rather than positioning these tools as mere utilities, this paper conceptualizes them as semi-autonomous collaborators – akin to junior team members who learn project context and evolve their behavior over time. The study also aims to address unresolved questions in the field, including:

– the effectiveness of agent-empowered IDEs across end-to-end development workflows;

– the role of agentic interface design in aligning with developer intent and trust;

– and the actual patterns of adoption in non-experimental, developer-centered environments.

Through a mixed-methods approach – combining a controlled experiment, comparative feature analysis, and reflective evaluation – this paper contributes empirical insights into the capabilities and limitations of Cursor AI and Windsurf as intelligent development agents.

Research

To provide a comprehensive understanding of Cursor AI and Windsurf, the comparison covers multiple dimensions: installation and setup, interface usability, memory and rule management, available models, agent behavior, and overall responsiveness.

Acquaintance

The initial step involves downloading and installing the respective tools, which is a straightforward process. Both Windsurf and Cursor AI provide direct installation options via their official websites: [9] and [2], respectively.

The installation process is similar for both, quite fast and convenient. As shown in the screenshots, each environment offers intuitive navigation, with the right-hand panel dedicated to interactive communication with the AI model. This design choice aligns with contemporary UI/UX practices, facilitating fluid developer-agent collaboration from the outset.

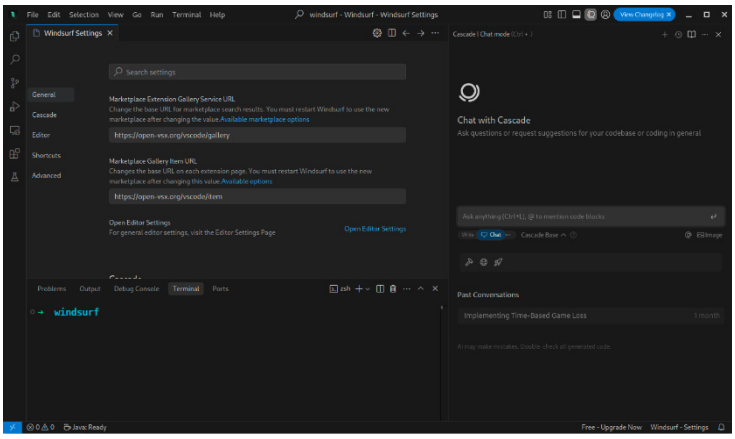


Fig. 1. Windsurf IDE

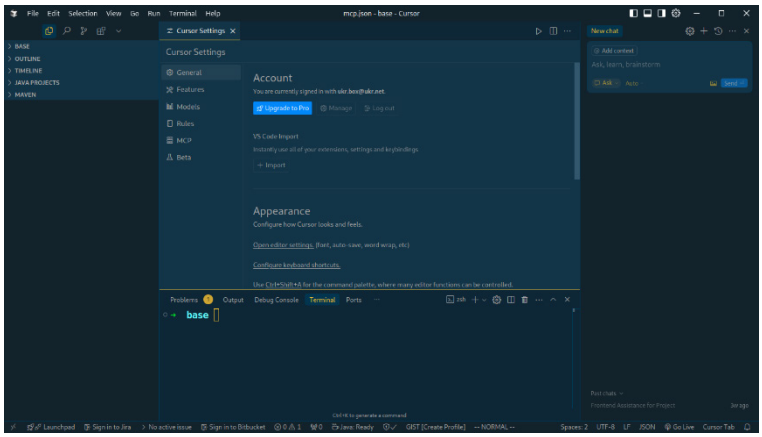


Fig. 2. Cursor AI IDE

Base

Both tools are forked from VS Code, providing a familiar user environment. After installation, each IDE conveniently offers to add all existing plugins from your current VS Code setup, greatly simplifying the transition. Both IDEs allow you to choose models for work, including ChatGPT, Claude, and others [3; 9].

Agent mode

Composer in Cursor – an agent that can be asked to formulate tasks in natural language right in the chat. It generates and executes commands in the terminal independently and can also determine what context it needs to understand [3]. If you're a Composer fan, you should thank the Windsurf team – they were the first to implement a similar idea in their Cascade [9]. While writing these materials, the Cursor has rolled out an update, and from now on Agent, Chat, and Composer are all in one [3; 9].

Memories and rules

For Cursor AI, we should have the **.cursorrules** file – it allows you to define project-specific instructions for Cursor and should be placed in the root directory of your project. Besides, Cursor has Notepads. They are ideal for capturing architectural decisions, coding standards, reusable snippets, team guidelines, and commonly accessed reference materials. Windsurf has a specific section where you can manage memories (like in ChatGPT) and tune your rules, which the user manually defines at local and global levels.

Differences

Until recently, Cursor had one unique feature – commit message generation, but on April 2, the Windsurf's team threw the same feature, though in beta. Also, Windsurf introduced Deploys (Beta), which helps to deploy your application with one prompt to Netlify under a **windsurf.build** domain. One more ability that looks great is a preview: previews in Windsurf allow you to view the local deployment of your app either in the IDE or in the browser with listeners, allowing you to iterate rapidly by easily sending elements and errors back to Cascade as context. An interesting thing I found in Cursor is ignoring files: **.cursorignore** blocks files from being added in chat or sent up for tab completions, in addition to ignoring them from indexing. Another update is that Cursor can now play a sound when a chat is ready for review. Both teams carefully study their competitors, so you don't have to worry that some cool things one team has won't be available to the other.

Price

The Windsurf Pro plan will cost you $15 per month, while the Cursor AI Pro plan will cost $20 (with an early plan, it will reduce to $16). Unfortunately, Cascade Base does not support Write mode on the Free plan, and you need a paid plan [3; 9].

Experiment

Details

As a backend engineer, my typical interaction with end-users is limited to designing and exposing APIs. Consequently, tasks that require a user interface (UI) layer often present a challenge, particularly in the absence of dedicated front-end resources. In team-based development, such responsibilities are typically delegated to front-end specialists. However, in contexts such as personal projects or proof-of-concept prototypes – where time and resources are constrained – developers frequently need to implement basic UI components themselves.

To simulate this scenario, an experiment was designed in which a frontend interface was created for an existing backend web service [7]. The selected use case was a simple tic-tac-toe game with standard functionality. The base project and resulting implementations are publicly available via the links provided at the end of the article. It is important to note that the frontend was developed by a non-specialist in UI/UX, and therefore, the visual and structural quality is left open for evaluation by experienced practitioners. Readers are encouraged to assess the resulting interfaces in terms of architectural clarity, adherence to best practices, and code readability [7].

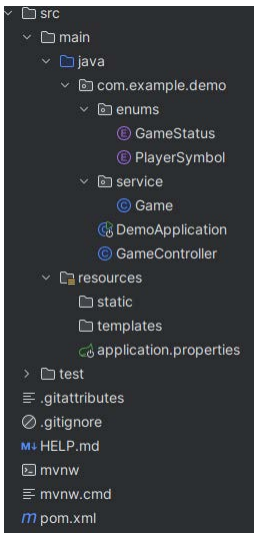


Fig. 3. The structure of base Java project

Configuration

1. Both AI tools will work on the same base project.

2. We will use the same prompts in both systems, except when the models are halting or making mistakes.

3. A comparison is performed on standard settings of environments using the same model. In our case, it is Claude 3.5 Sonnet from Anthropic.

Outcomes

The initial task assigned to both AI-powered IDEs mirrored a common developer practice when approaching unfamiliar code: performing a high-level analysis and generating a concise summary of the project. Both Cursor AI and Windsurf completed this task effectively, demonstrating their capability to comprehend code structure and provide meaningful overviews.

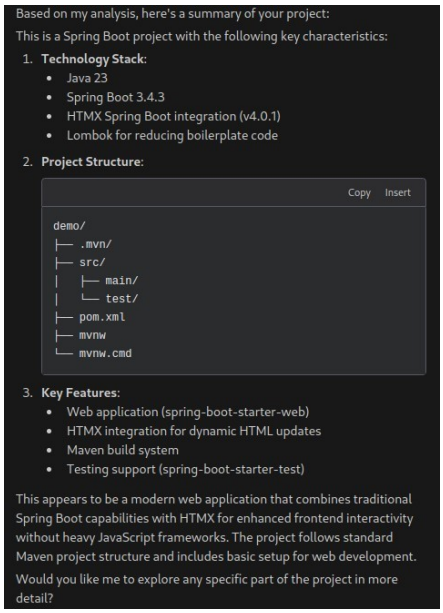


Fig. 4. Windsurf's first analysis

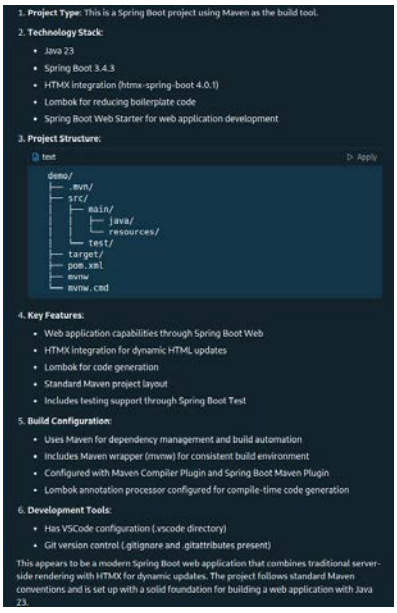


Fig. 5. Cursor's first analysis

Following the initial analysis, the assistants successfully identified the technology stack and overall project structure. The next step involved inferring the application's domain and underlying business requirements. Both AI tools demonstrated the ability to extract and articulate the intended purpose of the software with minimal guidance, effectively interpreting contextual cues and providing coherent summaries of the system's goals.

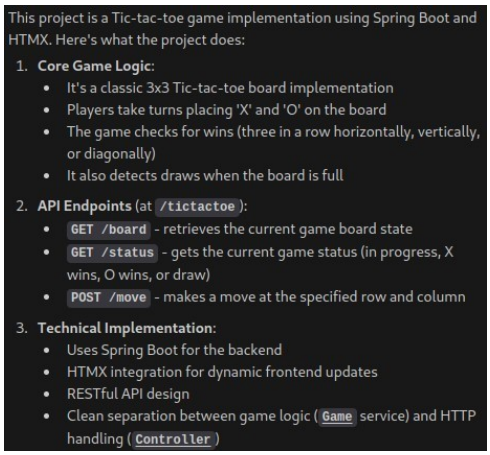


Fig. 6. Windsurf's respons

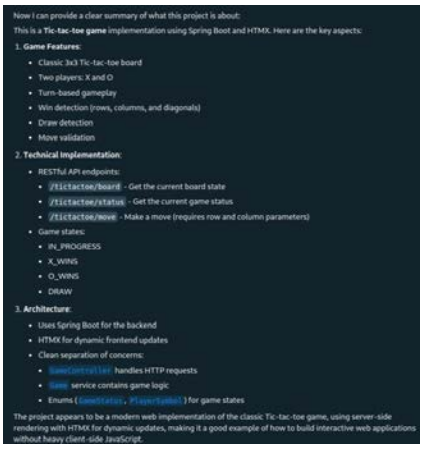


Fig. 7. Cursor's response

With the foundational analysis complete, the assistants were ready to begin addressing the core development tasks. As the interaction patterns in both AI-powered IDEs are largely analogous, the focus shifts to examining notable interaction instances, emergent behaviors, and opportunities arising during

agent-assisted development. Figure 8 illustrates the assistants' reasoning process and real-time code modifications, offering insight into their operational flow and decision-making strategies.

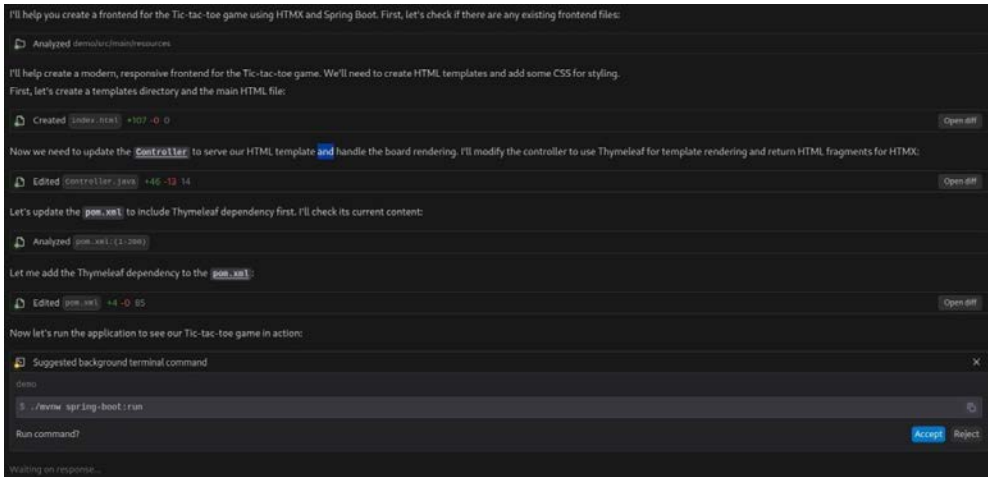


Fig. 8. Collaboration in the chat

You'll notice that you can check and compare differences for each change. Pay attention to how detailed each step is. It makes it easier to navigate faster and keep an eye on what's going on. In addition to that, when you don't like changes in a specific file, you can discard them. Even more, you can accept or reject them line by line (see Fig. 9).

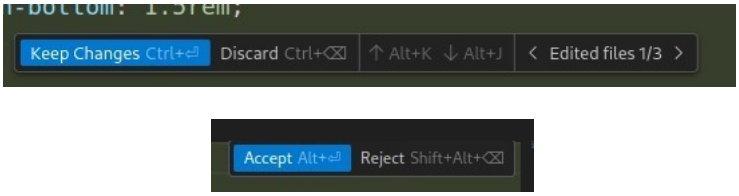


Fig. 9. Changes managing

We have finished with the changes, and IDE provides a command for the terminal to run our application and 2 buttons, accept or reject. After accepting, it will open the terminal window and run the command – looks great. In the course of deployment, the application encountered a compatibility issue related to the Java version. Although the backend service was developed using Java 23, this version was not available on the local machine. Remarkably, both AI-powered IDEs identified the issue from the terminal's stack trace, accurately diagnosed the root cause, detected the installed Java version in the current environment, and autonomously adjusted the project configuration to use Java 17. This behavior contrasts with typical responses from general-purpose AI chat interfaces, which often suggest upgrading or switching tools. In this case, the IDE agents demonstrated adaptive reasoning, behaving more like experienced developers by aligning the application's configuration with the actual runtime environment.

Once the application was successfully launched, functional testing commenced. Notably, although no explicit request was made for a "new game" button, Cursor AI autonomously included this feature. This behavior may be interpreted in two ways. On one hand, it could represent a hallucination – a known phenomenon where large language models generate content that was not prompted. On the other hand, it may reflect the model's proactive behavior, where the agent inferred the utility of such a feature based on common usage patterns and project context, thereby anticipating user needs.

The initial implementation resulted in a minimally functional application. With the core features in place, further enhancements were explored. Two seemingly simple features were selected for integration-deliberately chosen to evaluate the assistants' ability to interpret visual and contextual instructions. A screenshot was annotated with graphical markers and embedded comments to illustrate the intended modifications. This approach allowed for testing the agents' capacity to interpret visual prompts and translate them into actionable code changes (see Fig. 12).

Additionally, a request was made to Windsurf to implement a reset game button. The process of reviewing and accepting modifications in the generated files closely resembles the workflow of resolving merge conflicts in version control systems (Fig. 13).

Fig. 10. Windsurf's result

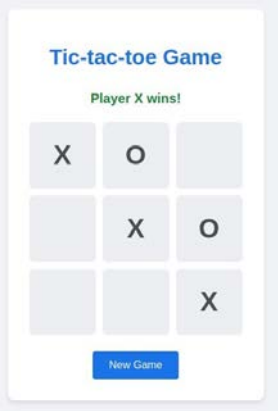


Fig. 11. Cursor's result

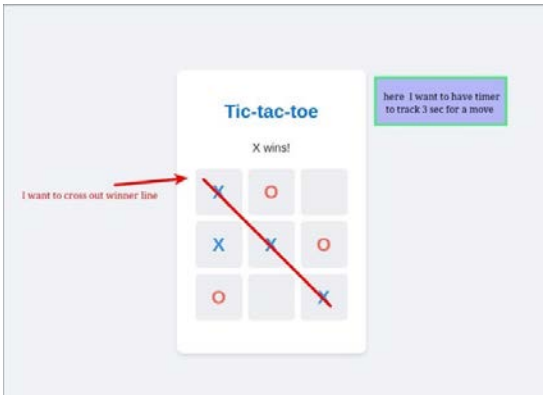


Fig. 12. Screen with instructions

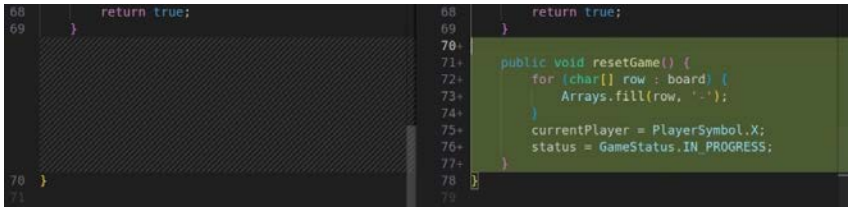


Fig. 13. Review changes

After one more round of modifications and adding new functionality, we need to run the app again and test it. And another feather in the cap: both agents detect that the previous server instance is still running, stop it, and restart the application.

We are approaching the end of our experiment. The application has a UI part, new functionality was added, and everything works as expected. I needed 20 requests to reach this result. Windsurf calculates prompts (each message) and flow action credits (each tool call: create, modify, analyze, search, terminal). Cursor AI is a bit simpler and calculates requests.

Fig. 14. Cursor's free account limits

Failed takes

1. It is not entirely clear what prompted Windsurf to modify the controller mappings. In the Java world, we have two approaches to annotating controllers. @RestController annotation to simplify the creation of RESTful web services. It's a convenient annotation that combines @Controller and @ResponseBody, and Windsurf replaced @RestController with @Controller and added @ResponseBody to each method.

2. Also, I'm not a big fan of such hardcoded parts that Windsurf did.

return String.format("%s %s ", board, status);

3. An interesting behavior that needs to be mentioned is that AI generates code and then detects errors in the code it just produced.

4. We are stuck in a cycle of hallucinations while fixing proper display cross out the winner line. When Cursor needed five shoots to fix the behavior, Windsurf did in 13. Also, on each iteration, AI made a lot of changes. Sometimes, it started checks, found issues, and automatically started to resolve them, but it took a lot of time. After 10 shorts, I asked to remove the line and start over. Lesson learned: we should interrupt the process when we notice such behavior.

Fig. 15. Hallucination examples

Conclusions

This study contributes to the growing body of research on AI-assisted software development by empirically examining Cursor AI and Windsurf – two modern integrated development environments enhanced with generative AI capabilities. Anchored in a case-based methodology, the evaluation focused on their capacity to support real-world development workflows, particularly in tasks related to code generation, contextual reasoning, and iterative refactoring.

Both tools demonstrated promising potential as agentic collaborators, autonomously adapting to environmental constraints and user input. The comparative analysis revealed several shared strengths, including rapid contextual understanding, multi-step interaction, and integration with LLM-based agents. At the same time, distinct differences emerged in terms of interface conventions, feature completeness, and pricing strategies.

Cursor AI's inclusion of a functional agent mode in its free plan offers an accessible entry point for experimentation, whereas Windsurf's reliance on a paid tier for core functionality presents a limitation for broader adoption. However, Windsurf compensates with innovative features such as real-time deployment previews and deploy-to-Netlify integration, reinforcing its appeal for rapid prototyping scenarios.

While both environments reflect active development and feature parity trends, the observed issues – such as hallucinated code modifications, redundant iterations, and non-intuitive decisions – highlight the need for further refinement in aligning agent behavior with developer expectations and mental models. These limitations underscore the importance of human oversight and interruption mechanisms in agentic IDEs.

Importantly, the findings affirm that AI-augmented IDEs should not be perceived as mere utilities but rather as evolving collaborators. Their value lies in supporting developers through high-level orchestration, not replacing human judgment. Future research could extend this evaluation to complex, large-scale software systems and team-based workflows to further investigate long-term adoption patterns and productivity impacts.

In conclusion, Cursor AI and Windsurf offer significant yet evolving contributions to the software engineering landscape. Their success will ultimately depend on continued enhancements in usability, transparency, and trustworthiness – qualities essential for integrating AI agents seamlessly into everyday development practices.

Bibliography

1. Agnia Sergeyuk, Yaroslav Golubev, Timofey Bryksin, Iftekhar Ahmed, Using AI-based coding assistants in practice: State of affairs, perceptions, and ways forward. *Information and Software Technology.* 2025. Vol. 178. P. 107610. ISSN 0950–5849. https://doi.org/10.1016/j.infsof.2024.107610.
2. Codeium. (n.d.). Windsurf – Codeium. URL: https://codeium.com/windsurf.
3. Cursor. (n.d.). Cursor. URL: https://www.cursor.com/.
4. Desolda G., Mazzini S., Polonio A., De Angeli A. Aligning AI programming assistants with developers' mental models: Lessons learned from empirical studies. *Advanced Engineering Informatics.* 2024. № 60. P. 102401. https://doi.org/10.1016/j.aei.2024.102401.
5. IBM. (n.d.). What is an AI agent? URL: https://www.ibm.com/think/topics/ai-agents.
6. Marron J. M., D'Antoni L., Teitelman R. Intelligent Development Environments: The Next Evolution of AI-Powered Software Engineering (arXiv:2404.12000v2). arXiv. 2024. https://doi.org/10.48550/arXiv.2404.12000.
7. Petrenko S. (n.d.). AI-powered [GitHub repository]. *GitHub.* URL: https://github.com/serhii-petrenko/ai-powered.
8. Weisz J. D., Kumar S., Muller M., Browne K.-E., Goldberg A., Heintze E., Bajpai S. Examining the Use and Impact of an AI Code Assistant on Developer Productivity and Experience in the Enterprise (arXiv:2412.06603v2). arXiv. 2025. https://doi.org/10.48550/arXiv.2412.06603.
9. Windsurf. (n.d.). Windsurf. URL: https://windsurf.com/.

References

1. Agnia Sergeyuk, Yaroslav Golubev, Timofey Bryksin, Iftekhar Ahmed, Using AI-based coding assistants in practice: State of affairs, perceptions, and ways forward, Information and Software Technology, Volume 178, 2025, 107610, ISSN 0950–5849, https://doi.org/10.1016/j.infsof.2024.107610.
2. Codeium. (б. д.). *Windsurf – Codeium*. https://codeium.com/windsurf.
3. Cursor. (б. д.). *Cursor*. https://www.cursor.com/.
4. Desolda, G., Mazzini, S., Polonio, A., & De Angeli, A. (2024). Aligning AI programming assistants with developers' mental models: Lessons learned from empirical studies. Advanced Engineering Informatics, 60, 102401. https://doi.org/10.1016/j.aei.2024.102401.
5. IBM. (n.d.). What is an AI agent? https://www.ibm.com/think/topics/ai-agents.
6. Marron, J. M., D'Antoni, L., & Teitelman, R. (2024). Intelligent Development Environments: The Next Evolution of AI-Powered Software Engineering (arXiv:2404.12000v2). arXiv. https://doi.org/10.48550/arXiv.2404.12000.
7. Petrenko, S. (б. д.). *AI-powered* [GitHub-репозиторій]. GitHub. https://github.com/serhii-petrenko/ai-powered.
8. Weisz, J. D., Kumar, S., Muller, M., Browne, K.-E., Goldberg, A., Heintze, E., & Bajpai, S. (2025). Examining the Use and Impact of an AI Code Assistant on Developer Productivity and Experience in the Enterprise (arXiv:2412.06603v2). arXiv. https://doi.org/10.48550/arXiv.2412.06603.
9. Windsurf. (б. д.). *Windsurf*. https://windsurf.com/.

Дата першого надходження рукопису до видання: 27.05.2025
Дата прийнятого до друку рукопису після рецензування: 30.06.2025
Дата публікації: 25.07.2025

УДК 621.396.67
DOI https://doi.org/10.36994/2788-5518-2025-01-09-05

ПОЛЯРИЗАЦІЙНІ ПАРАМЕТРИ 16-ЕЛЕМЕНТНОЇ ТА 64-ЕЛЕМЕНТНОЇ ФАР ДЛЯ РАДІОСИСТЕМ НВЧ ДІАПАЗОНУ

Почерняєв В. М., *д.т.н., проф., Національна академія Служби безпеки України, Київ, Україна. https://orcid.org/0000-0001-7130-8668, vpochernyaev@gmail.com*

Сивкова Н. М., *доктор філософії, Національна академія Служби безпеки України, Київ, Україна. https://orcid.org/0000-0002-4934-4109, natsivonat@gmail.com*

Магомедова М. С., *доктор філософії, Київський фаховий коледж зв'язку, Київ, Україна. https://orcid.org/0000-0003-1936-5555, kkz.praktika@ukr.net*

Гонтар С. В., *Київський фаховий коледж зв'язку, Київ, Україна. https://orcid.org/0009-0006-6862-3548, serg.hontar@gmail.com*

Анотація. *У статті рекомендується використання фазованих антенних решіток у мобільних цифрових комбінованих тропосферних системах військового та спеціального призначення. Відмічено, що для організації управління в військових конфліктах, при виникненні надзвичайних і позаштатних ситуацій у мирний час радіозв'язок НВЧ діапазону є основним типом зв'язку. Одним із важливих параметрів енергетичного потенціалу є коефіцієнт підсилення передавальної та приймальної антен. Використання високонаправлених фазованих антенних решіток у діапазоні НВЧ дає змогу суттєво підвищити надійність і стійкість зв'язку на тропосферних лініях зв'язку. Перспективним є створення мобільного комплексу цифрового тропосферного зв'язку, у складі якого є мобільна вузлова цифрова тропосферна станція й цифрові тропосферні станції різних модифікацій: типу Р-412 МУ, малогабаритні на автомобілях типу «джип», переносні (портативні). При цьому мобільна вузлова цифрова тропосферна станція є станцією, яка працює за схемою «точка-багатоточка». Виділено напрями розвитку радіотехнологій НВЧ діапазону в системі військового та спеціального зв'язку. Відмічено, що основними поляризаційними параметрами фазованих антенних решіток є коефіцієнт поляризації та кут орієнтації еліпса поляризації. Подано поляризаційні параметри фазованих антенних решіток і їх характеристики, розрахунки кутів сканування для ФАР-16 і ФАР-64. Приведено графічні залежності коефіцієнта поляризації та кута орієнтації еліпса поляризації від кутів сканування й геометричних параметрів ФАР-16 і ФАР-64. Отримано чисельні результати, що дають змогу розрахувати поляризаційні параметри фазованих антенних решіток, а також простежити динаміку зміни поляризаційних параметрів при різних кутах сканування. Показано, що амплітудний розподіл впливає на поляризаційні параметри залежно від кількості елементів фазованої антенної решітки практично однаково: чим більша відмінність в амплітудному розподілі правополяризаційних і лівополяризаційних елементів, тим більше коефіцієнт поляризації прямує до кругової поляризації того знака, амплітудний розподіл якого ближче до рівномірного.*

Ключові слова: *фазована антенна решітка; коефіцієнт поляризації; кут орієнтації еліпса поляризації; кут сканування; мобільні цифрові комбіновані тропосферні станції.*

POLARIZATION PARAMETERS OF 16-ELEMENTS AND 64-ELEMENTS PHASED ANTENNA ARRAYS FOR MICROWAVE RADIOSYSTEMS

Vitaly Pochernyaev, *Doctor of Technical Sciences, Professor, National Academy of Security Service of Ukraine, Kyiv, Ukraine. https://orcid.org/0000-0001-7130-8668, vpochernyaev@gmail.com*

Nataliia Syvkova, *Doctor of Philosophy, National Academy of Security Service of Ukraine, Kyiv, Ukraine. https://orcid.org/0000-0002-4934-4109, natsivonat@gmail.com*

Mariia Mahomedova, *Doctor of Philosophy, Kyiv Professional College of Communications, Kyiv, Ukraine. https://orcid.org/0000-0003-1936-5555, kkz.praktika@ukr.net*

Serhii Hontar*, Kyiv Professional College of Communications, Kyiv, Ukraine. https://orcid.org/0009-0006-6862-3548, serg.hontar@gmail.com*

Abstract. *The article recommends the use of phased antenna arrays in mobile digital troposcatter combined systems for military and special purposes. It is noted that for organizing control in military conflicts, in the event of emergency and abnormal situations in peacetime microwave radio communication is the main type of communication. One of the important parameters of the energy potential is the gain of the*

transmitting and receiving antennas. The use of highly directional phased antenna arrays in the microwave range allows for a significant increase in the reliability and stability of communications on tropospheric communication lines. A promising option is the creation of a mobile digital tropospheric communication complex, which includes a mobile nodal digital tropospheric station and digital tropospheric stations of various modifications: type R-412 MU, small-sized on vehicles such as "jeep" and portable. In this case, the mobile nodal digital tropospheric station is a station operating on the "point-to-multipoint" scheme. In the article highlights the development directions of microwave radio technologies in the military and special communications system. In the article notes that the main polarization parameters of phased antenna arrays are the polarization coefficient and the orientation angle of the polarization ellipse. The article presents the polarization parameters of phased antenna arrays and their characteristics, the calculation of scanning angles for PAA-16 and PAA-64. Graphic dependences of the polarization coefficient and the angle of orientation of the ellipse polarization on the scanning angles and geometric parameters of the FAR-16 and FAR-64 are given. Numerical results are obtained that allow calculating the polarization parameters of phased antenna arrays, as well as tracking the dynamics of changes in polarization parameters, both at different scans of the diagram DS and at different parameters of the amplitude distribution. It is shown that the amplitude distribution affects the polarization parameters depending on the number of elements of the phased antenna arrays in almost the same way: the greater the difference in the amplitude distribution of right-hand and left-hand polarization elements, the more the polarization coefficient is directed toward the circular polarization of the sign whose amplitude distribution is closer.

Key words: *phased antenna array; polarization coefficient; polarization ellipse orientation angle; scanning angle; mobile digital combined tropospheric stations.*

Вступ

Для організації управління у військових конфліктах, при виникненні надзвичайних і позаштатних ситуацій у мирний час радіозв'язок НВЧ діапазону є основним типом зв'язку. Питання підвищення надійності, стійкості та якості радіозв'язку (достовірності передачі інформації) пов'язані з енергетичним потенціалом радіолінії. Одним із важливих параметрів енергетичного потенціалу є коефіцієнт підсилення передавальної та приймальної антен. Широке впровадження фазованих антенних решіток (далі – ФАР) у радіолокації дало змогу рекомендувати використання ФАР у цифрових комбінованих тропосферних системах військового й спеціального призначення. Наприклад, корпорація Raytheon використовує діапазон 14 ГГц та кругову поляризацію для комбінованих тропосферно-космічних станцій військового призначення [1].

Використання високонаправлених ФАР у діапазоні НВЧ дає змогу суттєво підвищити надійність і стійкість зв'язку на тропосферних лініях зв'язку. Перспективним є створення мобільного комплексу цифрового тропосферного зв'язку, у складі якого є мобільна вузлова цифрова тропосферна станція (далі – МВЦТрС) і цифрові тропосферні станції (далі – ЦТрС) різних модифікацій: типу Р-412 МУ, малогабаритні на автомобілях типу «jeep», переносні (портативні). При цьому МВЦТрС є станцією, яка працює за схемою «точка-багатоточка». Схема організація зв'язку такого комплексу представлена на рис. 1.

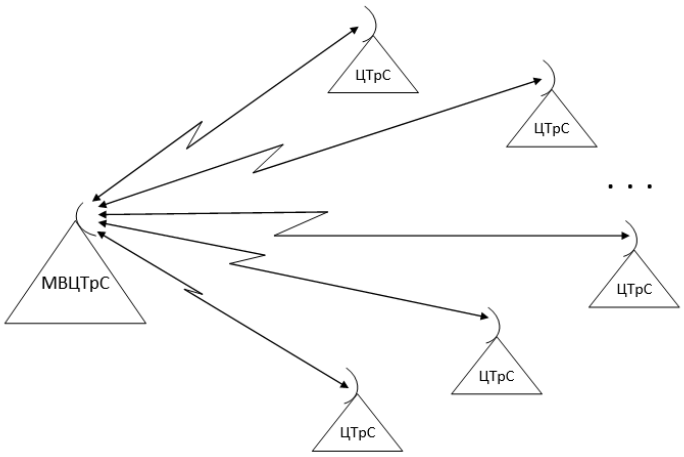


Рис. 1. Схема організації зв'язку «точка-багатоточка» на цифрових тропосферних станціях

Джерело: використано авторами за [1].

Відмітимо, що для телекомунікаційних НВЧ систем економічним є впровадження різних модифікацій ФАР та ефективність від такого впровадження буде вищою, ніж при використанні параболічних антен на штатних засобах зв'язку НВЧ діапазону в системах військового й спеціального зв'язку. Так,

пропонується використання 16-елементної ФАР (ФАР-16) і 64-елементної ФАР (ФАР-64) для мобільної комбінованої цифрової тропосферно-космічної станції [2] та мобільної вузлової цифрової тропосферної станції [3]. Тому комплекс цифрових тропосферних станцій включає таке: мобільну вузлову цифрову тропосферну станцію з ФАР-64 та кінцеві цифрові тропосферні станції у вигляді мобільних комбінованих цифрових тропосферно-космічних станцій з ФАР-16 і мобільних високошвидкісних цифрових тропосферних станцій з ФАР-16 [4].

Напрями розвитку радіотехнологій НВЧ діапазону в системі військового та спеціального зв'язку охоплюють таке:

1) передачу мультимедійної та графічної інформації в завадозахищеному режимі з високою пропускною заданістю;

2) реалізацію завадозахищених каналів цифрового радіозв'язку за рахунок розробки нових сигнально-кодових конструкцій і схем модуляцій з розширенням бази сигналів і мінімізації ширини смуги;

3) передачу інформації по каналах терагерцового діапазону частот і близького до нього ділянки спектру гігагерцового діапазону частот;

4) використання технологій ФАР різних модифікацій з вузьконаправленими діаграмами спрямованості;

5) уведення адаптивних систем контролю якості каналів радіозв'язку.

Аналіз літератури показує, що сучасні радіотехнічні системи НВЧ потребують великого різноманіття ФАР. Тому різноманітним питаннями створення ФАР присвячено багато публікацій:

– широкосмуговим і ширококутовим скануючим ФАР – [5; 8; 11; 16; 17; 19; 21; 28; 31; 33];
– створенню ФАР із високим коефіцієнтом підсилення – [6: 10];
– дослідженню та створення ФАР у високих діапазонах частот – [7; 15; 20; 29];
– використанню нових матеріалів при конструюванні ФАР – [9; 32];
– методиці випробувань ФАР – [12; 13];
– питанням синтезу ФАР – [14; 34];
– створенню активних ФАР [18; 22; 23; 24; 30];
– створенню моноімпульсної ФАР – [25];
– використанню фотонної технології при проєктуванні – [26].

Аналіз дає змогу зробити висновок, що одним із трендів розвитку цифрових радіосистем систем НВЧ діапазону є впровадження ФАР у різних діапазонах частот і для різних призначень [35].

Метою роботи є дослідження поляризаційних параметрів ФАР-64 та ФАР-16 для комплексу цифрових тропосферних станцій

Виконання досліджень

Основними поляризаційними параметрами ФАР є коефіцієнт поляризації P та кут орієнтації еліпса поляризації $\varphi_{ел}$.

Уважаємо, що ФАР складається з N^+ елементів кругової поляризації правого обертання (лінійної горизонтальної поляризації) та з N^- елементів кругової поляризації лівого обертання (лінійної вертикальної поляризації). У напрямку головного максимуму діаграми спрямованості (далі – ДС) ФАР її поляризаційні параметри запишуться в такому вигляді:

коефіцієнт поляризації:

$$P_{£} = \left| \frac{E^{+}_{£} - E^{-}_{£}}{E^{+}_{\Sigma} + E^{-}_{\Sigma}} \right|,$$

кут орієнтації еліпсу поляризації:

$$\varphi_{ел} = \frac{\Phi^{+}_{\Sigma} - \Phi^{-}_{\Sigma}}{2},$$

де $\overset{+}{E}{}^{-}_{£} = \left| \sum_{i=-(\overset{+}{N}{}^{-}-1/2)}^{\overset{+}{N}{}^{-}-1/2)} \overset{+}{A}{}^{-}_{i} \overset{+}{e}{}^{j\overset{+}{\Phi}{}^{-}_{i}} \right|$,

$$\overset{+}{\Phi}{}^{-}_{£} = arctg \frac{I_m \left\{ \overset{+}{E}{}^{-}_{\Sigma} \right\}}{R_e \left\{ \overset{+}{E}{}^{-}_{\Sigma} \right\}},$$

де $\overset{+}{A}{}^{-}_{i}$ – амплітудний розподіл праволяризованих (лінійно-горизонтально поляризованих) і лівополяризованих (лінійно-вертикально поляризованих) елементів, Φ^{-}_{i} – фазовий розподіл праволяризованих (лінійно-горизонтально поляризованих) і лівополяризованих (лінійно-вертикально поляризованих) елементів.

Запишемо поляризаційні параметри в напрямку максимуму ДС ФАР:

− коефіцієнт поляризації – P;
− кут орієнтації еліпса поляризації $\varphi_{ел}$;

− амплітудне розподілення на ФАР в *i*-елементі:

$$A_i^{\pm} = \delta^{\pm}\left[cos\left(\frac{\pi}{2}\frac{i}{N^2}\right)\right];$$

− фазовий розподіл на ФАР в *i*-елементі:

$$\psi_i^{\pm} = 2\pi\frac{d}{\lambda}\left(sin\varphi - sin\varphi_{ск}\right).$$

У ФАР будуть такі характеристики:

$$P = \left[\sum A^{+}e^{j\psi^{+}} - \sum A^{-}e^{j\psi^{-}}\right] / \left[\sum A^{+}e^{j\psi^{+}} + \sum A^{-}e^{j\psi^{-}}\right],$$

$$\varphi_{ел} = \frac{\sum \psi_i^{+} - \sum \psi_i^{-}}{2}.$$

Розрахунки кутів сканування для ФАР-16 для величини $0^0 \le \varphi_{ск} \le 45^0$ та $d/\lambda = 0{,}6;0{,}8$ показані на рис. 2a, b, c, d.

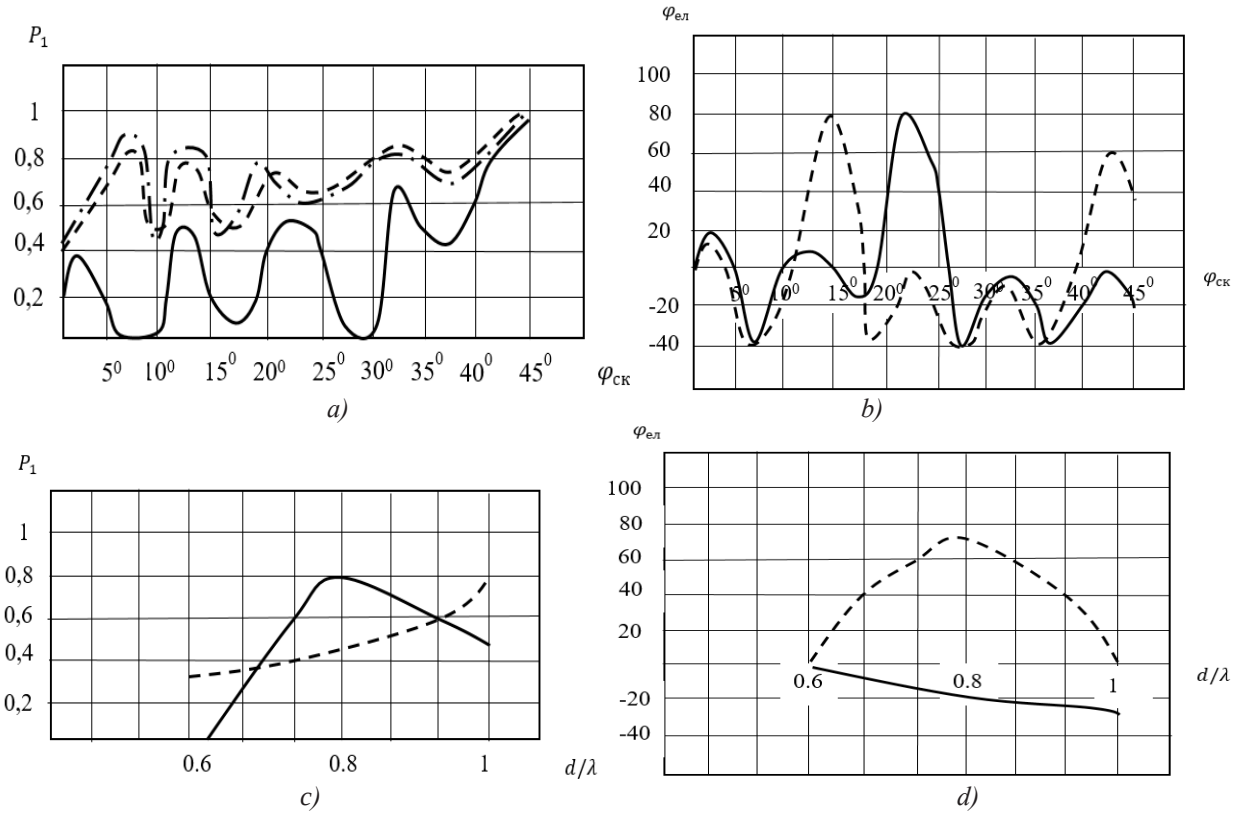


Рис. 2. Залежність коефіцієнта поляризації та кута орієнтації еліпса поляризації від кутів сканування й геометричних параметрів ФАР-16

Розглянемо залежність поляризаційних параметрів від кутів сканування, параметрів амплітудного розподілу й геометричних параметрів для ФАР-16.

При малих кутах сканування ($0^0 \le \varphi_{ск} \le 15^0$) і при $d/\lambda = 0{,}6$ функція $P = f\left(\varphi_{ск}\right)$ змінюється в межах 0 … 0,5, при середніх кутах сканування ($15^0 \le \varphi_{ск} \le 30^0$) – у межах 0,7 … 0, при великих кутах сканування ($30^0 \le \varphi_{ск} \le 45^0$) – у межах 0 … 0,95.

Залежність кута нахилу великої осі еліпса поляризації від кута сканування також має осцилюючий характер. Для тих самих геометричних параметрів $d/\lambda = 0{,}6$ кут $\varphi_{ел}$ змінюється в межах $-60^0 \ldots + 90^0$.

З графіків на рис. 2*a, b, c, d* видно, що при малих кутах сканування ($\varphi_{ск} \le 5^0$) коефіцієнт поляризації *P* суттєво залежить від відстані між елементами випромінювання. При середніх і великих кутах сканування ($\varphi_{ск} > 5^0$) коефіцієнт поляризації *P* майже залежить від відстані *d* між випромінюючими елементами.

Зазначимо, що при куті сканування $\varphi_{ск} = 0$ коефіцієнт поляризації *P=0,3*.

Розрахунки кутів сканування для ФАР-64 для величини $0^0 \le \varphi_{ск} \le 45^0$ та $d/\lambda = 0{,}6;0{,}8$ показані на рис. 3*a, b, c, d*.

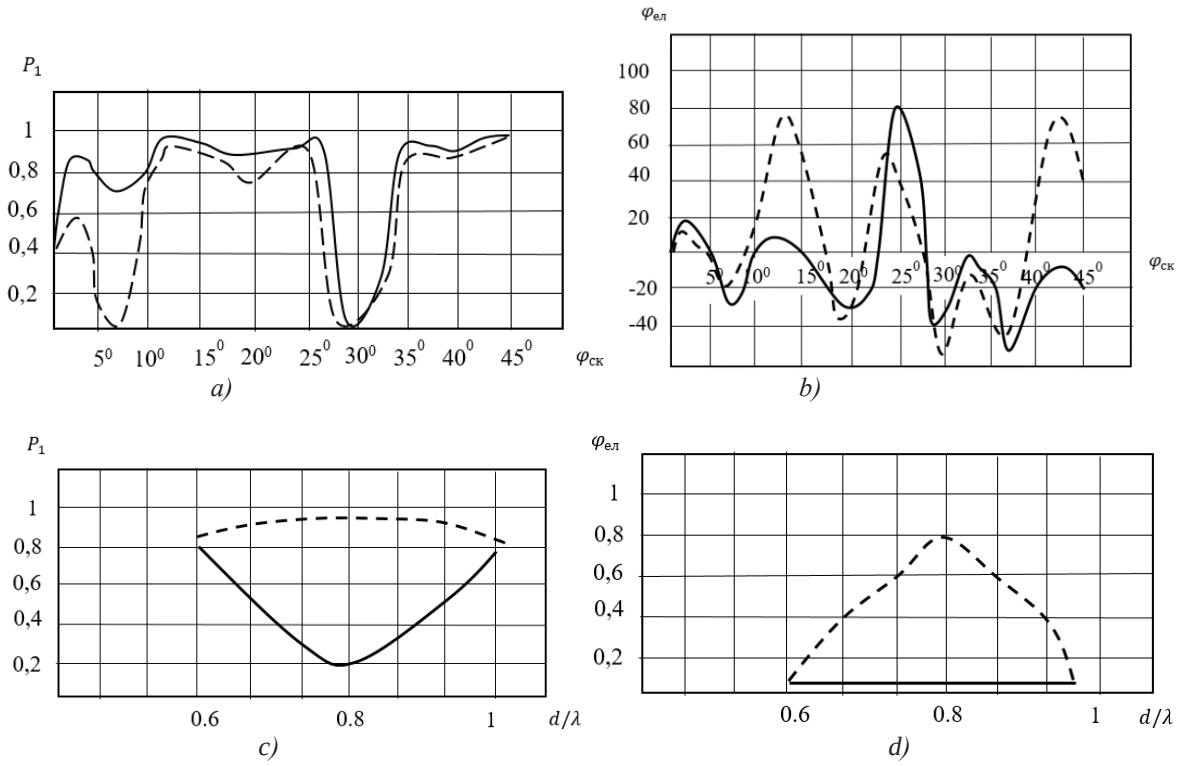


Рис. 3. Залежність коефіцієнта поляризації та кута орієнтації еліпса поляризації від кутів сканування й геометричних параметрів ФАР-64

Розглянемо залежність поляризаційних параметрів від кутів сканування, параметрів амплітудного розподілу й геометричних параметрів для ФАР-64.

При малих кутах сканування ($0^0 \le \varphi_{ск} \le 15^0$) і при $d/\lambda = 0{,}6$ функція $P = f(\varphi_{ск})$ змінюється в межах 0,25 … 0,95, при середніх кутах сканування ($15^0 \le \varphi_{ск} \le 30^0$) – у межах 0,95 … 0, при великих кутах сканування ($30^0 \le \varphi_{ск} \le 45^0$) – у межах 0 … 0,95.

При середніх і великих кутах сканування ($15^0 \le \varphi_{ск} \le 30^0$) коефіцієнт поляризації *P* майже не залежить від відстані *d* між випромінюючими елементами.

Залежність кута нахилу великої осі еліпса поляризації від кута сканування також має осцилюючий характер. Для тих самих геометричних параметрів $d/\lambda = 0{,}6$ кут $\varphi_{ел}$ змінюється в межах $-50^0 \ldots +80^0$.

При малих кутах сканування коефіцієнт поляризації *P* суттєво залежить від величини d/λ. При великих кутах сканування ($30^0 \le \varphi_{ск} \le 45^0$) коефіцієнт поляризації *P* майже не залежить від відстані *d* між випромінюючими елементами.

Висновок

Отримані чисельні результати, що дають змогу розрахувати поляризаційні параметри ФАР, а також простежити динаміку зміни поляризаційних параметрів як при різних скануванні ДС, так і при різних параметрах амплітудного розподілу.

Показано, що амплітудний розподіл впливає на поляризаційні параметри залежно від кількості елементів ФАР практично однаково: чим більша відмінність в амплітудному розподілі правополяризаційних і лівополяризаційних елементів, тим більше коефіцієнт поляризації прямує до кругової поляризації того знака, амплітудний розподіл якого ближче до рівномірного.

Коефіцієнт поляризації *P* практично незалежно від кількості випромінюючих елементів при малих кутах сканування ($0^0 \le \varphi_{ск} \le 15^0$) має залежність від відстані *d* між випромінюючими елементами, а при середніх ($15^0 \le \varphi_{ск} \le 30^0$) і великих ($30^0 \le \varphi_{ск} \le 45^0$) кутах сканування – практично не залежить.

Залежність кута нахилу головної вісі еліпса поляризації від кута сканування – осцилююча, глибина осциляції залежить від виду амплітудного розподілу, кількості елементів ФАР і відстані між елементами. Характер залежності кута нахилу великої вісі еліпса поляризації від відстані між елементами при N=16 та N=64 однаковий. При малих кутах сканування ($\varphi_{ск} \leq 5^0$) кут орієнтації еліпсу майже не залежить від відстані між елементами, тоді як при середніх і великих кутах сканування ($\varphi_{ск} \leq 10^0$) значно змінюється.

Перспективним є ФАР у мобільних комбінованих цифрових тропосферно-радіорелейних станціях [36; 37], мобільних комбінованих цифрових тропосферно-іоносферних станціях [38] і мобільних вузлових цифрових тропосферних станціях [39].

Література

1. DART-T. Dual-Mode All-Band Relocatable-Communications Transport Terminal. HC-BLOS Family of Products. URL: https://vovklict.nl/intercom/2010/4/47.pdf.
2. Патент України на винахід № 126206 Україна: С2. заявл. 10.03.2020; опубл. 31.08.2022. Бюл. № 35. Мобільна вузлова цифрова тропосферна станція.
3. Патент України на винахід № 126203 Україна: С2. заявл. 17.01.2020; опубл. 31.08.2022. Бюл. № 35. Мобільна цифрова тропосферно-космічна станція.
4. Патент України на винахід № 122168 Україна: С2. заявл. 01.08.2018; опубл. 25.09.2020. Бюл. № 18. Мобільна високошвидкісна цифрова тропосферна станція.
5. Bakan T., Yilmaz B. A., Çetintepe Ç., Alatan L. A Dual-Polarized Low-Profile Wideband Antenna Array with Wide-Scan Ability. 2022 IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P.1-4. DOI: 10.1109/PAST49659.2022.9975098.
6. Su Y., Soares I. V., Balon S., Cao J., Nikolayev D., Skrivervik A. K. Compact Irregular Conformal Phased Array: High-Gain Wide-Scan Onboard System for Enhancing UAV Communication. *IEEE Transactions on Vehicular Technology.* 2024. Vol. 73. № 12. P. 18005–18016. DOI: 10.1109/TVT.2024.3445456.
7. Aljuhani A. H., Kanar T., Zihir S., Rebeiz G. M. A 256-Element Ku-Band Polarization Agile SATCOM Receive Phased Array With Wide-Angle Scanning and High Polarization Purity. *IEEE Transactions on Microwave Theory and Techniques.* 2021. Vol. 69, № 5. P. 2609–2628. DOI: 10.1109/TMTT.2021.3056439.
8. Bu D., Qu S.-W. Millimeter-Wave Endfire Circularly Polarized Phased Array With Wideband Wide-Angle Scanning. IEEE Transactions on Antennas and Propagation, 2024. Vol.72, No.10, P.7644-7650. DOI: 10.1109/TAP.2024.3451158.
9. Fang S.-G., Qu S.-W., Yang S., Li X.-Q., Sun H.-B., Zhou Z.-P. Low Scattering Patch Array Antenna Based on Grooved Ground. *IEEE Antennas and Wireless Propagation Letters.* 2021. Vol. 20, № 3. P. 308–312. DOI: 10.1109/LAWP.2020.3048779.
10. Omam Z. R., Abdel-Wahab W. M., Gigoyan S., Safavi-Naeini S. High Gain 4×4 SIW Passive Phased Array Antenna. IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 2020. P. 45–46. DOI: 10.1109/IEEECONF35879.2020.9329752.
11. Li A., Qu S.-W., Yang S. Conformal Array Antenna for Applications in Wide-Scanning Phased Array Antenna Systems. *IEEE Antennas and Wireless Propagation Letters.* 2022. Vol. 21, № 9. P. 1762–1766. DOI: 10.1109/LAWP.2022.3179404.
12. Ghaffarian N. et al. Characterization and Calibration Challenges of an K-Band Large-Scale Active Phased-Array Antenna with a Modular Architecture. 50th European Microwave Conference (EuMC), Utrecht, Netherlands, 2021, P. 1039–1042. DOI: 10.23919/EuMC48046.2021.9338111.
13. Ciociola A., Infante L., Ricciardella N., Solimene R., Felaco M., Pellegrini G. Digitally Synthesized Antenna Test Bench for Next Generation Phased Array Systems. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022, P. 1–6. DOI: 10.1109/PAST49659.2022.9975028.
14. Cash I., Tong K.-F., Al-Armaghany A. Passive Phase Multiplication as a Means Towards Low-Cost, High-Performance Phased Array Antennas. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024. P. 1–6. DOI: 10.1109/ARRAY58370.2024.10880421.
15. Sano M., Higaki M., Wada K., Hashimoto K. A Patch Antenna Array With a Rotatable Polarization Plane for Ku-Band Phased Arrays. International Symposium on Antennas and Propagation (ISAP), Osaka, Japan, 2021. P. 7–8. DOI: 10.23919/ISAP47053.2021.9391459.
16. Kim J.-W., Chae S.-C., Jo H.-W., Yeo T.-D., Yu J.-W. Wideband Circularly Polarized Phased Array Antenna System for Wide Axial Ratio Scanning. *IEEE Transactions on Antennas and Propagation.* 2022. Vol. 70, № 2. P. 1523–1528. DOI: 10.1109/TAP.2021.3078509.
17. Chikkodi R., Vangol P., A. M., Kedar A. Wide-Scan and Wideband Antenna Array Design for Phased Array applications in X-band. IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, 2021. P. 197–201. DOI: 10.1109/MysuruCon52639.2021.9641092.
18. Xue X., Shi H., Jiang T., Han Y., Zhi G. Design of a Low-Profile Polarization Tracking Active Phased Array Antenna for Satcom On the Move. IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2022. P. 1765–1769. DOI: 10.1109/ITOEC53115.2022.9734657.
19. Peng J.-J., Qu S.-W., Xia M., Yang S. Wide-Scanning Conformal Phased Array Antenna for UAV Radar Based on Polyimide Film. *IEEE Antennas and Wireless Propagation Letters.* 2020. Vol. 19, № 9. P. 1581–1585. DOI: 10.1109/LAWP.2020.3011412.
20. Omam Z. R., Abdel-Wahab W. M., Ghafarian N., Gigoyan S., Safavi-Naeini S. Two-way Passive Phased Array Antenna for Simultaneous Transmit and Receive Signals. IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI), Singapore, Singapore, 2021. P. 1401–1402. DOI: 10.1109/APS/URSI47566.2021.9704568.
21. Wang D., Zhang X., Xia X., Feng C. An Ultra Wideband Wide-Angle Scanning Phased Array Antenna. IEEE 12th Asia-Pacific Conference on Antennas and Propagation (APCAP), Nanjing, China, 2024. P. 1–2. DOI: 10.1109/APCAP62011.2024.10881986.
22. Kedar A. Indigenous State-of-the-Art Development of Wide Band and Wide Scan Active Phased Array Antennas. IEEE Microwaves, Antennas, and Propagation Conference (MAPCON), Bangalore, India, 2022. P. 1572–1577. DOI: 10.1109/MAPCON56011.2022.10046969.
23. Gupta N., Jha B. M. Wideband Phased Array Radar Antenna Using Low Profile Cavity-Backed Slots. IEEE Microwaves, Antennas, and Propagation Conference (MAPCON), Ahmedabad, India, 2023. P. 1–4. DOI: 10.1109/MAPCON58678.2023.10464007.
24. Abdel-Wahab W. M., Palizban A., Ehsandar A., Safavi-Naeini S. SIW-Feed for 20 GHz Active Phased Array (1024-Element) Antenna with Modular Architecture. IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 2020. P. 629–630. DOI: 10.1109/IEEECONF35879.2020.9329789.

25. Waters J., Pour M. Dual-Mode Antenna Element with Application in a Linear Monopulse Phased Array. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024. P. 1-3. DOI: 10.1109/ARRAY58370.2024.10880334.
26. Kataria C. Y., Cantley L. E., Fenn A. J., Terranova D. F., Yegnanarayanan S. S., Zhang B. Design of a Flexible Thermally-Drawn Photonics Receive Linear Dipole Phased Array Antenna at UHF. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P. 1–8. DOI: 10.1109/PAST49659.2022.9974961.
27. Zhu Y., Y Hao., Deng C. Single-Layer Dual-Polarized End-Fire Phased Array Antenna for Low-Cost Millimeter-Wave Mobile Terminal Applications. IEEE Conference on Antenna Measurements and Applications (CAMA), Guangzhou, China, 2022. P. 1–3. DOI: 10.1109/CAMA56352.2022.10002551.
28. Liu M., Li W. An X-band Dual-polarized Microstrip Phased Array Antenna with Wide-angle Scanning. 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Zhuhai, China, 2021. P. 1–3. DOI: 10.1109/ISAPE54070.2021.9753153.
29. Jihao L. Antenna on board package for 5G millimeter wave phased array antenna. 21st International Conference on Electronic Packaging Technology (ICEPT), Guangzhou, China, 2020. P. 1–5. DOI: 10.1109/ICEPT50128.2020.9202652.
30. Deng Y., Li B., Zhang J., Sun L., Zhou Z. A Compact W-band Tapered Slot Antenna on Silicon Substrate for Active Phased Array Antenna. International Conference on Microwave and Millimeter Wave Technology (ICMMT), Shanghai, China, 2020. P. 1–3. DOI: 10.1109/ICMMT49418.2020.9386634.
31. Tadayon H., Ardakani M. D., Karimian R., Ahmadi S., M Zaghloul. A Wideband Non-Reciprocal Phased Array Antenna with Side Lobe Level Suppression. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P. 1–4. DOI: 10.1109/PAST49659.2022.9974969.
32. Banerjee R., Sharma S. K., Nguyen P., Chieh J.-C. S., Olsen R. Investigations of All Metal Heat Sink Dual Linear Polarized Phased Array Antenna for Ku-Band Applications. International Applied Computational Electromagnetics Society Symposium (ACES), Monterey, CA, USA, 2020. P. 1–2. DOI: 10.23919/ACES49320.2020.9196052.
33. Yang G., Zhang Y., Zhang S. Wide-Band and Wide-Angle Scanning Phased Array Antenna for Mobile Communication System. *IEEE Open Journal of Antennas and Propagation.* 2021. Vol. 2. P. 203–212. DOI: 10.1109/OJAP.2021.3057062.
34. Schvartzman D., Palmer R. D., Yu T.-Y., Nai F. Phase-Only Antenna Pattern Synthesis for Polarimetric Phased Array Radar. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024. P. 1–6. DOI: 10.1109/ARRAY58370.2024.10880311.
35. Почерняєв В. М., Магомедова М. С., Сивкова Н. М. Плоска фазована антенна решітка для мобільних цифрових станцій зв'язку «точка-багатоточка» НВЧ діапазону. *Системи управління, навігації та зв'язку.* 2024. Т. 2. № 76. С. 187–190. DOI: 10.26906/SUNZ.2024.2.187.
36. Патент України на винахід № 112217 Україна: С2. заявл. 12.09.2014; опубл. 10.08.2016. Бюл. №15 Мобільна цифрова тропосферно-радіорелейна станція.
37. Патент України на винахід № 120288 Україна: С2. заявл. 29.08.2017; опубл. 11.11.2019. Бюл. № 21. Мобільна цифрова тропосферно-радіорелейна станція.
38. Патент України на винахід № 127524 Україна: С2. публ. 20.09.2023. Бюл № 38. Мобільна цифрова тропосферно-іоносферна станція.
39. Почерняєв В. М., Сивкова Н. М., Магомедова М. С. Мобільна вузлова цифрова тропосферна станція. *Системи озброєння і військова техніка.* 2024. № 4(76). С. 6–15. https://journal-hnups.com.ua/index.php/soivt/article/view/1542/1408.

References

1. DART-T. Dual-Mode All-Band Relocatable-Communications Transport Terminal. HC-BLOS Family of Products. URL: https://vovklict.nl/intercom/2010/4/47.pdf.
2. Patent Ukrainy na vynakhid № 126206 Ukraina: C2. zaiavl. 10.03.2020; opubl. 31.08.2022, Biul. № 35. Mobilna vuzlova tsyfrova troposferna stantsiia.
3. Patent Ukrainy na vynakhid № 126203 Ukraina: C2. zaiavl. 17.01.2020; opubl. 31.08.2022, Biul. № 35. Mobilna tsyfrova troposferno-kosmichna stantsiia.
4. Patent Ukrainy na vynakhid № 122168 Ukraina: C2. zaiavl. 01.08.2018; opubl. 25.09.2020, Biul. № 18. Mobilna vysokoshvydkisna tsyfrova troposferna stantsiia.
5. Bakan T., Yilmaz B. A., Çetintepe Ç., Alatan L. A Dual-Polarized Low-Profile Wideband Antenna Array with Wide-Scan Ability. 2022 IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P. 1–4. DOI: 10.1109/PAST49659.2022.9975098.
6. Su Y., Soares I. V., Balon S., Cao J., Nikolayev D., Skrivervik A. K. Compact Irregular Conformal Phased Array: High-Gain Wide-Scan Onboard System for Enhancing UAV Communication. IEEE Transactions on Vehicular Technology, 2024. Vol. 73, № 12, P. 18005–18016. DOI: 10.1109/TVT.2024.3445456.
7. Aljuhani A. H., Kanar T., Zihir S., Rebeiz G. M. A 256-Element Ku-Band Polarization Agile SATCOM Receive Phased Array With Wide-Angle Scanning and High Polarization Purity. IEEE Transactions on Microwave Theory and Techniques, 2021. Vol. 69, № 5, P. 2609–2628. DOI: 10.1109/TMTT.2021.3056439.
8. Bu D., Qu S.-W. Millimeter-Wave Endfire Circularly Polarized Phased Array With Wideband Wide-Angle Scanning. IEEE Transactions on Antennas and Propagation, 2024. Vol. 72, № 10, P. 7644–7650. DOI: 10.1109/TAP.2024.3451158.
9. Fang S.-G., Qu S.-W., Yang S., Li X.-Q., Sun H.-B., Zhou Z.-P. Low Scattering Patch Array Antenna Based on Grooved Ground. IEEE Antennas and Wireless Propagation Letters, 2021. Vol. 20, № 3, P. 308–312. DOI: 10.1109/LAWP.2020.3048779.
10. Omam Z. R., Abdel-Wahab W. M., Gigoyan S., Safavi-Naeini S. High Gain 4×4 SIW Passive Phased Array Antenna. IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 2020, P. 45–46. DOI: 10.1109/IEEECONF35879.2020.9329752.
11. Li A., Qu S.-W., Yang S. Conformal Array Antenna for Applications in Wide-Scanning Phased Array Antenna Systems. IEEE Antennas and Wireless Propagation Letters, 2022. Vol. 21, № 9, P. 1762–1766. DOI: 10.1109/LAWP.2022.3179404.
12. Ghaffarian N. et al. Characterization and Calibration Challenges of an K-Band Large-Scale Active Phased-Array Antenna with a Modular Architecture. 50th European Microwave Conference (EuMC), Utrecht, Netherlands, 2021, P. 1039–1042. DOI: 10.23919/EuMC48046.2021.9338111.
13. Ciociola A., Infante L., Ricciardella N., Solimene R., Felaco M., Pellegrini G. Digitally Synthesized Antenna Test Bench for Next Generation Phased Array Systems. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022, P. 1–6. DOI: 10.1109/PAST49659.2022.9975028.

14. Cash I., Tong K.-F., Al-Armaghany A. Passive Phase Multiplication as a Means Towards Low-Cost, High-Performance Phased Array Antennas. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024, P. 1–6. DOI: 10.1109/ARRAY58370.2024.10880421.

15. Sano M., Higaki M., Wada K., Hashimoto K. A Patch Antenna Array With a Rotatable Polarization Plane for Ku-Band Phased Arrays. International Symposium on Antennas and Propagation (ISAP), Osaka, Japan, 2021, P. 7–8. DOI: 10.23919/ISAP47053.2021.9391459.

16.Kim J.-W., Chae S.-C., Jo H.-W., Yeo T.-D., Yu J.-W. Wideband Circularly Polarized Phased Array Antenna System for Wide Axial Ratio Scanning. IEEE Transactions on Antennas and Propagation, 2022. Vol. 70, № 2, P. 1523–1528. DOI: 10.1109/TAP.2021.3078509.

17. Chikkodi R., Vangol P., A M., Kedar A. Wide-Scan and Wideband Antenna Array Design for Phased Array applications in X-band. IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, 2021. P. 197–201. DOI: 10.1109/MysuruCon52639.2021.9641092.

18. Xue X., Shi H., Jiang T., Han Y., Zhi G. Design of a Low-Profile Polarization Tracking Active Phased Array Antenna for Satcom On the Move. IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2022. P. 1765–1769. DOI:10.1109/ITOEC53115.2022.9734657.

19. Peng J.-J., Qu S.-W., Xia M., Yang S. Wide-Scanning Conformal Phased Array Antenna for UAV Radar Based on Polyimide Film. IEEE Antennas and Wireless Propagation Letters, 2020. Vol. 19, № 9, P. 1581–1585. DOI: 10.1109/LAWP.2020.3011412.

20. Omam Z. R., Abdel-Wahab W. M., Ghafarian N., Gigoyan S., Safavi-Naeini S. Two-way Passive Phased Array Antenna for Simultaneous Transmit and Receive Signals. IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI), Singapore, Singapore, 2021. P. 1401–1402. DOI: 10.1109/APS/URSI47566.2021.9704568.

21. Wang D., Zhang X., Xia X., Feng C. An Ultra Wideband Wide-Angle Scanning Phased Array Antenna. IEEE 12th Asia-Pacific Conference on Antennas and Propagation (APCAP), Nanjing, China, 2024. P. 1–2. DOI: 10.1109/APCAP62011.2024.10881986.

22. Kedar A. Indigenous State-of-the-Art Development of Wide Band and Wide Scan Active Phased Array Antennas. IEEE Microwaves, Antennas, and Propagation Conference (MAPCON), Bangalore, India, 2022. P. 1572–1577. DOI: 10.1109/MAPCON56011.2022.10046969.

23. Gupta N., Jha B. M. Wideband Phased Array Radar Antenna Using Low Profile Cavity-Backed Slots. IEEE Microwaves, Antennas, and Propagation Conference (MAPCON), Ahmedabad, India, 2023. P. 1–4. DOI: 10.1109/MAPCON58678.2023.10464007.

24. Abdel-Wahab W. M., Palizban A., Ehsandar A., Safavi-Naeini S. SIW-Feed for 20 GHz Active Phased Array (1024-Element) Antenna with Modular Architecture. IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 2020. P. 629–630. DOI:10.1109/IEEECONF35879.2020.9329789.

25. Waters J., Pour M. Dual-Mode Antenna Element with Application in a Linear Monopulse Phased Array. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024. P. 1-3. DOI: 10.1109/ARRAY58370.2024.10880334.

26. Kataria C. Y., Cantley L. E., Fenn A. J., Terranova D. F., Yegnanarayanan S. S., Zhang B. Design of a Flexible Thermally-Drawn Photonics Receive Linear Dipole Phased Array Antenna at UHF. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P. 1–8. DOI: 10.1109/PAST49659.2022.9974961.

27. Zhu Y., Y Hao., Deng C. Single-Layer Dual-Polarized End-Fire Phased Array Antenna for Low-Cost Millimeter-Wave Mobile Terminal Applications. IEEE Conference on Antenna Measurements and Applications (CAMA), Guangzhou, China, 2022. P. 1–3. DOI: 10.1109/CAMA56352.2022.10002551.

28. Liu M., Li W. An X-band Dual-polarized Microstrip Phased Array Antenna with Wide-angle Scanning. 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Zhuhai, China, 2021. P. 1–3. DOI: 10.1109/ISAPE54070.2021.9753153.

29. Jihao L. Antenna on board package for 5G millimeter wave phased array antenna. 21st International Conference on Electronic Packaging Technology (ICEPT), Guangzhou, China, 2020. P. 1–5. DOI: 10.1109/ICEPT50128.2020.9202652.

30. Deng Y., Li B., Zhang J., Sun L., Zhou Z. A Compact W-band Tapered Slot Antenna on Silicon Substrate for Active Phased Array Antenna. International Conference on Microwave and Millimeter Wave Technology (ICMMT), Shanghai, China, 2020. P. 1–3. DOI: 10.1109/ICMMT49418.2020.9386634.

31. Tadayon H., Ardakani M.D., Karimian R., Ahmadi S., M Zaghloul. A Wideband Non-Reciprocal Phased Array Antenna with Side Lobe Level Suppression. IEEE International Symposium on Phased Array Systems & Technology (PAST), Waltham, MA, USA, 2022. P. 1–4. DOI: 10.1109/PAST49659.2022.9974969.

32. Banerjee R., Sharma S. K., Nguyen P., Chieh J.-C. S., Olsen R. Investigations of All Metal Heat Sink Dual Linear Polarized Phased Array Antenna for Ku-Band Applications. International Applied Computational Electromagnetics Society Symposium (ACES), Monterey, CA, USA, 2020. P. 1–2. DOI: 10.23919/ACES49320.2020.9196052.

33. Yang G., Zhang Y., Zhang S. Wide-Band and Wide-Angle Scanning Phased Array Antenna for Mobile Communication System. IEEE Open Journal of Antennas and Propagation, 2021. Vol. 2, P. 203–212. DOI: 10.1109/OJAP.2021.3057062.

34. Schvartzman D., Palmer R. D., Yu T.-Y., Nai F. Phase-Only Antenna Pattern Synthesis for Polarimetric Phased Array Radar. IEEE International Symposium on Phased Array Systems and Technology (ARRAY), Boston, MA, USA, 2024. P. 1–6. DOI: 10.1109/ARRAY58370.2024.10880311.

35. Pocherniaiev V. M., Mahomedova M. S., Syvkova N. M. Ploska fazovana antenna reshitka dlia mobilnykh tsyfrovykh stantsii zviazku «tochka-bahatotochka» NVCh diapazonu. Systemy upravlinnia, navihatsii ta zviazku, 2024. T. 2, № 76, S. 187-190. DOI: 10.26906/SUNZ.2024.2.187.

36. Patent Ukrainy na vynakhid № 112217 Ukraina: C2. zaiavl. 12.09.2014; opubl. 10.08.2016, Biul. № 15. Mobilna tsyfrova troposferno-radioreleina stantsiia.

37. Patent Ukrainy na vynakhid № 120288 Ukraina: C2. zaiavl. 29.08.2017; opubl. 11.11.2019, Biul. № 21. Mobilna tsyfrova troposferno-radioreleina stantsiia.

38.Patent Ukrainy na vynakhid № 127524 Ukraina: S2. publ. 20.09.2023 Biul № 38. Mobilna tsyfrova troposferno-ionosferna stantsiia.

39. Pocherniaiev V. M., Syvkova N. M., Mahomedova M. S. Mobilna vuzlova tsyfrova troposferna stantsiia. Systemy ozbroiennia i viiskova tekhnika. 2024, № 4(76). C. 6–15. https://journal-hnups.com.ua/index.php/soivt/article/view/1542/1408.

Дата першого надходження рукопису до видання: 10.05.2025
Дата прийнятого до друку рукопису після рецензування: 12.06.2025
Дата публікації: 25.07.2025

УДК 004.8:628.1/.3:502.55
DOI https://doi.org/10.36994/2788-5518-2025-01-09-06

АНАЛІЗ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИКОРИСТАННЯ У СФЕРІ ОЦІНЮВАННЯ ЯКОСТІ ВОДИ

***Сиса В. О.**, аспірант, Національний університет «Львівська політехніка», Львів, Україна. https://orcid.org/0009-0007-6782-5671, vladyslav.o.sysa@lpnu.ua*

***Яковенко Є. І.,** к.т.н., доц., Національний університет «Львівська політехніка», Львів, Україна. https://orcid.org/0000-0001-9065-5649, yevheniia.i.yakovenko@lpnu.ua*

***Анотація.** Якість води безпосередньо впливає на життя людини. Розвиток інструментів штучного інтелекту розширює можливість для ефективного моніторингу за змінами її параметрів та автоматизування процесу виявлення непридатної для споживання води. Застосування штучного інтелекту у сфері водних ресурсів має вагоме значення для безпеки водопостачання та прогнозування стану води. Алгоритми штучного інтелекту можуть швидко обробляти великі масиви даних, виявляти закономірності, які неможливо або важко помітити традиційними методами.*

У роботі проаналізовано наукові джерела, спрямовані на дослідження можливостей використання штучного інтелекту у сфері водних ресурсів. Досліджено моделі машинного навчання та нейронних мереж, які найчастіше зустрічаються в експериментах з дослідженням якості води, а саме: моделі Random Forest, SVM, k-NN та LSTM. Для розуміння кожної моделі проаналізовано літературу й описано основні їх особливості. Особливо важливим є порівняння ефективності цих моделей і точності, яку вони показують. Також розглянуто приклади практичного застосування нейронних мереж у реальних системах моніторингу якості води.

Метод «випадковий ліс» є ансамблевим методом, що базується на побудові множини дерев рішень і демонструє високу ефективність у малих вибірках. SVM будує оптимальну гіперплощину для розділення класів, використовуючи опорні вектори та ядрові функції. Алгоритм k-NN класифікує нові точки за найближчими сусідами, що базується на обраній матриці відстані. Модель LSTM здатна ефективно працювати з послідовними даними, зберігати довгострокові залежності завдяки механізму пам'яті й воротам контролю потоку інформації.

Проведено оцінювання експериментів стосовно ефективності моделей, описаних у наукових працях. У результаті визначено найефективнішу модель з погляду точності й швидкості надання інформації про стан води, яка має назву «випадковий ліс». Згідно з експериментами ця модель показує точність визначення якості води більше ніж 90%. Відповідно, її можна використовувати для подальшого впровадження моніторингу якості води з використання пристроїв інтернету речей.

***Ключові слова:** якість води; штучний інтелект; інтернет речей; модель «випадковий ліс»; LSTM.*

ANALYSIS OF ARTIFICIAL INTELLIGENCE MODELS FOR USE IN THE FIELD OF WATER QUALITY ASSESSMENT

***Vladyslav Sysa**, Ph.D. student, Lviv Polytechnic National University, Lviv, Ukraine. https://orcid.org/0009-0007-6782-5671, vladyslav.o.sysa@lpnu.ua*

***Eugenia Yakovenko**, Ph.D. in Technical Sciences, Associate Professor, Lviv Polytechnic National University, Lviv, Ukraine. https://orcid.org/0000-0001-9065-5649, yevheniia.i.yakovenko@lpnu.ua*

***Abstract.** Water quality has a direct impact on human life. The development of artificial intelligence tools expands the possibilities for effective monitoring of changes in water parameters and automating the process of detecting water that is unsuitable for consumption. The application of artificial intelligence in the field of water resources is of great importance for the safety of water supply and the prediction of water conditions. Artificial intelligence algorithms can quickly process large volumes of data and detect patterns that are difficult or impossible to identify using traditional methods.*

This paper analyzes scientific sources focused on exploring the potential of artificial intelligence in the field of water resources. Machine learning and neural network models most frequently used quality research experiments are examined, namely Random Forest, SVM, k-NN and LSTM. To better understand each model, a literature review was conducted and their main features were described. Particularly important is the comparison of the effectiveness of these models and the accuracy they demonstrate. Practical examples of the application of neural networks in real water quality monitoring systems are also considered.

The Random Forest method is an ensemble approach based on constructing multiple decision trees and demonstrates high efficiency with small datasets. SVM builds an optimal hyperplane to separate classes using support vector and kernel functions. The k-NN algorithm classifies new data points based on the nearest neighbors using a selected distance metric. The LSTM model is capable of effectively working with sequential data, maintaining long-term dependencies thanks to its memory cell and gated information control mechanism.

An evaluation of experiments regarding the effectiveness of these models described in scientific works was conducted. As a result, the most efficient model in terms of accuracy and speed of providing information about water quality was identified as the Random Forest model. According to experiments, this model achieves water quality implementation in water quality monitoring systems using Internet of Things devices.

Key words: *water quality; artificial intelligence; Internet of Things; Random Forest model; LSTM.*

Вступ

Загальновідомо, що вода є одним із джерел життя для організмів, які живуть на планеті Земля. У період стрімкої глобалізації зменшуються ресурси, у тому числі й водні. Крім того, погіршення екологічної ситуації також впливає на стан водойм по всій планеті, від глобального забруднення океану до локальних засмічень річок, які є джерелами води для людей.

Людство навчилося очищувати воду для своєї діяльності й пиття. Однак, з огляду на постійне забруднення навколишнього середовища та неощадливе використання води, варто адаптувати технології очищення під сучасний стан водних ресурсів. Одним із інструментів, який може допомогти в цьому, є контроль якості питної води.

З огляду на широке застосування нейронних мереж у багатьох сферах діяльності людини, варто інтегрувати їх у сфери водних ресурсів і водопостачання. Це поступово впроваджують в інших країнах світу [3; 8; 11; 18]. Враховуючи перспективи використання інструментів на базі нейронних мереж, доцільно проаналізувати, які моделі нейронних мереж і методи машинного навчання вже використовують науковці.

Мета роботи – проаналізувати наукові роботи стосовно використання штучного інтелекту у сфері водних ресурсів; виокремити найбільш перспективні моделі нейронних мереж і методи машинного навчання для подальшої інтеграції їх у системи аналізу якості води.

Виконання досліджень

Дослідження моделей і методів штучного інтелекту у сфері аналізу якості води

Віднедавна штучний інтелект (далі – ШІ) став доступним широкому колу людей. Нині він має вплив на полегшення багатьох процесів у різних сферах діяльності людини. Програми на основі ШІ значно вплинули на формування бізнес-процесів, а також його все частіше почали використовувати у сфері екології та управлінні водними ресурсами. Однак роль ШІ в управлінні екологічними ресурсами в бізнесі для досягнення сталого розвитку залишається обмеженою [15]. Упровадження інструментів, пов'язаних із нейронними мережами у сферу дослідження якості води, дасть змогу ефективно контролювати зміни в складниках питної води й розробляти технології для її очищення. Цей факт є позитивним як для добробуту людей, так і для комерційних підприємств, які зможуть швидко реагувати на зміни в якості води.

Сьогодні існує багато наукових праць, які описують використання штучного інтелекту для аналізу якості водних ресурсів. Його застосовують для моделювання можливих змін у природних водоймах, аналізу стічних вод, прогнозування витрат води й аналізу якісних параметрів води. Зростання використання машинного навчання та ШІ спонукає до інновацій у сфері управління якістю води. Ці досягнення спричиняють зміну методів, що використовуються для виконання аналізу, а також інтерпретації даних про якість води. Використання систем, керованих ШІ, постійно зростає в галузі управління якістю води, що включає не тільки моніторинг, а й прогнозування та реагування на такі проблеми [15].

Для повноцінного функціонування таких систем усе частіше використовують пристрої інтернету речей, які постійно передають отримані дані. Саме тут системи зі ШІ полегшують аналіз отриманих даних для науковців або інженерів. У результаті спеціалісти оперативно отримують більш точні результати, спрямовані на дослідження водних ресурсів. Відповідно, є статистика, яка засвідчує, що моніторинг із використанням ШІ є точнішим (рис. 1).

Зважаючи на користі, яку приносять інструменти ШІ у сфері дослідження водних ресурсів, логічним є аналіз цих інструментів. Це дасть змогу глибше досліджувати ефективні методи й надалі вдосконалювати їх.

Як говорив один із перших дослідників у галузі Марвін Лі Мінські: «Штучний інтелект – це наука про те, як зробити машини такими, щоб вони робили речі, які вимагали б інтелекту, якби їх робила людина». Ця цитата датована 1968 роком і згадана у його книзі «Semantic Information Processing. MIT Press». З того часу пройшло більше ніж пів століття, і люди почали використовувати ШІ у повсякденних справах. Термін «штучний інтелект» описує в статті О. Баранов: «Штучний інтелект – це певна сукупність методів, способів, засобів і технологій, насамперед комп'ютерних, що імітує (моделює) когнітивні функції, які мають критерії, характеристики та показники еквівалентні критеріям, характеристикам та показникам відповідних когнітивних функцій людини» [1].

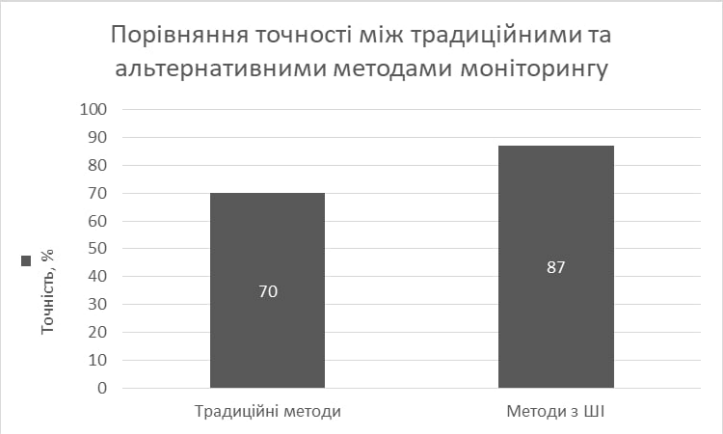


Рис. 1. Різниця між точністю традиційних методів моніторингу якості води та систем зі ШІ [15]

З опрацьованих статей стосовно дослідження водних ресурсів зустрічається безліч інструментів ШІ, як машинне навчання, так і нейронні мережі. Машинне навчання дає змогу системам автоматично вчитися на основі даних без явного програмування. Машинне навчання забезпечує ШІ здатністю знаходити закономірності, робити прогнози й покращувати свою ефективність з часом. Нейронні мережі – сукупність пов'язаних між собою елементів, які можуть «навчатися» шляхом коригування вагових коефіцієнтів під час тренування.

У сфері моніторингу якості води здебільшого застосовують методи машинного навчання Random Forest, SVM, k-NN або Decision Trees, а нейронні мережі – ANN, CNN, RNN, LSTM, Autoencoder. Кожен із цих інструментів так чи інакше використовується у сфері дослідження водних ресурсів. Оскільки нейронні мережі потребують великого обсягу даних і ресурсів для їх обробки, для моніторингу якості води доцільніше використовувати методи машинного навчання, адже вони потребують менше даних і швидко навчаються, тоді як нейронні мережі широко застосовуються для моделювання та прогнозування стану якості водних ресурсів [6].

Згідно з дослідженнями індійського вченого Апарна Понуру та його колег, доцільно застосовувати Random Forest, SVM, k-NN та LSTM. Використання того чи іншого методу зумовлене поставленим завданням і швидкістю отримання інформації. Ознайомимося детальніше з кожним із згаданих вище методів.

Random Forest – це метод ансамблевого машинного навчання, який використовується як для класифікації, так і для регресійного аналізу. Він застосовує техніку пакетування (або завантажувальну агрегацію), яка є методом генерації нового набору даних із заміною наявного набору даних. Метод машинного навчання «випадковий ліс» має особливості [13]:

– вивчення ансамблю, що використовується у «випадковому лісі», запобігає його надмірному підбору;

– пакування дає змогу «випадковому лісу» добре працювати з невеликим набором даних;

– випадкові дерева лісу можна навчати паралельно;

– автоматичний вибір функції вмикається шляхом вивчення дерева рішень у «випадковому лісі».

Support Vector Machine, або SVM – це метод машинного навчання, який використовується для класифікації та регресії. Основна ідея полягає в побудові гіперплощини, яка максимально розділяє об'єкти різних класів у просторі ознак. У загальному випадку SVM реалізує емпіричну мінімізацію ризику з додатковою регуляризацією, що допомагає уникнути перенавчання [17].

Особливості використання методу SVM [17]:

– шукає лінію, яка найкраще розділяє різні класи даних, щоб мінімізувати помилки при класифікації;

– може працювати з даними в дуже складних просторах, навіть якщо вони не лінійно роздільні, завдяки спеціальним математичним функціям (ядрам), які дають змогу «перенести» дані в інший простір, де їх можна краще розділити;

– у процесі навчання SVM використовує лише деякі точки з навчальних даних, які називаються «опорними векторами», що дає змогу моделі бути компактною й ефективною;

– завдяки механізмам регуляризації SVM може працювати, коли деякі дані містять помилки;

– чим більше даних для навчання, тим краще модель може знайти оптимальне рішення, тому цей метод добре працює, коли обсяг даних збільшується.

Алгоритм k-найближчих сусідів (k-NN) – це непараметричний класифікатор із контрольованим навчанням, який використовує близькість для класифікацій або прогнозів щодо групування окремої точки даних. Це один із популярних і найпростіших класифікаторів і класифікаторів регресії, які використовуються сьогодні в машинному навчанні [10].

Рішення про класифікацію приймається на основі відстаней між точкою, що класифікується, і точками в тренувальній вибірці. Це робить алгоритм універсальним для завдань класифікації, регресії та навіть аналізу аномалій, особливо в умовах, коли відсутнє апріорне знання про структуру даних. Ефективність алгоритму істотно залежить від вибору метрики відстані й кількості сусідів k. Вибір метрики визначає форму локального простору навколо точки, що класифікується, а параметр k – чутливість моделі до шуму [3].

Long Short-Term Memory – це тип рекурентної нейронної мережі (RNN), розроблений для ефективного моделювання послідовностей даних і збереження довготривалих залежностей у часових рядах. Уперше запропонований Сеппом Хохрайтером і Юргеном Шмідхубером у 1997 році. LSTM усуває проблеми згасаючих і вибухаючих градієнтів, які часто виникають у класичних RNN при навчанні на довгих послідовностях [9; 12].

Основна особливість LSTM – це наявність комірки пам'яті, яка дає змогу зберігати важливу інформацію протягом тривалого часу. Потік інформації в мережі регулюється за допомогою трьох воріт: Forget Gate – вирішує, яку частину інформації потрібно зберегти, а яку – відкинути; Input Gate – визначає, які нові дані потрібно додати до комірки пам'яті; Output Gate – обирає, яку частину інформації з комірки пам'яті передати далі як вихід мережі.

LSTM-мережі використовують комбінацію сигмоїдальної функції активації, яка обмежує значення від 0 до 1, а також гіперболічного тангенса, що дає змогу мати як додатні, так й від'ємні значення. Це допомагає м'яко керувати потоком інформації через мережу [12].

Ці архітектурні особливості дають LSTM можливість:

– ефективно працювати з довгими послідовностями;

– зберігати важливу інформацію, уникаючи її «забування»;

– уникати проблеми загасання або вибуху градієнтів під час навчання.

Розуміючи особливості різних методів машинного навчання, повертаємося до дослідження Апарна Понуру, який експериментальним шляхом дослідив ефективність кожного з них. Експеримент полягав у виявленні найбільш ефективного інструмента нейронних мереж для використання його в парі з IoT.

У статті [14] чітко описано, що саме опрацьовує той чи інший метод. Так, метод «випадковий ліс» краще використовувати для визначення, є вода чистою чи забрудненою, на основі вхідних характеристик, таких як pH, каламутність і концентрація розчиненого кисню. За допомогою SVM можна класифікувати зразки води як «безпечні» або «забруднені» відповідно до отриманих даних про якість води. У свою чергу, LSTM запобігає втраті довготривалих залежностей, що робить його придатним для відстеження сезонних змін трендів забруднюючих речовин протягом тривалих періодів часу [14]. K-Means групує інформацію про якість води в окремі категорії. Тому його можна застосовувати для розподілу на кластери в межах інформації про забруднення води [5].

В експерименті фіксувалися такі показники води: кислотність води (pH), каламутність, кількість розчиненого кисню, кількість розчинених солей (TDS) і кількість органічних забруднень. Дані попередньо опрацьовано для видалення відсутніх значень, нормалізації числових функцій і розділено на 70% навчальних, 15% перевірних і 15% тестових наборів [14].

Відповідно, проведено кілька експериментів, які дали змогу оцінити продуктивність моделей, автентичність результатів порівняно з уже відомими й ефективність моделей. Крім того, учені на практиці протестували моделі та визначили, як кожна з них класифікує воду на безпечну, помірно забруднену й забруднену. Крім того, вивчено використання цих моделей у режимі реального часу для виявлення середньодобової точності роботи кожної моделі. Характеристики зразків даних показано на рис. 2.

ID зразка	pH	Каламутність (NTU)	Розчинений кисень (mg/L)	Жорсткість (ppm)	Органічні забруднення (mg/L)	Характеристика якості
1	7.2	1.5	8.1	500	10	Безпечно
2	6.5	3.2	6.9	800	25	Забруднено
3	8.0	1.0	9.2	450	8	Безпечно
4	5.9	4.5	5.5	1000	40	Забруднено

Рис. 2. Характеристики зразків набору даних для експерименту [14]

Розглянемо отримані результати експериментів. Першим із них був експеримент на дослідження ефективності вибраних моделей за точністю, запам'ятовуванням та оцінкою F1 (рис. 3) [14].

Модель	Точність (%)	Прицезійність (%)	Запам'ятовування (%)	F1-оцінка (%)
Random Forest	92.5	91.8	93.1	92.4
SVM	89.7	88.5	90.2	89.3
LSTM	94.2	93.6	94.8	94.2
K-Means	85.3	84.1	86.0	85.0

Рис. 3. Результати дослідження ефективності моделей ШІ

З результатів видно, що модель LSTM має найкращі показники по всіх досліджуваних характеристиках. Найгірші результати показав k-NN. Варто зазначити, що модель «випадковий ліс» показала досить хороші результати, які є наближеними до лідера в цьому експерименті.

Далі вчені оцінюють швидкість обчислення кожної моделі, що продемонстровано на рис. 4.

Модель	Час навчання (с)	Швидкість передбачення (мс/вибірка)	Придатність для використання в реальному часі
Random Forest	120	5	Висока
SVM	180	12	Середня
LSTM	300	7	Висока
K-Means	90	4	Висока

Рис. 4. Результати дослідження обчислювальної ефективності моделей [14]

Як бачимо, найшвидшим є метод k-NN, який демонструє швидкість у навчанні й видачі результатів. LSTM потребує багато часу для навчання, це можна обґрунтувати тим, що це модель глибокого навчання. Зважаючи на показники точності з попереднього експерименту, варто виділити метод «випадковий ліс», який є точним і швидким.

Наступний крок після навчання – кожна з моделей повинна класифікувати зразки води як безпечні, помірно забруднені й забруднені (рис. 5). Цей експеримент вкотре показав, що модель LSTM найточніше класифікує опрацьовані зразки, але не набагато гірші результати продемонстровано методом «випадковий ліс». Істотні помилки зроблено моделлю k-NN.

Модель	Безпечна вода (%)	Помірно забруднена (%)	Забруднена (%)
Random Forest	68.4	20.3	11.3
SVM	66.2	21.5	12.3
LSTM	71.5	18.7	9.8
K-Means	63.1	23.2	13.7

Рис. 5. Результати експерименту з класифікації зразків води [14]

Одним із найважливіших показників є точність при роботі в реальному часі. Тому протягом 30 днів виконували моніторинг якості води. Основною характеристикою для цієї оцінки стала середньодобова точність класифікації води (рис. 6).

День	Random Forest (%)	SVM (%)	LSTM (%)	K-Means (%)
1	91.8	89.0	93.9	84.2
10	92.3	88.7	94.1	84.8
20	92.6	89.5	94.5	85.0
30	92.5	89.7	94.2	85.3

Рис. 6. Показники середньодобової точності моделей при класифікації води [14]

Як видно з рис. 6, найкращі результати показали моделі LSTM і «випадковий ліс»: їхня точність перевищує 90%. Близькою до таких значень була модель SVM, однак, з огляду на її швидкість роботи в реальному часі, вона не є кращим варіантом.

Отже, згідно з проведеними дослідженнями, можна виділити дві моделі, які показують хороші результати. Відповідно, це LSTM й Random Forest. Однак варто розуміти, яку ж з них краще широко застосувати для контролю якості води в парі з IoT. Зважаючи на простішу реалізацію, порівняно з LSTM, модель «випадковий ліс», на думку авторів, варто застосовувати для дослідження якості води, використовуючи пристрої інтернету речей. Точність і швидкість цієї моделі дасть змогу ефективно збирати й опрацьовувати інформацію для оцінювання стану води. Проте варто зазначити, що використання LSTM є доцільним при довготривалому прогнозуванні змін якості води та її моніторингу за наявності потужностей для такого спостереження.

Висновок

Аналіз наукових статей показав, що впровадження інструментів ШІ у сферу моніторингу якості води є актуальним. Оцінено ефективність сучасних моделей машинного навчання та нейронних мереж для використання їх разом з пристроями інтернету речей. У підсумку можна стверджувати, що варто використовувати модель «випадковий ліс», яка показує високу точність і швидкість. Саме на основі цієї моделі варто розвивати пристрої для моніторингу якості води, а також створювати нові гібридні моделі ШІ.

Література

1. Баранов О. А. Визначення терміну «Штучний інтелект». *Інформація і право.* 2023. № 1(44). С. 32–49. https://orcid.org/0000-0003-3233-6687.

2. Awwad A., Husseini G., Albasha L. AI-Aided Robotic Wide-Range Water Quality Monitoring System. MDPI, 2024. DOI: https://doi.org/10.3390/su16219499.

3. Daulay R. S. A., Efendi S., Suherman. Review of Literature on Improving the KNN Algorithm. *Transactions on Engineering and Computing Sciences.* 2023. № 11(3). С. 63–72. DOI: https://doi.org/10.14738/tecs.113.14768.

4. Durgun Y. Real-time water quality monitoring using AI-enabled sensors: Detection of contaminants and UV disinfection analysis in smart urban water systems. *Journal of King Saud University – Science.* 2024. № 36(9). DOI: 10.1016/j.jksus.2024.103409.

5. Elazab R., Dahab A. A., Adma M. A., Hassan H. A. Reviewing the frontier: modeling and energy management strategies for sustainable 100% renewable microgrids. *SN Applied Sciences.* 2024. № 6(4). P. 168. DOI: 10.1007/s42452-024-05820-6.

6. Gaya Muhammad Sani, Abba Sani Isah, Abdu Aliyu Muhammad, Tukur Abubakar Ibrahim, Saleh Mubarak Auwal, Esmaili Parvaneh, Wahab Norhaliza Abdul. Estimation of water quality index using artificial intelligence approaches and multi-linear regression. *IAES International Journal of Artificial Intelligence (IJ-AI).* 2020. Vol. 9, № 1. P. 126–134. DOI: 10.11591/ijai.v9.i1.pp126-134.

7. Godavarthi B., Nalajala P., Ganapuram V. Design and Implementation of Vehicle Navigation System in Urban Environments using Internet of Things (Iot). IOP Conference Series Materials Science and Engineering, 2017. DOI: 10.1088/1757-899X/225/1/012262 225(1):012262.

8. Gunda N., Gautam S., Mitra S. Artificial Intelligence Based Mobile Application for Water Quality Monitoring. ECS Meeting Abstracts, 2018. DOI: 10.1149/MA2018-02/56/1997.

9. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation.* 1997. № 9(8). P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

10. IBM Think: What is the k-nearest neighbors (KNN) algorithm? URL: https://www.ibm.com/think/topics/knn.

11. Kim Hoo Hugo, Byeongwook Choi, Zahid Ullah, Nahyeon Jeong, Kyung Hwa Cho, Sanghun Park, Sang-Soo Baek, Moon Son. Decoupling ion concentrations from effluent conductivity profiles in capacitive and battery electrode deionizations using an artificial intelligence model. Water Research, 2024. DOI: https://doi.org/10.1016/j.watres.2024.122092.

12. Okut H., Deep Learning: Long-Short Term Memory. Deep Learning Applications, 2021. DOI: 10.5772/intechopen.96180.

13. Pandey Rajiv, Sunil Kumar Khatri, Neeraj Kumar Singh, Parul Verma. Artificial Intelligence and Machine Learning for EDGE Computing. 1st Edition – April 26, 2022.

14. Ponnuru Aparna J. V., Madhuri Dr. S., Saravanan T., Vijayakumar Dr. V., Manimegalai Dr. Abhijeet Das. Data-Driven Approaches to Water Quality Monitoring: Leveraging AI, Machine Learning, and Management Strategies for Environmental Protection. *Journal of Neonatal Surgery.* 2025. № 14 (5s). P. 664–675. DOI: https://doi.org/10.52783/jns.v14.2107.

15. Rajitha Akula, Aravinda K., Nagpal Amandeep, Kalra Ravi, Maan Preeti, Ashish Kumar, Dalael Saad Abdul-Zahra. Machine Learning and AI-Driven Water Quality Monitoring and Treatment. E3S Web of Conferences 505, 03012 (2024). https://doi.org/10.1051/e3sconf/202450503012.

16. Rudjord Z., Reid M., Schwermer C., Lin Y. Laboratory Development of an AI System for the Real-Time Monitoring of Water Quality and Detection of Anomalies Arising from Chemical Contamination. MDPI, 2022. DOI:10.3390/w14162588.

17. Steinwart Ingo and Christmann Andreas. Support Vector Machines, 2008.

18. Zhu M., Wang J., Yang X., Zhang Y. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 2022. DOI: 10.1016/j.eehl.2022.06.001.

References

1. Baranov O. A. Definition of the term «Artificial Intelligence». ORCID: https://orcid.org/0000-0003-3233-6687. INFORMATION AND LAW. № 1(44)/2023 (p. 32–49).

2. Awwad A., Husseini G., Albasha L. AI-Aided Robotic Wide-Range Water Quality Monitoring System. MDPI, 2024. DOI: https://doi.org/10.3390/su16219499.

3. Daulay, R. S. A., Efendi, S., & Suherman. Review of Literature on Improving the KNN Algorithm. Transactions on Engineering and Computing Sciences, 2023. 11(3). 63–72. DOI: https://doi.org/10.14738/tecs.113.14768.

4. Durgun Y. Real-time water quality monitoring using AI-enabled sensors: Detection of contaminants and UV disinfection analysis in smart urban water systems. Journal of King Saud University – Science 36(9), 2024. DOI: 10.1016/j.jksus.2024.103409.

5. Elazab, R., Dahab, A. A., Adma, M. A., Hassan, H. A., 2024. Reviewing the frontier: modeling and energy management strategies for sustainable 100% renewable microgrids. SN Applied Sciences, 6(4), pp. 168. DOI: 10.1007/s42452-024-05820-6.

6. Gaya Muhammad Sani, Abba Sani Isah, Abdu Aliyu Muhammad, Tukur Abubakar Ibrahim, Saleh Mubarak Auwal, Esmaili Parvaneh, Wahab Norhaliza Abdul. Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 9, № 1, March 2020, pp. 126–134. DOI: 10.11591/ijai.v9.i1.pp126-134.

7. Godavarthi B., Nalajala P., Ganapuram V. Design and Implementation of Vehicle Navigation System in Urban Environments using Internet of Things (Iot). IOP Conference Series Materials Science and Engineering, 2017. DOI: 10.1088/1757-899X/225/1/012262 225(1):012262.

8. Gunda N., Gautam S., Mitra S. Artificial Intelligence Based Mobile Application for Water Quality Monitoring. ECS Meeting Abstracts, 2018. DOI: 10.1149/MA2018-02/56/1997.

9. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation. November 1997, 9(8):1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

10. IBM Think: What is the k-nearest neighbors (KNN) algorithm? URL: https://www.ibm.com/think/topics/knn.

11. Kim Hoo Hugo, Byeongwook Choi, Zahid Ullah, Nahyeon Jeong, Kyung Hwa Cho, Sanghun Park, Sang-Soo Baek, Moon Son. Decoupling ion concentrations from effluent conductivity profiles in capacitive and battery electrode deionizations using an artificial intelligence model. Water Research, 2024. DOI: https://doi.org/10.1016/j.watres.2024.122092.

12. Okut H., Deep Learning: Long-Short Term Memory. Deep Learning Applications , 2021. DOI: 10.5772/intechopen.96180.

13. Pandey Rajiv, Sunil Kumar Khatri, Neeraj Kumar Singh, Parul Verma. Artificial Intelligence and Machine Learning for EDGE Computing. 1st Edition – April 26, 2022.

14. Ponnuru Aparna, J. V. Madhuri, Dr. S. Saravanan, T. Vijayakumar, Dr. V. Manimegalai, Dr. Abhijeet Das. Data-Driven Approaches to Water Quality Monitoring: Leveraging AI, Machine Learning, and Management Strategies for Environmental Protection., 2025. Journal of Neonatal Surgery, 14 (5s), 664–675. DOI: https://doi.org/10.52783/jns.v14.2107.

15. Rajitha Akula, Aravinda K, Nagpal Amandeep, Kalra Ravi, Maan Preeti, Ashish Kumar, Dalael Saad Abdul-Zahra. Machine Learning and AI-Driven Water Quality Monitoring and Treatment. E3S Web of Conferences 505, 03012 (2024). https://doi.org/10.1051/e3sconf/202450503012.

16. Rudjord Z., Reid M., Schwermer C., Lin Y. Laboratory Development of an AI System for the Real-Time Monitoring of Water Quality and Detection of Anomalies Arising from Chemical Contamination. MDPI, 2022. DOI: 10.3390/w14162588.

17. Steinwart Ingo and Christmann Andreas. Support Vector Machines. 2008.

18. Zhu M, Wang J., Yang X., Zhang Y. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 2022. DOI: 10.1016/j.eehl.2022.06.001.

Дата першого надходження рукопису до видання: 25.05.2025
Дата прийнятого до друку рукопису після рецензування: 27.06.2025
Дата публікації: 25.07.2025

УДК 004.9
DOI https://doi.org/10.36994/2788-5518-2025-01-09-07

ПРИНЦИП ФУНКЦІОНУВАННЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ ДОБОРУ ОЛИВ

***Тягунова М. Ю.**, к.т.н., доц., Національний університет «Запорізька політехніка», Запоріжжя, Україна. https://orcid.org/0000-0002-9166-5897, mary.tyagunova@gmail.com*

***Бобирь Д. С.,** аспірант, Національний університет «Запорізька політехніка», Запоріжжя, Україна. https://orcid.org/0009-0009-6635-3582, dimabobyr@gmail.com*

***Анотація.** В умовах нестримного розвитку автомобільної галузі питання вибору мастильного матеріалу стає дедалі актуальнішим, адже сучасні силові агрегати вимагають точного дотримання технологічних норм і стандартів. Невідповідна олива може призвести до зниження ефективності двигуна, зростання витрат пального й передчасного зносу. Автоматизовані системи підбору моторної оливи дають змогу підвищити ефективність технічного обслуговування транспортних засобів, забезпечуючи точність, швидкість і персоналізацію вибору [1; 2]. Метою роботи є розроблення алгоритму, що має адаптуватися до змінних умов експлуатації, технічних параметрів двигуна й індивідуальних запитів користувача. У статті розглянуто створення алгоритму автоматизованої системи, яка враховує технічні характеристики транспортного засобу (тип двигуна, об'єм, турбонаддув, ступінь зношення), експлуатаційні умови (сезонність, стиль керування, навантаження), нормативні вимоги (API, ACEA, ILSAC, OEM-допуски), а також суб'єктивні вподобання споживача (бренд, бюджет, інтервал заміни). Алгоритм добору базується на системі експертних правил, що реалізовані у вигляді дерева прийняття рішень з багаторівневою фільтрацією. Для кожного варіанта оливи обчислюється рейтинг відповідності, який ураховує пріоритетність критеріїв. З метою підвищення продуктивності системи запропоновано оптимізаційні підходи: індексацію бази даних, кешування результатів, паралельну обробку запитів. Інтеграція машинного навчання дає змогу системі самонавчатися на основі накопиченої інформації, підвищуючи точність персоналізованих рекомендацій. Валідація охоплює аналітичні методи, порівняння з офіційними специфікаціями й експертну оцінку. Розроблена система демонструє високу точність добору, відповідність міжнародним стандартам і має значний потенціал для практичного застосування в галузі сервісного обслуговування автомобільного транспорту.*

***Ключові слова:** автоматизована система; експлуатаційні умови; база даних; моторна олива; алгоритм підбору олив.*

THE PRINCIPLE OF OPERATION OF AN AUTOMATED OIL SELECTION SYSTEM

***Mariia Tiahunova,** Ph.D, Ass. Prof., National University Zaporizhzhia Polytechnic, Zaporizhzhia, Ukraine. https://orcid.org/0000-0002-9166-5897, mary.tyagunova@gmail.com*

***Dmytro Bobyr,** PhD student, National University Zaporizhzhia Polytechnic, Zaporizhzhia, Ukraine. https://orcid.org/0009-0009-6635-3582, dimabobyr@gmail.com*

***Abstract.** In the context of rapid development in the automotive industry, the issue of selecting an appropriate lubricant is becoming increasingly relevant, as modern power units require strict compliance with technological standards and regulations. Inappropriate engine oil can lead to decreased engine efficiency, increased fuel consumption, and premature wear. Automated engine oil selection systems help improve the efficiency of vehicle maintenance by ensuring accuracy, speed, and personalization in the selection process. The aim of this work is to develop an algorithm that adapts to changing operating conditions, technical parameters of the engine, and individual user preferences. This study considers the creation of an algorithm for an automated system that takes into account the technical characteristics of the vehicle (engine type, displacement, turbocharging, degree of wear), operating conditions (seasonality, driving style, load), regulatory requirements (API, ACEA, ILSAC, OEM approvals), as well as subjective consumer preferences (brand, budget, oil change interval). The selection algorithm is based on a system of expert rules implemented in the form of a decision tree with multi-level filtering. For each oil option, a conformity rating is calculated, taking into account the prioritization of various criteria. To increase system performance, optimization approaches are proposed: database indexing, result caching, and parallel request processing. The integration of machine learning allows the system to self-learn based on accumulated data, improving the accuracy of personalized recommendations. Validation includes analytical methods, comparison with*

official specifications, and expert evaluation. The developed system demonstrates high selection accuracy, compliance with international standards, and significant potential for practical application in the field of automotive maintenance.

Key words: *automated system; operating conditions; database; engine oil; oil selection algorithm.*

Вступ

Оптимальний підбір моторної оливи є одним із ключових факторів забезпечення довготривалої, ефективної та екологічно безпечної експлуатації двигуна внутрішнього згоряння [3; 4]. У сучасних умовах розвитку інженерії нестримно зростає актуальність створення автоматизованих систем підтримки прийняття рішень у цій сфері. Ефективність таких систем визначається точністю побудови критеріїв підбору мастильних матеріалів, що мають охоплювати широкий спектр параметрів.

При побудові наших алгоритмів доцільно буде виокремити чотири ключові групи критеріїв вибору моторної оливи в інформаційно-аналітичних платформах.

Передусім необхідно розглянути технічні параметри силового агрегату.

Вони охоплюють конструкційно-функціональні характеристики двигуна, які безпосередньо впливають на вимоги до властивостей мастильного матеріалу. Розглянемо основні технічні характеристики. Тип двигуна – бензинові, дизельні, гібридні та газові установки мають різні потреби у в'язкості, рівні зольності, хімічному складі й температурній стабільності олив. Робочий об'єм двигуна визначає теплове навантаження та швидкість деградації оливи; потужніші двигуни зазвичай потребують більш розширений склад моторного пакету. Наявність турбіни вимагає застосування термостійких олив із підвищеними антиокислювальними властивостями. Стан агрегату (пробіг, знос) – двигуни з підвищеним зношенням можуть потребувати більш в'язких олив або продуктів із модифікованими присадками для підтримки компресії та зменшення витрат мастила.

По-друге, потрібно звернути увагу на експлуатаційні умови використання транспортного засобу. Щоденне використання автомобіля визначає динаміку старіння мастильного матеріалу, ступінь термічного й механічного навантаження та загальну потребу в певних типах захисту. Експлуатація автомобіля в зимовий або літній період вимагає олив із відповідним температурним діапазоном. Температурні коливання навколишнього середовища зумовлюють вибір олив із відповідним індексом в'язкості (SAE). Агресивне керування, інтенсивне гальмування та прискорення, часті поїздки в міському циклі сприяють швидкому зносу й потребують олив із високими протизношувальними характеристиками. Регулярне перевезення вантажів, короткі дистанції, тривалі простої або підвищена запиленість впливають на частоту заміни й вимоги до складу оливи.

По-третє, відповідність нормативним вимогам і стандартам. Автоматизована система має інтегрувати актуальні технічні регламенти, які забезпечують безпеку й відповідність моторної оливи конструкційним вимогам автовиробника. Розглянемо основні з них.

Міжнародні класифікації (API, ACEA, ILSAC) регламентують категорії якості, термостійкість, рівень забруднення, окислювальну стабільність тощо. OEM-допуски – вимоги, установлені виробниками автомобілів (наприклад, Mercedes-Benz MB 229.5, VW 504.00/507.00), які гарантують відповідність оливи конкретним силовим агрегатам. Екологічні стандарти (EURO-4, EURO-5, EURO-6) установлюють обмеження на хімічний склад та емісію забруднюючих речовин, що передбачає застосування Low SAPS-олив із низьким умістом сульфатної золи, фосфору й сірки. Також необхідно враховувати особисті пріоритети й уподобання кінцевих споживачів. Незважаючи на технічну другорядність, ці параметри часто мають вирішальне значення у фінальному рішенні користувача: суб'єктивне ставлення до торгової марки, маркетингові кампанії та досвід попереднього використання; фінансова доступність мастильного матеріалу визначає його придатність для широких груп споживачів; тривалість інтервалу між замінами оливи залежить від її базового складу (мінеральна, синтетична) та умов експлуатації.

Таким чином, ефективне впровадження автоматизованої системи добору моторної оливи передбачає не лише наявність детальної бази даних, а й побудову чітко визначеної логіки прийняття рішень. Запропонована система ґрунтується на підході, що базується на експертних правилах, які описують умови використання конкретних видів мастильних матеріалів залежно від численних вхідних параметрів.

Виконання досліджень

Розглянемо основні принципи алгоритмічного підходу. Кожне експертне правило реалізується у формі логічної конструкції, що дає змогу поетапно звужувати перелік можливих варіантів шляхом аналізу одного або кількох параметрів. На цій основі формується дерево прийняття рішень, яке забезпечує поступову фільтрацію даних із подальшим формуванням впорядкованого списку відповідних олив [5].

Розглянемо кожен крок такого алгоритму. Спочатку відбувається збирання вхідної інформації. На цьому етапі система отримує дані, які можуть бути внесені вручну користувачем або імпортовані автоматично. До них належать технічні характеристики транспортного засобу, умови експлуатації, стандарти автовиробника, а також побажання самого користувача. Первинна фільтрація даних ви-

користовується з метою зменшення вибірки. Система виконує відсів невідповідних варіантів олив за такими критеріями: тип двигуна (бензиновий, дизельний, гібридний), температурні умови експлуатації (параметри в'язкості за SAE), наявність допусків автовиробника (наприклад, VW, MB, BMW) [6, 7].

Наступний крок – оцінювання й розрахунок рейтингу. Для кожного залишеного у вибірці варіанта розраховується коефіцієнт відповідності, що узагальнює ступінь придатності конкретної оливи для заданих умов. Рейтинг розраховується за формулою:

$$Рейтинг = W_1 \cdot D + W_2 \cdot T + W_3 \cdot S + W_4 \cdot C,$$

де

– D – відповідність технічним характеристикам двигуна;
– T – відповідність температурному режиму;
– S – відповідність нормативам і стандартам (API, ACEA, ILSAC);
– C – урахування індивідуальних уподобань користувача;
– $W_1 ... W_4$ – вагові коефіцієнти, які відображають пріоритетність кожного критерію.

На основі розрахованих рейтингів система формує список рекомендованих мастильних матеріалів. Кожна позиція супроводжується коротким поясненням вибору й описом основних характеристик продукту.

Якщо рекомендована олива недоступна в наявності, користувачу пропонуються найближчі альтернативи, які відрізняються мінімальною різницею в рейтингу.

Побудована модель відзначається високим рівнем адаптивності. Система дає змогу динамічно змінювати вагові коефіцієнти з урахуванням типу автомобіля, стилю кермування чи індивідуальних пріоритетів користувача. Завдяки цьому забезпечується як гнучкий підхід до добору, так і висока точність персоналізованих рекомендацій. На рис. 1 представлено блок-схему послідовності роботи алгоритму.

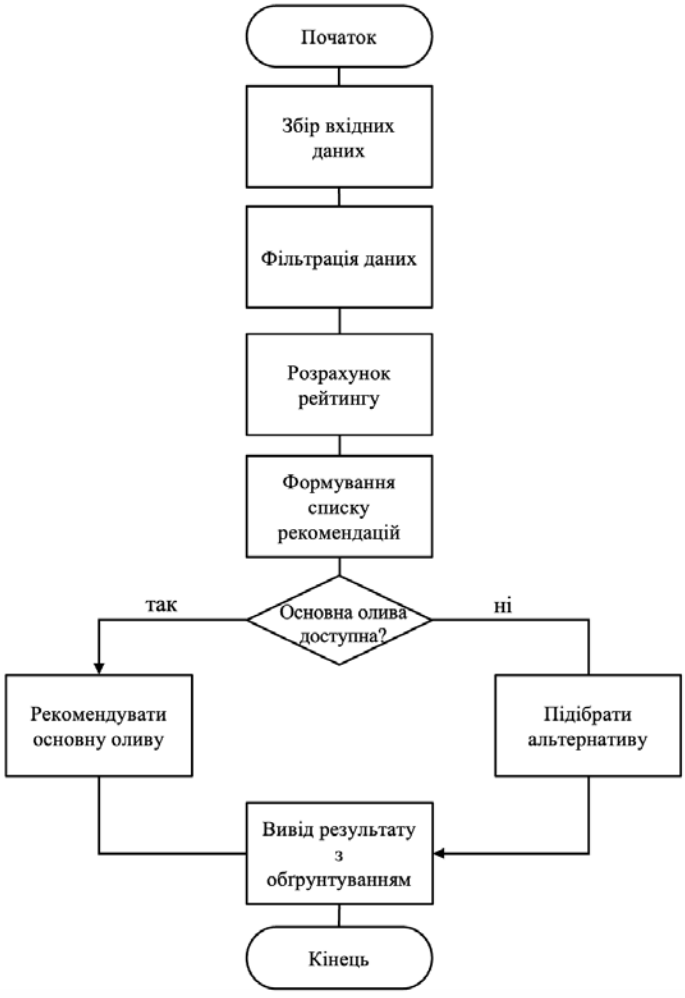


Рис. 1. Блок-схема алгоритму функціонування системи

Щоб забезпечити високу продуктивність і надійну роботу автоматизованої системи добору моторних олив, необхідно застосувати сучасні підходи до оптимізації обробки даних. Удосконалення охоплює як логіку алгоритмів, так й інфраструктурні аспекти. Одним із базових способів підвищення ефективності є індексація структур бази даних, яка значно пришвидшить вибірку інформації. Додатково, з метою зменшення навантаження на серверну частину, пропонується впровадити механізм кешування результатів запитів, що дасть змогу уникнути повторних обчислень у разі схожих параметрів.

Підтримка паралельної обробки даних дасть змогу одночасно обслуговувати велику кількість запитів, оптимізуючи використання апаратних ресурсів і знижуючи час очікування відповіді. Окрім того, зібрані історичні дані можуть бути використані як основа для навчання предиктивних моделей [7]. Застосування методів машинного навчання підвищить точність результатів і дасть змогу формувати більш індивідуалізовані рекомендації.

Адаптація системи до конкретного користувача є важливим складником сучасного сервісу. У цьому контексті в системі використовуватимуться дані про попередні запити, звички керування транспортом, історію заміни оливи й умови експлуатації. Усе це дасть можливість системі з кожним новим зверненням точніше відповідати потребам користувача. Таким чином, поєднання оптимізаційних механізмів з адаптивними можливостями формує інтелектуальну платформу, здатну до самонавчання на основі накопиченого досвіду [5].

Для гарантування коректної роботи системи необхідно провести ретельну валідацію всіх її компонентів. Перевірка точності запропонованих рекомендацій здійснюється в кілька етапів, що охоплюють як аналітичні, так і практичні методи оцінювання. Серед першочергових підходів – ручне тестування на типових прикладах, що відображають реальні сценарії експлуатації. Це дає змогу виявити критичні помилки в логіці алгоритмів. Також важливим є порівняння результатів системи з офіційними рекомендаціями виробників техніки, що забезпечує відповідність нормативним вимогам.

Додаткову цінність дає експертна оцінка спеціалістів, які можуть визначити практичну доцільність запропонованих рішень. Не менш важливим є й зворотний зв'язок від кінцевих користувачів. Аналіз таких відгуків дає змогу швидко вносити зміни в логіку добору або уточнювати параметри алгоритмів.

До основних критеріїв якості належать відповідність запропонованої оливи міжнародним стандартам (API, ACEA, ILSAC), безпечність її використання в конкретних умовах, зменшення зношення елементів двигуна та стабільність витрат палива після її застосування [6; 7].

Нарешті, частота потреби в ручному втручанні в процес підбору слугує показником ступеня автономності системи. Чим рідше виникає така необхідність, тим вища ефективність функціонування платформи.

Після впровадження запропонованої системи підбору олив отримаємо інструмент, завдяки якому значно знизиться рівень помилок при підборі оливи в техніку. Нижче наведено графік оцінювання ефективності системи.

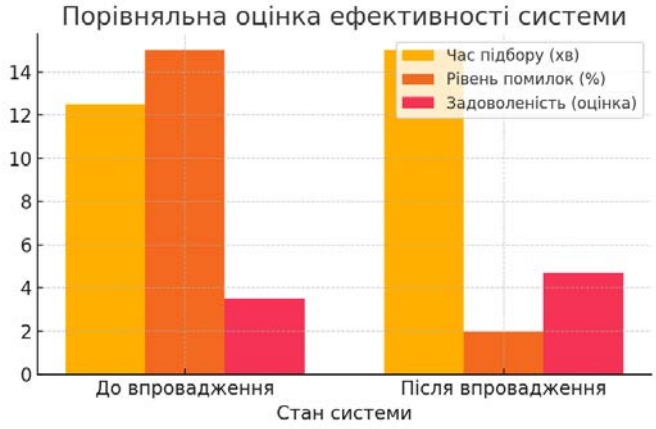


Рис. 2. Порівняльна оцінка ефективності системи

Комплексна реалізація зазначених механізмів перевірки дає змогу досягти високого рівня достовірності результатів і сформувати довіру до системи серед кінцевих користувачів.

Висновок

У результаті дослідження розроблено концепцію інтелектуального алгоритму підбору моторної оливи, який здатен адаптуватися до широкого спектру вхідних параметрів. Запропоновані методи оптимізації й валідації забезпечують високу точність та ефективність системи, що дає змогу інтегрувати її у сферу сервісного обслуговування автомобілів і сприяти підвищенню якості експлуатації техніки.

Література

1. Kamble A. R., Kamble A. D. Expert system for automotive engine oil recommendation based on driving conditions. *International Journal of Computer Applications*, 2019. Vol. 178 (9). P. 1–5. https://doi.org/10.5120/ijca2019919017.
2. Zhou Y., Xian, Y. (2020). Application of machine learning in automotive predictive maintenance: A review. *Procedia Computer Science*. 2020. Vol. 175. P. 539–546. https://doi.org/10.1016/j.procs.2020.07.077.
3. Mang T., Dresel W. (Eds.). Lubricants and lubrication (3rd ed.). Wiley-VCH, 2017.
4. Totten G. E. Lubrication and lubricant selection: A practical guide (2nd ed.). CRC Press, 2011.
5. API. Engine oil guide. *American Petroleum Institute*. 2023. URL: https://www.api.org/products-and-services/engine-oil.
6. ILSAC. ILSAC GF-6 and API SP standards. *International Lubricant Standardization and Approval Committee*. 2023. URL: https://www.ilma.org/ilsac.
7. Smith D. W., Spikes H. A. The science of engine oils – Composition, function, and performance. *Journal of ASTM International*. 2005. Vol. 2 (1). P. 1–10. https://doi.org/10.1520/JAI12534.

References

1. Kamble, A. R., & Kamble, A. D. (2019). Expert system for automotive engine oil recommendation based on driving conditions. International Journal of Computer Applications, 178 (9), 1–5. https://doi.org/10.5120/ijca2019919017.
2. Zhou, Y., & Xiang, Y. (2020). *Application of machine learning in automotive predictive maintenance: A review*. Procedia Computer Science, 175, 539–546. https://doi.org/10.1016/j.procs.2020.07.077.
3. Mang, T., & Dresel, W. (Eds.). (2017). *Lubricants and lubrication* (3rd ed.). Wiley-VCH.
4. Totten, G. E. (2011). *Lubrication and lubricant selection: A practical guide* (2nd ed.). CRC Press.
5. API. (2023). *Engine oil guide*. American Petroleum Institute. https://www.api.org/products-and-services/engine-oil.
6. ILSAC. (2023). *ILSAC GF-6 and API SP standards*. International Lubricant Standardization and Approval Committee. https://www.ilma.org/ilsac.
7. Smith, D. W., & Spikes, H. A. (2005). *The science of engine oils – Composition, function, and performance*. Journal of ASTM International, 2 (1), 1–10. https://doi.org/10.1520/JAI12534.

Дата першого надходження рукопису до видання: 12.05.2025
Дата прийнятого до друку рукопису після рецензування: 11.06.2025
Дата публікації: 25.07.2025

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

UDC 004.4`4
DOI https://doi.org/10.36994/2788-5518-2025-01-09-08

SEPARATING STRUCTURE AND DATA IN ABSTRACT SYNTAX TREES FOR EFFICIENT IMMUTABLE TRANSFORMATIONS

Bilyk V. R., *IT STEP University, Lviv, Ukraine. https://orcid.org/0009-0002-0173-0120, vladbilyk@posteo.net*

Abstract. *In compiler design, the abstract syntax tree (AST) is the primary data structure used to represent the syntactic structure of a program during translation and analysis. Traditional AST representations model structure and data as a single entity, where structural relationships are encoded implicitly via object references, and data is embedded directly within the objects. While intuitive, this approach complicates the design and implementation of AST transformations, particularly under immutability constraints. Even simple modifications often require explicit AST traversal and extensive copying, introducing accidental complexity that may obscure the essential logic of transformations and reduce performance.*

This work explores an alternative approach that explicitly separates structure from data in the AST representation. The internal form of the AST is modeled as a pair of an array of data-only nodes and an adjacency matrix that explicitly encodes the AST structure. This separation enables efficient transformation by enabling independent updates to data without modifying the structure. Structural changes can also be performed in a principled manner by modifying the adjacency matrix and the node array independently. This separation of concerns simplifies reasoning about transformations and enables concise expression without the accidental traversal logic.

To demonstrate the utility of this design, a compiler for a minimal data query language is presented. Three representative AST operations are explored: verification, data-only updates, and structural modification, illustrating how different kinds of transformations benefit from the proposed model. The proposed approach is compared to traditional object-oriented representations and existing rewriting systems. The extensibility and first-class role of the proposed design makes it applicable to a wide range of compiler tasks beyond the example presented. This model serves as a useful foundation for building compilers and transformation tools that are both robust and practical. Future directions include optimizing the internal representation and supporting alternative traversal strategies.

Key words: *compilers, abstract-syntax tree, transformation, immutability, Java.*

ВІДОКРЕМЛЕННЯ СТРУКТУРИ ВІД ДАНИХ В АБСТРАКТНИХ СИНТАКСИЧНИХ ДЕРЕВАХ ДЛЯ ЕФЕКТИВНИХ ПЕРЕТВОРЕНЬ НЕЗМІННИХ ОБ'ЄКТІВ

Владислав Білик*, ІТ СТЕП Університет, Львів, Україна. https://orcid.org/0009-0002-0173-0120, vladbilyk@posteo.net*

Анотація. *У дизайні компіляторів абстрактне синтаксичне дерево (АСД) є основною структурою даних, яка використовується для представлення синтаксичної структури програми під час трансляції й аналізу. Традиційні представлення АСД моделюють структуру й дані як єдине ціле, де структурні зв'язки неявно кодуються через посилання на об'єкти, а дані є частиною самих об'єктів. Хоча цей підхід є інтуїтивним, він ускладнює моделювання та реалізацію перетворень АСД, особливо для незмінних об'єктів. Навіть прості модифікації часто потребують явного обходу АСД та значного копіювання, що додає другорядну складність до перетворень і знижує швидкодію.*

У роботі досліджується альтернативний підхід, який явно відокремлює структуру від даних у представленні АСД. Внутрішня форма АСД моделюється як пара масиву вершин, що містять лише дані, і матриці суміжності, яка явно кодує структуру АСД. Така модель уможливлює ефективні перетворення завдяки незалежним оновленням даних без зміни структури. Структурні зміни також є можливими, виконуються незалежно, змінюючи матрицю суміжності й масив вершин. Таке відокремлення спрощує розуміння перетворень та уможливлює лаконічне вираження без другорядної складності, що стосується обходу АСД.

Аби продемонструвати застосовність наведеного підходу, представлено компілятор для мінімальної мови запитів до даних. Дослідили три типові операції над АСД: верифікацію, оновлення лише даних, структурну модифікацію, ілюструючи, як така модель сприяє різним видам перетво-

рень. Запропонований підхід порівнюється з традиційними об'єктно-орієнтованими представленнями й наявними системами рерайтингу. Розширюваність і першокласна природа запропонованої моделі роблять її застосовною до низки компіляційних процесів, що виходять за рамки представленого прикладу. Ця модель слугує основою для створення надійних, практичних компіляторів і засобів перетворення. Серед майбутніх напрямів дослідження – оптимізація внутрішнього представлення АСД та підтримка альтернативних стратегій обходу.

Ключові слова: *компілятори, абстрактне синтаксичне дерево, трансформація, незмінні об'єкти, Java.*

Introduction

Programming languages provide the ability to describe computations at a high level of abstraction. For a program to be executed, it must ultimately be interpreted or translated into a lower-level language – typically machine code that will be executed directly by a CPU. This translation is the task of a compiler, which transforms source code from one representation into another. While it is typical for compilers to target machine code, the codomain of compilation may be arbitrary. A compiler may translate one high-level language into another to connect different domains, facilitating integration with external systems or leveraging specialized execution engines (e.g., compiling a domain-specific language embedded in Java [1] into SQL for database execution).

The compilation process is centered around an abstract syntax tree (AST) – a core data structure that represents the syntactic structure of the source program. After parsing, the AST becomes the intermediate representation upon which subsequent compiler stages operate. These stages typically include verification, optimization, and code generation, each of which may involve some form of analysis or transformation of the AST.

The process of transforming an AST is often the most complex and performance-critical part of a compiler. This work focuses on improving how AST transformations are defined and performed, specifically in the context of immutable data structures.

Two key challenges arise during AST transformation:

1. **Simplicity of definition and implementation.** Ideally, the definition of a transformation should be concise, tackling the core of the problem, without having to explicitly encode the rules for AST traversal. Approaches such as *rewriting systems* [2] and *source transformation systems* [3] try to achieve this goal by separating the core transformation logic from the details of the strategy for matching parts of an AST and applying the changes. This principle is in line with Brooks's distinction between *essential* and *accidental* complexity [4].

2. **Performance under immutability.** When all data is immutable – an increasingly common choice in modern software engineering [5] – informal reasoning becomes easier at the cost of performance penalties of varying degree. In compiler design, each AST transformation must produce a new AST object. Naïve implementations that rebuild entire trees on each change are inefficient. Techniques such as sharing [6], the Zipper [7], or flattened AST representations [8] aim to optimise operations on immutable data.

This work explores an alternative approach: decoupling AST structure and data into separate representations, enabling more localized and efficient updates while preserving immutability. We propose a model for AST design and transformation that explicitly separates structure from data. It is shown how this design simplifies the definition of AST transformations without incurring heavy performance penalties, and how it aligns with the goals of clarity, locality, and immutability. These ideas are illustrated using a minimal compiler for a data query language, implemented in Java.

Query Language

To illustrate the proposed approach, we introduce a minimal language for data querying, called Query Language (QL). QL serves as the source language in our compiler, with SQL as its target. While the final compilation stage that emits SQL code is beyond the scope of this work, we focus on intermediate stages involving AST transformations.

The syntax of QL is defined using a grammar in extended Backus–Naur form (BNF):

```
Query   ::= (from Source)? yield Yield+
Source  ::= <name:string>
Yield   ::= Expr (as <alias:string>)?
Expr    ::= Val | Column | Query
Val     ::= <value:object>
Column  ::= <name:string>
```

This grammar uses conventions extended from traditional BNF: capitalized identifiers represent non-terminals, lowercase ones represent terminals, and postfix operators "?", "*", and "+" denote optionality, repetition (zero or more), and repetition (one or more), respectively. Angle bracket notation *<label:type>* (e.g., *<name:string>*) denotes arbitrary data of a corresponding type with a label for descriptive purposes. *<value:object>* denotes a literal value of an arbitrary type supported by QL, which is expected to include common data types used in SQL: temporal types (date, datetime), numeric (integer, decimal), etc.

Here are a few example queries that conform to the QL grammar:

1. *from users yield name, email*
2. *yield 1*
3. *from users yield (yield 1) as id, “admin” as name*

To demonstrate AST transformation in this context, three key operations are defined that a QL compiler might perform:

1. **Scalar subquery verification.** A scalar subquery is a *Query* node that appears within a *Yield* node. Since such subqueries are intended to produce a single scalar value, they must contain exactly one *Yield*. If this condition is violated, compilation should fail with an error. This is an example of a *verification* operation.
2. **Alias renaming.** This operation changes the alias associated with a *Yield* node, without modifying its structure. This is an example of a *data-only transformation.*
3. **Subquery inlining.** If a scalar subquery yields a constant *Value* and does not use a *Source*, it is semantically equivalent to that value and can be inlined by replacing the *Query* with the *Value*. This is an example of a *structural transformation*.

These operations serve as specific use cases for the approach developed in the rest of this work.

A naïve approach

A common and straightforward way to model an AST is to tightly couple its structure and data in a single representation. For the Query Language (QL), such a model might look like the following:

```
interface Expr
record Query(Source source, List<Yield> yields) implements Expr
record Yield(Expr expr, String alias) implements Expr
```

In this encoding, *Query* and *Yield* are both nodes in the AST. The *Query* node is purely structural, consisting of a source and a list of yields, while each *Yield* node combines a structural sub-expression *expr* with a data component *alias*.

The limitation of this design becomes apparent when implementing AST transformations. Because the structure and data are coupled, any transformation, regardless of scope, must begin at the root node (Query) and recursively traverse the tree to locate and manipulate the desired parts.

For example, consider the implementation of operation 1, which verifies that each subquery yields exactly one value:

```
verifyScalarSubquery(Query q)
  for (Yield y : q.yields())
    if (y.expr() instanceof Query sq)
      verifyScalarSubquery_(sq);

verifyScalarSubquery_(Query q)
  if (q.yields().size() != 1)
    error();
  else // (1)
    verifyScalarSubquery(q);
```

In this implementation, the top-level *Query* node must be explicitly walked. The compiler must inspect every *Yield* node, detect subqueries, and recursively apply the same verification. Line (1) illustrates a subtle point: a subquery may itself contain nested subqueries, so the traversal logic must be repeated.

This naïve representation may suffice for a toy language like QL, but in a realistic scenario with more constructs (e.g., *Where*, *OrderBy*), the number of places where subqueries might occur would increase significantly. The programmer would have to manually extend the traversal logic for each new language construct. As a result, the complexity of defining even simple transformations grows rapidly because of the traversal overhead. This is an example of accidental complexity overshadowing the essential complexity of the task.

The proposed solution

To overcome the limitations of tightly coupled AST representations, we propose a design that explicitly separates structure from data, enabling AST transformations to focus on the essential complexity, simplifying both the definition and execution of transformations under immutability constraints.

The AST is defined using interfaces that distinguish structural and data components. This separation is expressed through method declarations, with methods representing data annotated using the *@Data* annotation:

```
interface Query
  Source source();
  List<Yield> yields();

interface Yield
  Expr expr();
  @Data String alias();
```

Structural relationships, such as between *Yield* and *Expr*, are represented by unannotated methods, while data components, such as *Yield.alias*, are explicitly marked. These interfaces are processed at runtime to produce an internal representation of the AST, which is hidden from the programmer. The implementation details of this internal data structure will be presented in a subsequent section.

Defining transformations

An AST is defined by a set of node types. An AST transformation is a function from ASTs to ASTs – an endomorphism over the AST domain.

For example, in the QL compiler, let the AST type be *QL*. Then, any transformation can be expressed as a function *f: QL → QL.* This general definition, while expressive, is often impractical for defining specific transformations that target specific node types instead of the whole AST. Therefore, a notion of specific transformations, which are then lifted into general transformations via structural traversal, is also required.

A specific AST transformation is characterized by:

1. **A node type as the entry point.** Let *g* be a transformation defined over a node type *N*. It can be lifted to an AST transformation *f* by applying *g* to every subtree whose root is a node of type *N*. For example, renaming a yield alias (operation 2) is defined over *Yield* nodes, but can be lifted to operate on the entire AST by applying it to all nodes of type *Yield.*

2. **The nature of the change.** Transformations may involve two kinds of changes: *data-only changes*, which modify just the contents of a node, without altering the tree structure; *structural changes*, which affect the relationships between nodes (e.g., replacing one subtree with another).

Some operations, such as verification, may have no structural or data changes, relying solely on side effects (e.g., error reporting). Below it is shown how each of the three QL transformations introduced earlier can be implemented using this model.

Operation 1: Scalar subquery verification

```
verifyScalarSubquery(Yield yield)
  if (yield.expr() instanceof Query q && q.yields().size() != 1)
    error();
```

This operation is defined on *Yield* nodes and reports an error whenever a subquery does not match the requirements. Since the traversal is automatically handled by the abstraction, the implementation is minimal and free of accidental complexity. Only the essential logic needs to be specified, no traversal logic is required. This resembles a simple form of pattern matching in rewriting systems, where the transformation is applied only at relevant nodes, decoupled from traversal concerns.

Operation 2: Alias renaming

```
renameYield(Yield yield, String oldAlias, String newAlias, Yield.Factory yieldFactory)
  if (yield.alias().equals(oldAlias))
    replace(yield, yieldFactory.make(newAlias));
```

This operation demonstrates a data-only transformation: it changes the alias of a *Yield* without affecting its structure. The abstraction also provides a factory for each node type, which is simply a data constructor. This enables the programmer to construct new nodes using only the relevant data components. In this case, since *Yield* has only one data component (*alias*), the factory method has a single corresponding parameter. The structural component (*expr*) remains unchanged and can be reused in the resulting AST.

Operation 3: Subquery inlining

```
inlineSubquery(Yield y)
  if (y.expr() instanceof Query q && q.yields().size() == 1 && q.yields().getFirst() instanceof Value v)
    replace(y.expr(), v);
```

This example illustrates a structural transformation. When a scalar subquery yields a constant value and has no source, it can be inlined by replacing the subquery with the value it yields. Thanks to the abstraction, this replacement modifies the AST structure without requiring the programmer to manually reconstruct the entire tree. Notably, the transformation only replaces the relationship between a *Yield* and its *Expr*, leaving the *Yield* node itself unchanged. This illustrates the benefit of separating structure from data: structural updates can be performed without touching the data, while retaining the ability to compose them with data changes.

AST encoding

The internal representation of an AST consists of two components:

1. An array of nodes, where each element contains only data. The nodes are stored in the order of a depth-first pre-order traversal of the AST. This array captures the contents of the AST while remaining independent of its structure.

2. An adjacency matrix that encodes the structural relationships between nodes. The matrix is of size *n x n*, where *n* is the number of nodes in the array. For a given parent-child relationship between nodes *x* and *y*, the matrix entry at position *(i, j)* is present if node *x = a[i]* is the parent of node *y = a[j]*.

An AST is defined as a tuple *(a, m)*, where *a* is the array of data-only nodes, and *m* is the adjacency matrix capturing the structure.

Alternatively, an adjacency list could be used instead of an adjacency matrix, which would be most optimal for large but sparse ASTs, where the number of nodes tends to be large but the number of connections between nodes tends to be small. If ASTs do not have a clear tendency towards sparsity or density, an adaptive approach could be used, where the representation would be chosen dynamically based on the properties of the parser's input, such as its length. For illustrative purposes, an adjacency matrix is used, as it is trivial to map the transformations from one representation onto another.

Data-only changes

When performing data-only transformations, the structure of the AST remains unchanged. Therefore, the adjacency matrix *m* can be reused. To update an AST *(a, m)* by replacing a node *x* with a new version *x'* that differs only in its data components, we construct a new array *a'* as a copy of *a* with *x* replaced by *x'* at the same index. The resulting AST is *(a', m)*.

For example, consider the AST in Figure 1. A data-only change that renames the alias *"n"* to *"num"* in the *Yield* at index 2 produces a new AST as shown in Figure 2. Since the structure is unchanged, the same adjacency matrix is retained, and only the array of nodes is updated. This example illustrates the alias renaming operation presented earlier.

The following algorithm describes the necessary steps to perform this AST transformation:

1. Copy the array of nodes *a* into a new array *a'*.
2. For each *Yield y* in *a*, call the transformation function, and let its result be *y'*.
3. For each *y,* let its index in the array be *i*. Write *y'* to *a'[i].*
4. The result is a new AST – *(a', m)*, where *m* is the original adjacency matrix.

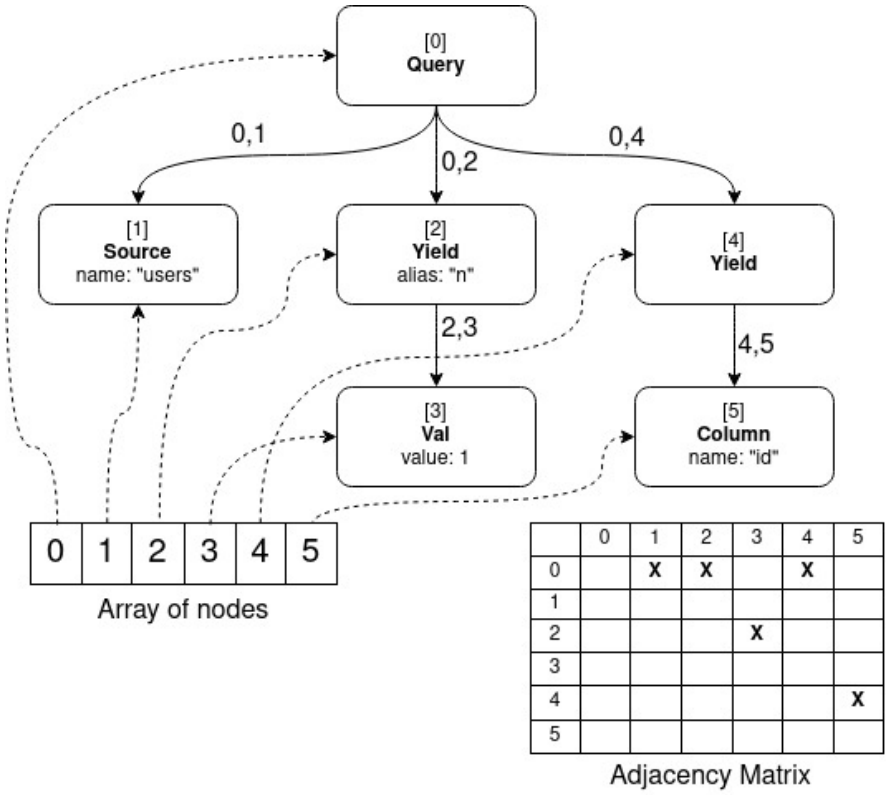


Fig. 1. AST before the data-only change

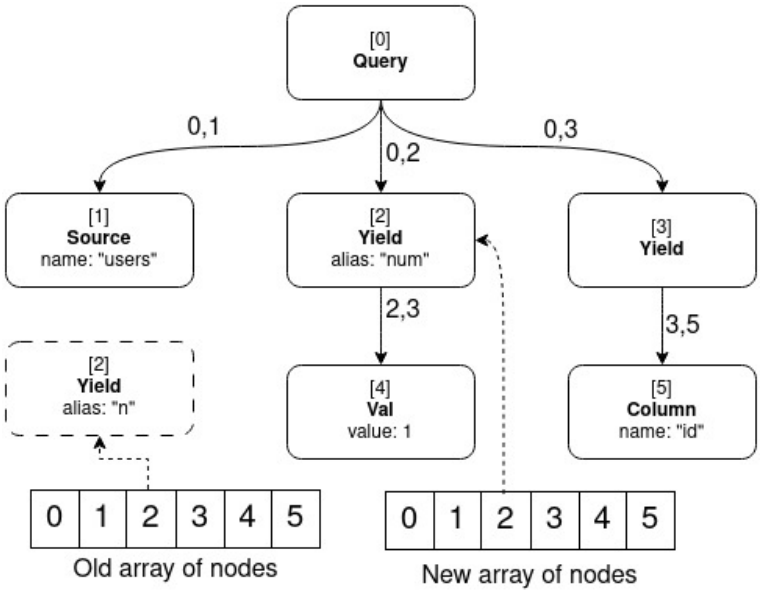


Fig. 2. AST after the data-only change

Structural changes

In contrast, structural transformations affect the shape of an AST and thus require modifications to both the node array and the adjacency matrix.

To replace a subtree rooted at node *x* with a new subtree rooted at *y*, the following steps are taken:

1. Identify subtrees *Tx* and *Ty*, rooted at *x* and *y*, respectively.
2. Construct a new AST *(a', m')* where *a'* is a copy of *a*, with the nodes in *Tx* removed and the nodes in *Ty* inserted in their place; *m'* is a copy of *m*, with the rows and columns corresponding to nodes in *Tx* replaced by those for *Ty*.

For instance, consider the AST in Figure 3, where the *Query* at index 3 is replaced by the *Value* at index 5. The result is shown in Figure 4. Nodes associated with the original subtree are removed from the array, the corresponding rows and columns in the adjacency matrix are pruned, and matrix entries are updated to reflect the new structure. This example illustrates the subquery inlining operation presented earlier.

The following algorithm describes the necessary steps to perform this AST transformation:

1. Copy the array of nodes *a* into a new array *a'*, and copy the adjacency matrix *m* into a new matrix *m'*.
2. For each *Yield y* in *a*, call the transformation function, and let its result be *(old, new)*, where *old* is the node to be removed from the AST, and *new* is to be inserted in its place.
3. Find the index of *old* in *a*, and let it be *i*.
4. Write *new* to *a'[i]*.
5. Find the index of *new* in *a*, and let it be *j*.
6. Write the contents of *m[j]* to *m'[i]*.
7. Trace and delete orphaned nodes in *a'* and their corresponding entries in *m'*.
8. The result is a new AST – *(a', m')*.

Related work

Sampson's work [8] explores an alternative AST representation where all nodes are stored in a contiguous array, and structural relationships are maintained through integer indices rather than pointers. This "flattened" layout improves memory locality, and enables efficient allocation and deallocation. While this representation is driven by performance and memory constraints, our approach complements performance gains by explicitly separating structure and data, providing additional flexibility and clarity in transformation definitions.

Balland et al. [9] present Tom, a strategy-based term rewriting system embedded in Java. Their work emphasizes the separation of rewriting rules from traversal strategies, enabling modular and declarative transformations. While Tom is a standalone language built on top of Java, our approach fits into the Java language and can be implemented as a library. Tom relies on pattern matching over algebraic terms and strategy interpretation at runtime, while our method encodes the AST structure explicitly via an adjacency matrix and provides an interface to control data and structure separately.

Hemel et al. [10] present an architecture for domain-specific language compilation centered around code generation via model-to-model transformation based on AST rewriting. Their work is focused on modular transformation pipelines that enable modular normalization rules that operate on immutable data. Similarly

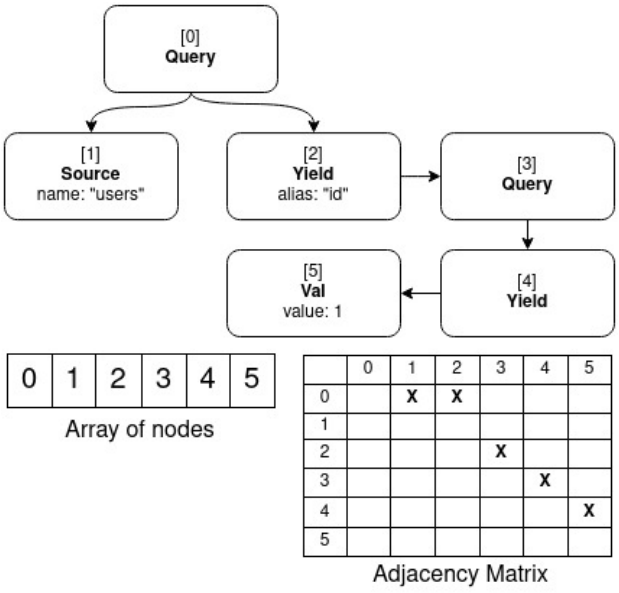


Fig. 3. AST before the structural change

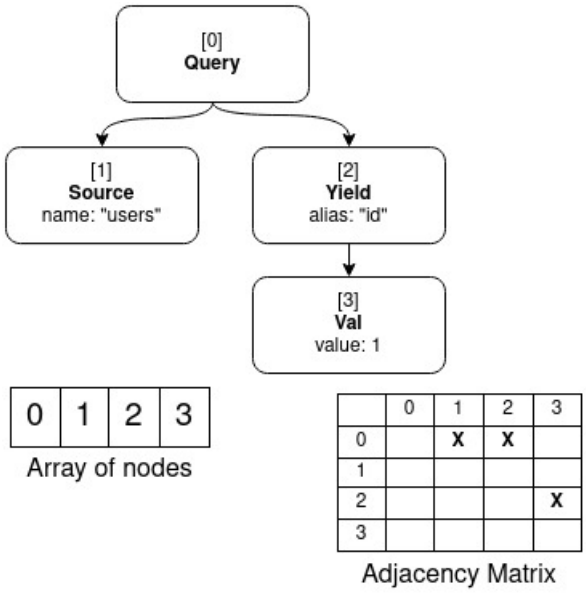


Fig. 4. AST after the structural change

to our approach, they seek to separate the essential complexity of AST transformation from traversal logic. While they focus on a standalone transformation language, our approach presents an encoding that can be integrated into Java directly.

Table 1 summarizes the key properties of the approach presented in this work in comparison with existing solutions. While the comparisons are mainly qualitative, focusing on expressiveness, ease of integration, and performance, they highlight the essential differences between the approaches.

Table 1

Comparison of AST transformation frameworks

Framework	Integration	Interface for transformations	Performance properties
This work	Library	Concise, high-level definition, decoupled from traversal	Adaptive choice of the AST representation based on language properties
Sampson [8]	Manual, must be adopted as a programming style	Unspecified	Helps avoid recursion during traversal, exploits cache locality
Balland et al. [9]	Standalone language	Rewrite rules based on pattern matching, decoupled from traversal	Transformations are optimized and compiled into Java bytecode
Hemel et al. [10]	Standalone language	Rewrite rules with custom traversal strategies	Multi-stage compilation of transformations

Conclusion

This work presented an approach to AST transformation that separates structure from data, enabling efficient operations on immutable data structures. By storing all AST nodes in an array and structural relationships in an adjacency matrix, the encoding supports localized updates without heavy performance penalties. Through AST transformation for a minimal query language as an example, it was shown how this approach eliminates accidental complexity and simplifies the implementation, supporting data-only and structural changes to the AST.

Compared to traditional object-oriented design, where transformation logic is often coupled with traversal concerns, our abstraction takes care of the accidental complexity. This facilitates informal reasoning, which is particularly important in modern software engineering. Unlike external systems that rely on additional rewriting languages, our approach stays within the Java language, providing the benefits of static typing and integration.

Future work may explore optimizations to the internal AST encoding and composability of transformations. The principles and architecture presented in this work provide a foundation for building compilers that are both simple and robust.

References

1. Ghosh, D. (2010). DSLs in action. Manning.
2. Bravenboer, M., Kalleberg, K. T., Vermaas, R., & Visser, E. (2006). Stratego/XT 0.16. In PEPM '06: Proceedings of the 2006 ACM SIGPLAN symposium on Partial evaluation and semantics-based program manipulation, 95–99. https://doi.org/10.1145/1111542.1111558.
3. Cordy, J. R., Dean, T. R., Malton, A. J., & Schneider, K. A. (2002). Source transformation in software engineering using the TXL transformation system. Information and Software Technology, 44(13), 827–837. https://doi.org/10.1016/s0950-5849(02)00104-0.
4. Brooks, F. (1987). No Silver Bullet Essence and Accidents of Software Engineering. Computer, 20(4), 10–19. https://doi.org/10.1109/mc.1987.1663532.
5. Helland, P. (2015). Immutability changes everything. Communications of the ACM, 59(1), 64–70. https://doi.org/10.1145/2844112.
6. Okasaki, C. (1998). Purely functional data structures. Cambridge University Press. https://doi.org/10.1017/cbo9780511530104.
7. Huet, G. (1997). The Zipper. Journal of Functional Programming, 7(5), 549–554. https://doi.org/10.1017/s0956796897002864.
8. Sampson, A. (2023, May 1). Flattening ASTs (and Other Compiler Data Structures). https://www.cs.cornell.edu/~asampson/blog/flattening.html.
9. Balland, E., Moreau, P., Reilles, A. (2008). Rewriting strategies in Java. Electronic Notes in Theoretical Computer Science, 219, 97–111. https://doi.org/10.1016/j.entcs.2008.10.037.
10. Hemel, Z., Kats, L. C. L., Groenewegen, D. M., & Visser, E. (2009). Code generation by model transformation: a case study in transformation modularity. Software & Systems Modeling, 9(3), 375–402. https://doi.org/10.1007/s10270-009-0136-1.

Дата першого надходження рукопису до видання: 05.05.2025
Дата прийнятого до друку рукопису після рецензування: 04.06.2025
Дата публікації: 25.07.2025

УДК 004.8:378.147; 378.147:004.42; 004.41
DOI https://doi.org/10.36994/2788-5518-2025-01-09-09

ІНТЕГРОВАНІ СЕРЕДОВИЩА РОЗРОБКИ З ПІДТРИМКОЮ ШТУЧНОГО ІНТЕЛЕКТУ: НОВІ МОЖЛИВОСТІ ДЛЯ ВИВЧЕННЯ ПРОГРАМУВАННЯ СТУДЕНТАМИ

Богун Р. І., *к.ф.-м.н., Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0002-8519-2234, bohun.roma@gmail.com*

Одрібець Н. В., *к.ф.-м.н., доц., Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0008-0176-1211, sosn@ukr.net*

Веденєєва О. А., *Відкритий міжнародний університет розвитку людини, Київ «Україна», Україна. https://orcid.org/0009-0006-0941-165X, vedenieieva@cyber.uu.edu.ua*

Анотація. *У статті проведено комплексний аналіз інтеграції середовищ розробки з підтримкою штучного інтелекту (ШІ) в процес вивчення програмування студентами технічних спеціальностей. Дослідження фокусується на тому, як інструменти GitHub Copilot, Cursor, Windsurf та інші трансформують традиційні підходи до навчання, пропонуючи нові можливості для персоналізації й підвищення ефективності. Наведено детальний порівняльний аналіз ключових платформ за такими критеріями, як базові великі мовні моделі (LLM), підтримка мов програмування, інтеграція з різними IDE й, що особливо важливо для освітнього середовища, умови надання освітніх ліцензій. При цьому окремо наголошується на практичних аспектах отримання пільгового доступу: складнощах валідації статусу студента, можливих регіональних обмеженнях і змінності умов. Розглянуто весь спектр режимів взаємодії із ШІ: від базового автодоповнення коду (Code Completion) і діалогового режиму (Chat), що перетворює IDE на інтерактивного тьютора, до просунутої генерації (Generate) й агентного режиму (Agent), здатного виконувати комплексні багатоетапні завдання. Окрему увагу приділено розширенню функціонала ШІ за межі IDE – в інструменти командного рядка (CLI), що відкриває нові горизонти для автоматизації та вивчення системного адміністрування. На основі цього аналізу запропоновано низку інноваційних навчальних завдань, спрямованих не на просте копіювання коду, а на розвиток критичного мислення, навичок пошуку й виправлення помилок та аналізу якості згенерованих рішень. Фінальна частина статті присвячена глибокому аналізу ризиків, зокрема надмірній залежності від інструментів, що може призвести до атрофії фундаментальних навичок, і викликам для академічної доброчесності. Запропоновано конкретні педагогічні стратегії для мінімізації цих загроз, що включають зміну парадигми оцінювання, упровадження усних захистів проєктів і формування в студентів культури відповідального й етичного використання ШІ.*

Ключові слова: *генеративний штучний інтелект, інтегроване середовище розробки, навчання програмування, програмування, GitHub Copilot, академічна доброчесність, генерація коду, командний рядок, великі мовні моделі, освітній процес.*

AI-POWERED INTEGRATED DEVELOPMENT ENVIRONMENTS: NEW OPPORTUNITIES FOR STUDENT PROGRAMMING EDUCATION

Roman Bohun, *Ph.D., Open International University of Human Development “Ukraine”, Kyiv, Ukraine. https://orcid.org/0009-0002-8519-2234, bohun.roma@gmail.com*

Nataliia Odribets, *Ph.D., Ass. Prof., Open International University of Human Development “Ukraine”, Kyiv, Ukraine. https://orcid.org/0009-0008-0176-1211, sosn@ukr.net*

Olga Vedenieieva, *Open International University of Human Development “Ukraine”, Kyiv, Ukraine. https://orcid.org/0009-0006-0941-165X, vedenieieva@cyber.uu.edu.ua*

Abstract. *The article provides a comprehensive analysis of the integration of AI-powered development environments into the process of learning programming for students in technical fields. The research focuses on how tools like GitHub Copilot, Cursor, Windsurf, and others are transforming traditional teaching approaches by offering new opportunities for personalization and efficiency. A detailed comparative analysis of key platforms is presented based on criteria such as their underlying large language models (LLMs), programming language support, integration with various IDEs, and, crucially for the educational context, the terms of educational licenses. Special emphasis is placed on the practical aspects of obtaining preferential access: the complexities of student status validation, potential regional restrictions, and the changing nature of the terms. The full spectrum of AI interaction modes is examined: from basic code*

completion and chat mode, which turns the IDE into an interactive tutor, to advanced generation and agent mode, capable of performing complex multi-step tasks. Special attention is given to the extension of AI functionality beyond the IDE into command-line interface (CLI) tools, opening new horizons for automation and learning system administration. Based on this analysis, a series of innovative educational assignments are proposed, aimed not at simple code copying but at developing critical thinking, debugging skills, and the ability to analyze the quality of generated solutions. The final part of the article is devoted to a deep analysis of the risks, particularly over-reliance on these tools, which can lead to the atrophy of fundamental skills, and the challenges to academic integrity. Specific pedagogical strategies are proposed to minimize these threats, including shifting assessment paradigms, implementing oral project defenses, and fostering a culture of responsible and ethical AI use among students.

Key words: *generative artificial intelligence; integrated development environment; learning to program; programming; GitHub Copilot; academic integrity; code generation; command-line interface; large language models; educational process.*

Вступ

Інтегровані середовища розробки (далі – IDEs) з підтримкою штучного інтелекту (далі – ШІ), зокрема GitHub Copilot, Cursor, Windsurf тощо, відкривають нові можливості для персоналізованого навчання програмування, що підтверджується результатами емпіричних досліджень у різних освітніх контекстах [1; 2; 3]. Практичне використання ШІ як тьютора або асистента демонструє зростання якості студентських проєктів, підвищення продуктивності та зменшення когнітивного навантаження [4; 5]. Водночас дослідники звертають увагу на необхідність критичного осмислення ролі ШІ в освітньому процесі, зокрема щодо формування довіри до результатів генерації коду й збереження мотивації до самостійного мислення [6; 1]. Це вимагає адаптації навчальних стратегій, які поєднують переваги автоматизації з розвитком алгоритмічного мислення й етичної відповідальності.

Згідно з аналітичними звітами, кількість користувачів ШІ-помічників зростає експоненційно, а компанії-розробники інвестують рекордні ресурси в їхній розвиток. За словами генерального директора Microsoft Сатьї Наделли, станом на 2025 рік до 30% коду в компанії вже генерується за допомогою ШІ [7]. GitHub Copilot, наприклад, уже має понад 15 мільйонів активних користувачів, а інші інструменти демонструють стрімке зростання популярності серед студентів і молодих фахівців. Цей тренд відображається й у практиці працевлаштування: усе більше роботодавців включають питання про досвід використання ШІ-середовищ у технічні співбесіди, оцінюючи здатність кандидатів ефективно працювати з інтелектуальними інструментами. IDE з підтримкою ШІ дедалі активніше використовуються не лише в професійній розробці [8], а й у навчальному процесі. Ці інструменти виступають як персоналізовані цифрові асистенти, що інтегруються в навчальний процес і суттєво полегшують засвоєння програмування. Особливо ефективними ці інструменти є на початкових етапах навчання, коли студенти стикаються з типовими труднощами – синтаксичними помилками, нерозумінням логіки алгоритмів, браком прикладів.

Виконання досліджень

IDEs з підтримкою ШІ забезпечують студентам доступ до інтелектуальної підтримки в реальному часі: автозаповнення коду, генерація функцій, пояснення логіки, пошук і виправлення помилок і рефакторинг. Завдяки цьому здобувачі освіти можуть швидше засвоювати нові мови програмування, експериментувати з кодом, учитися на помилках і водночас розвивати критичне мислення. За результатами досліджень, використання ШІ-помічників дає змогу зменшити час на виконання завдань, а також підвищити якість коду й упевненість студентів у власних силах [1; 2; 4].

Порівняльний аналіз ШІ-інструментів для розробки

Для об'єктивної оцінки можливостей сучасних ШІ-інструментів проведено порівняльний аналіз найпопулярніших рішень на ринку. Критеріями порівняння слугували доступність для закладів освіти, підтримка мов програмування й IDE, функціональні можливості й особливості використання різних великих мовних моделей (LLM).

Критерій	GitHub Copilot	Cursor	Windsurf (раніше Codeium)	Amazon CodeWhisperer	Tabnine
Основна LLM	OpenAI GPT-4.1/*	auto	OpenAI GPT-4.1/*, SWE-1*	Власна модель	Власна модель
Використання інших LLM	Так	Так	Так	Так (платф. Amazon Bedrock)	Так
Приблизна кількість користувачів	> 15 млн.	> 1 млн.	> 1 млн.	> 1 млн.	> 1 млн.
Підтримувані IDE	VS Code, JetBrains, ...	Власна IDE (форк VS Code)	VS Code, JetBrains, ...	VS Code, JetBrains, others	VS Code, JetBrains...
Підтримувані мови	Усі популярні мови (Python, JS, TS, Java, C++, Go, Ruby тощо)				

Безкоштовний план (з лімітами)	Так	Так	Так	Так	Ні (2025)
Ліцензія для студентів/ викладачів	Безкоштовно (Copilot Pro)/+	Безкоштовно 1 рік (Cursor Pro), знижки для студентів/-	Знижка 50% на Pro план/-	Безкоштовно (Individual tier)/+	Н/Д
Окрема IDE	Так	Ні	Так	Ні	Ні
Плагіни до IDE	Так	Ні	Так	Так	Ні
Основні можливості	Автодоповнення коду, чат, генерація функцій, пояснення, рефакторинг, CLI-помічник, агент	«AI-first» IDE, чат з кодовою базою, генерація «з нуля», авто-дебаг, міграція коду, агент	Швидке автодоповнення, чат в IDE, генерація коду за коментарями, агент	Автодоповнення, сканування безпеки, генерація коду, робота з AWS SDK	Контекстне автодоповнення, працює локально (підвищена приватність), інші
Переваги	Найкраща інтеграція, величезна спільнота, безкоштовно для освіти	Потужні функції для роботи з усім проєктом	Дуже широка підтримка IDE, швидкість	Глибока інтеграція з екосистемою AWS, акцент на безпеці	Висока швидкість, можливість роботи офлайн
Недоліки		Платні функції для повного доступу		Найкраще працює з AWS, менш універсальний	Відсутній безкоштовний план
Наявність підтримки	Документація, спільнота, підтримка				

Примітки. Варто зазначити, що ринок ШІ-інструментів для розробки активно розвивається, тому деякі дані, наведені в таблиці, можуть швидко змінюватися й потребують уточнення на офіційних ресурсах.

Варто зазначити, що отримання безкоштовного або пільгового доступу до цих інструментів за освітніми програмами зазвичай вимагає валідації статусу студента чи викладача. Цей процес може займати певний час і потребує надання відповідних документів. Крім того, умови таких програм лояльності можуть періодично змінюватися, мати часові обмеження або географічні рамки, розповсюджуючись, наприклад, лише на певні регіони, як-от США, відрізнятися для студентів і викладачів. Тому перед початком використання рекомендується ретельно ознайомитися з актуальними правилами на офіційних сайтах розробників.

Сучасні ШІ-помічники пропонують кілька фундаментальних режимів взаємодії, які можна адаптувати для освітніх цілей. Варто зазначити, що, хоча більшість сучасних інструментів підтримують перелічені режими, їхня реалізація, функціонал і рівень «агентивності» можуть суттєво відрізнятися, пропонуючи різну кількість і глибину можливостей. Загалом можна виділити такіі режими:

1. Режим асистента (Code Completion) – найпоширеніший режим, де ШІ пропонує варіанти завершення рядка або цілого блоку коду в реальному часі. Для студентів це миттєвий спосіб уникнути синтаксичних помилок і швидше писати шаблонний код.

2. Режим діалогу (Chat/Ask) – інтегрований чат, що дає змогу ставити запитання мовою, близькою до природної. Студент може запитати: «Поясни, як працює цей фрагмент коду», «Яка різниця між let і const у JavaScript?» або «Запропонуй кращий спосіб реалізації цього алгоритму». Це перетворює IDE на інтерактивного тьютора.

3. Режим генерації (Generate/Edit) – студент описує завдання (наприклад, «Створи функцію, що приймає URL і повертає JSON-дані»), а ШІ генерує повний код. Цей режим також використовується для рефакторингу (покращення наявного коду) або виправлення помилок (дебагінгу).

4. Режим агента (Agent) – найпросунутіший режим, де ШІ може виконувати складні, багатоетапні завдання. Наприклад, студент може поставити завдання: «Проаналізуй весь проєкт, знайди потенційні вразливості безпеки, запропонуй виправлення та напиши для них unit-тести». ШІ-агент самостійно розбиває завдання на кроки й виконує їх, взаємодіючи з файловою системою проєкту.

ШІ-помічники активно використовуються не лише в IDE, а і як інструменти командного рядка (CLI), що дає змогу отримувати інтелектуальну підтримку безпосередньо в терміналі для керування версіями, виконання скриптів та інших завдань. Основні можливості таких інструментів включають генерацію й пояснення системних команд, створення скриптів автоматизації та допомогу в роботі з Git. Серед ключових інструментів виділяються офіційні GitHub Copilot CLI й нещодавно представлений Google Gemini CLI, які перетворюють термінал на інтерактивне навчальне середовище. Anthropic

також надає офіційну документацію для використання Claude з командного рядка, а для моделей OpenAI існують численні інструменти від спільноти, що розвиває в студентів навички роботи з API й автоматизації.

Дослідження показують, що ефективність ШІ-інструментів залежить від складності завдання та LLM, що лежить в їх основі. Наприклад, у стандартних тестах на генерацію коду, як-от HumanEval, моделі GPT-4 (Copilot, Cursor) і Claude 3 Opus (Cursor) часто демонструють найвищі результати. Водночас Amazon CodeWhisperer може бути ефективнішим у завданнях, пов'язаних з інфраструктурою AWS.

На основі цих можливостей можна сформулювати такі види навчальних робіт для студентів:

1. Критик ШІ: студент дає ШІ завдання згенерувати код для вирішення певної проблеми, а потім має проаналізувати отриманий результат: знайти помилки, неефективні місця, потенційні вразливості й запропонувати власне, покращене рішення.

Приклад. Попроси CodeWhisperer написати функцію для обробки завантажених файлів. Проаналізуй її на наявність вразливостей типу «Path Traversal».

Переваги: формує критичне мислення, учить не довіряти сліпо технологіям і розуміти глибинні аспекти проблеми.

2. Рефакторинг-челендж: викладач надає застарілий або «погано написаний» код. Завдання студента – за допомогою ШІ-інструментів провести його повний рефакторинг: покращити читабельність, оптимізувати продуктивність, додати коментарі та документацію.

Приклад. Використовуючи функцію «Edit» у Cursor, перетвори цей код на основі колбеків на сучасний синтаксис з async/await.

Переваги: навчає сучасних стандартів кодування й важливості підтримки кодової бази.

3. Дебагінг-квест: викладач навмисно вносить у код неочевидну помилку. Завдання студента – використовуючи можливості ШІ для налагодження (пояснення коду, пропозиції виправлень, аналіз помилок), знайти й виправити її.

Приклад. У цьому коді є помилка, пов'язана з асинхронними операціями, яка призводить до «race condition». Використовуй чат GitHub Copilot, щоб проаналізувати логіку та знайти причину проблеми.

Переваги: розвиває навички системного налагодження, учить використовувати ШІ як діагностичний інструмент.

4. TDD-спринт (Test-Driven Development): студент за допомогою ШІ спочатку генерує набір unit-тестів для ще не написаної функціональності, а потім пише код, який має задовольнити ці тести.

Приклад. Сформулюй для Tabnine запит на створення unit-тестів для функції калькулятора, що покривають додавання, віднімання, ділення на нуль і роботу з нечисловими даними.

Переваги: прищеплює культуру тестування та розробки через тестування (TDD).

5. API-дослідник: студентам дається завдання інтегрувати в проєкт новий, незнайомий сторонній API. Вони повинні використовувати ШІ-асистент для вивчення документації, генерації прикладів запитів і написання коду інтеграції.

Приклад. Використовуючи Cursor, інтегруй API для прогнозу погоди в цей додаток. Згенеруй код для запиту поточної погоди за назвою міста й відобрази результат.

Переваги: імітує поширене реальне завдання, навчає швидко освоювати й застосовувати нові технології.

6. Навігатор по коду: студенти отримують великий, незнайомий проєкт (наприклад, open-source бібліотеку) і за допомогою ШІ-чату повинні пояснити архітектуру, призначення ключових модулів і логіку роботи основних функцій.

Приклад. Використовуючи Copilot Chat, опиши, як реалізовано механізм маршрутизації в цьому веб-фреймворку.

Переваги: розвиває навички аналізу чужого коду, що є ключовим для роботи в команді.

7. Архітектор рішень: студенти отримують опис проблеми високого рівня й за допомогою ШІ повинні розробити архітектуру застосунку: обрати технології, спроєктувати схему бази даних, структурувати проєкт.

Приклад. Спроєктуй архітектуру для простої блог-платформи. Запитай у Claude поради щодо вибору між монолітною та мікросервісною архітектурою. Згенеруй базову структуру папок і файлів для обраного підходу.

Переваги: навчає системного проєктування високого рівня, спонукає думати про компроміси й архітектурні патерни.

Попри значні переваги, інтеграція ШІ в навчання несе суттєві ризики, які потребують детального аналізу та пропрацьованих стратегій мінімізації:

1. Надмірна залежність та «атрофія» навичок: студенти можуть утратити здатність самостійно вирішувати проблеми, оскільки звикають отримувати готові рішення від ШІ. Це може призвести до браку фундаментального розуміння алгоритмів і структур даних.

2. Галюцинації та неякісний код: ШІ-моделі іноді генерують код, який виглядає правдоподібно, але містить логічні помилки, вразливості безпеки або є неоптимальним. Студент без належного досвіду може не помітити цих недоліків.

3. Плагіат та академічна недоброчесність: межа між допомогою інструмента й плагіатом стає розмитою. Визначити, студент самостійно дійшов до рішення чи просто скопіював відповідь ШІ, стає надзвичайно складно для викладача.

4. Зниження креативності: постійне використання готових шаблонів може пригнічувати здатність студентів до пошуку нестандартних, інноваційних рішень.

Способи контролю й вирішення проблеми бездумного використання

Проблема «скопіював-вставив» є центральним викликом. Для її вирішення потрібен комплексний підхід, що поєднує технологічні, педагогічні й адміністративні заходи.

Запропоновані варіанти вирішення:

1. Зміна парадигми завдань: замість завдань, що мають одну правильну відповідь (наприклад, «напишіть функцію сортування»), варто пропонувати завдання відкритого типу, що вимагають аналізу й обґрунтування.

Приклад. Дослідіть три різні алгоритми сортування, згенеруйте їх реалізації за допомогою ШІ. Порівняйте їхню продуктивність на різних наборах даних. Обґрунтуйте, який алгоритм є найкращим для кожного випадку, і поясніть, чому згенерований ШІ код може бути неоптимальним.

2. Упровадження усної захисту робіт: студент повинен бути готовим пояснити кожний рядок свого коду, логіку прийнятих рішень та альтернативні шляхи вирішення проблеми. Це унеможливлює бездумне копіювання.

3. Використання інструментів для детекції ШІ-згенерованого коду: хоча такі інструменти не є ідеальними, вони можуть слугувати допоміжним сигналом для викладача, указуючи на роботи, що потребують ретельнішої перевірки.

4. Навчання «ШІ-грамотності»: включення в навчальні програми модулів, присвячених етичному й ефективному використанню ШІ. Студентів варто навчати, як правильно формулювати запити (prompt engineering), як критично оцінювати результати та як посилатися на використання ШІ-інструментів у своїх роботах, аналогічно до цитування наукових джерел.

5. Розроблення чітких інституційних політик: заклад вищої освіти має розробити й донести до студентів і викладачів чіткий кодекс академічної доброчесності в епоху ШІ, де буде визначено, що є допустимим використанням, а що вважається плагіатом.

Висновок

Отже, інтегровані середовища розробки з підтримкою ШІ є не лише інструментами автоматизації, а й потужними засобами розвитку цифрової грамотності, аналітичного мислення та професійної готовності студентів до викликів сучасної ІТ-індустрії.

Інтегровані середовища розробки з підтримкою ШІ створюють нові умови для ефективного, адаптивного й мотиваційного навчання програмування, сприяючи розвитку продуктивності, критичного мислення та цифрової грамотності студентів. На особливу увагу заслуговує GitHub Copilot завдяки своєму широкому функціоналу, підтримці багатьох мов програмування, інтеграції з популярними IDE та наявності безкоштовного доступу для викладачів і студентів через програму GitHub Education [9]. Подальші дослідження мають бути спрямовані на розроблення методик оцінювання навчального прогресу в умовах ШІ-підтримки, адаптації матеріалів навчальних курсів і вивчення впливу таких технологій на професійну готовність здобувачів освіти.

Література

1. Alanazi M., Soh B., Samra H., Li A.L.C. The Influence of Artificial Intelligence Tools on Learning Outcomes in Computer Programming: A Systematic Review and Meta-Analysis. *Computers*. 2025. Vol. 14, № 5. Article 185. DOI: https://doi.org/10.3390/computers14050185.

2. Gardella N., Pettit R., Riggs S.L. Performance, Workload, Emotion, and Self-Efficacy of Novice Programmers Using AI Code Generation. *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education (ITiCSE 2024).* Vol. 1. Milan, Italy, 2024. P. 290–296. DOI: https://doi.org/10.1145/3649217.3653615.

3. Avramovic S., Avramovic I., Wojtusiak J. Exploring the Impact of GitHub Copilot on Health Informatics Education. *Applied Clinical Informatics*. 2024. Vol. 15, № 5. P. 1121–1129. DOI: https://doi.org/10.1055/a-2414-7790.

4. Timcenko O. Case Study: Using Artificial Intelligence as a Tutor for a Programming Course. *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)*. Paris, France, 2024. P. 1–4. DOI: https://doi.org/10.1109/ITHET61869.2024.10837624.

5. Mesaros G.O. Learning Web Development Using GitHub Copilot in and Outside Academia: A Blessing or a Curse? *Interdisciplinary Description of Complex Systems*, 2024. Vol. 22, № 3. P. 355–359. DOI: https://doi.org/10.7906/indecs.22.3.10.

6. Shah A., Chernova A., Tomson E., Porter L., Griswold W.G., Raj A.G.S. Students' Use of GitHub Copilot for Working with Large Code Bases. *Proceedings of the 56th ACM Technical Symposium on Computer Science Education (SIGCSE TS 2025).* Vol. 1. Pittsburgh, PA, 2025. P. 1050–1056. DOI: https://doi.org/10.1145/3641554.3701800.

7. Novet J., Vanian J. Satya Nadella says as much as 30% of Microsoft code is written by AI. URL: https://www.cnbc.com/2025/04/29/satya-nadella-says-as-much-as-30percent-of-microsoft-code-is-written-by-ai.html (date of access: 06.07.2025).

8. Подвишенна Д. Вайбкодинг: чому всі про нього говорять і коли він справді потрібен. URL: https://dou.ua/lenta/articles/what-is-vibecoding-developement/ (дата звернення: 06.07.2025).

9. Getting free access to Copilot Pro as a student, teacher, or maintainer. URL: https://docs.github.com/en/copilot/managing-copilot/managing-copilot-as-an-individual-subscriber/getting-started-with-copilot-on-your-personal-account/getting-free-access-to-copilot-pro-as-a-student-teacher-or-maintainer (date of access: 06.07.2025).

References

1. Alanazi, M., Soh, B., Samra, H., Li, A.L.C. (2025). The Influence of Artificial Intelligence Tools on Learning Outcomes in Computer Programming: A Systematic Review and Meta-Analysis. *Computers*, 14 (5), 185. https://doi.org/10.3390/computers14050185.

2. Gardella, N., Pettit, R., Riggs, S.L. (2024). Performance, Workload, Emotion, and Self-Efficacy of Novice Programmers Using AI Code Generation. *In Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education (ITiCSE 2024), Vol. 1* (pp. 290–296). Milan, Italy. https://doi.org/10.1145/3649217.3653615.

3. Avramovic, S., Avramovic, I., Wojtusiak, J. (2024). Exploring the Impact of GitHub Copilot on Health Informatics Education. *Applied Clinical Informatics,* 15 (5), 1121–1129. https://doi.org/10.1055/a-2414-7790.

4. Timcenko, O. (2024). Case Study: Using Artificial Intelligence as a Tutor for a Programming Course // *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1–4). Paris, France. https://doi.org/10.1109/ITHET61869.2024.10837624.

5. Mesaros, G.O. (2024). Learning Web Development Using GitHub Copilot in and Outside Academia: A Blessing or a Curse? *Interdisciplinary Description of Complex Systems*, 22 (3), 355–359. https://doi.org/10.7906/indecs.22.3.10.

6. Shah, A., Chernova, A., Tomson, E., Porter, L., Griswold, W.G., Raj, A.G.S. (2025). Students' Use of GitHub Copilot for Working with Large Code Bases. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education, (SIGCSE TS 2025), Vol. 1* (pp. 1050–1056). Pittsburgh, PA. https://doi.org/10.1145/3641554.3701800.

7. Novet, J., Vanian, J. (2025). Satya Nadella says as much as 30% of Microsoft code is written by AI. *cnbc.com*. Retrieved from https://www.cnbc.com/2025/04/29/satya-nadella-says-as-much-as-30percent-of-microsoft-code-is-written-by-ai.html (date of access: 06.07.2025).

8. Podvyshenna, D. (2025). Vaibkodynh: chomu vsi pro noho hovoriat i koly vin spravdi potriben [Vibecoding: why everyone's talking about it and when it actually matters]. *dou.ua*. Retrieved from https://dou.ua/lenta/articles/what-is-vibecoding-developement/ (date of access: 06.07.2025) [in Ukrainian].

9. Getting free access to Copilot Pro as a student, teacher, or maintainer. *docs.github.com*. Retrieved from https://docs.github.com/en/copilot/managing-copilot/managing-copilot-as-an-individual-subscriber/getting-started-with-copilot-on-your-personal-account/getting-free-access-to-copilot-pro-as-a-student-teacher-or-maintainer (date of access: 06.07.2025).

Дата першого надходження рукопису до видання: 14.05.2025
Дата прийнятого до друку рукопису після рецензування: 12.06.2025
Дата публікації: 25.07.2025

UDC 004.8
DOI https://doi.org/10.36994/2788-5518-2025-01-09-10

COMPARATIVE ANALYSIS OF ADVERSARIAL-DIFFUSION HYBRID ARCHITECTURES OF GENERATIVE MODELS

Bondar V. V., *Ph.D. student, Cherkasy State Technological University, Cherkasy, Ukraine. https://orcid.org/0009-0002-3230-5979, v.v.bondar.asp24@chdtu.edu.ua*

Babenko V. H., *D.Scs., Prof., Cherkasy State Technological University, Cherkasy, Ukraine. https://orcid.org/0000-0003-2039-2841, v.babenko@chdtu.edu.ua*

Abstract. *This paper examines the emerging integration of adversarial objectives into diffusion-based generative frameworks, with the purpose of systematically analyzing how hybrid architectures can overcome the limitations of individual approaches while combining their respective strengths. We identify and compare three distinct categories of hybrid models: distillation-based approaches that compress diffusion models using adversarial feedback, in-loop methods that incorporate discriminators during diffusion training, and diffusion-assisted generative adversarial networks that leverage noise processes to stabilize adversarial learning. Our comparative analysis reveals critical trade-offs across generation speed, sample fidelity, training stability, and application suitability. Distillation-based hybrids achieve near-instant high-quality generation but require intensive training processes, while in-loop adversarial diffusion methods offer superior image quality with moderate sampling efficiency. Diffusion-assisted generative adversarial networks maintain single-pass generation while addressing traditional generative adversarial networks instabilities through diffusion mechanisms. Empirical evidence demonstrates that these hybrid approaches consistently outperform pure diffusion or adversarial methods in quality and efficiency metrics. We identify significant research gaps, particularly regarding the theoretical unification of adversarial and score-matching objectives, scaling challenges for high-resolution generation, and expansion beyond image synthesis to other modalities. This analysis provides a foundation for understanding how complementary strengths of diffusion and adversarial techniques can be leveraged to develop more effective generative frameworks combining distribution coverage with perceptual realism.*

Key words: *diffusion models, generative adversarial networks, generative models, training stability, sampling efficiency, hybrid architectures, deep learning.*

ПОРІВНЯЛЬНИЙ АНАЛІЗ ГІБРИДНИХ ГЕНЕРАТИВНИХ МОДЕЛЕЙ З ДИФУЗІЙНИМИ ТА ГЕНЕРАТИВНО-ЗМАГАЛЬНИМИ ЕЛЕМЕНТАМИ

Віталій Бондар, *здобувач ступеня доктор філософії, Черкаський державний технологічний університет, Черкаси, Україна. https://orcid.org/0009-0002-3230-5979, v.v.bondar.asp24@chdtu.edu.ua*

Віра Бабенко, *д.т.н., проф., Черкаський державний технологічний університет, Черкаси, Україна. https://orcid.org/0000-0003-2039-2841, v.babenko@chdtu.edu.ua*

Анотація. *Робота присвячена дослідженню інтеграції принципів генеративно-змагальних мереж у генеративні фреймворки на основі дифузії з метою систематичного аналізу щодо того, як гібридні архітектури можуть подолати обмеження окремих підходів, поєднуючи їхні переваги. Виокремлено та порівняно три категорії гібридних моделей: підходи на основі дистиляції, які стискають дифузійні моделі за допомогою змагального зворотного зв'язку; методи поєднання в циклі навчання, що включають дискримінатори під час тренування дифузії; генеративно-змагальні мережі з підтримкою дифузії, які використовують шумові процеси для стабілізації змагального навчання. Порівняльний аналіз дав змогу виявляти критичні компроміси між швидкістю генерації, точністю зразків, стабільністю навчання й придатністю для застосування. Гібриди на основі дистиляції досягають майже миттєвої високоякісної генерації, але потребують інтенсивних процесів навчання, тоді як додавання дискримінатора в цикли дифузійних методів пропонують вищу якість зображень з помірною ефективністю семплювання. Генеративно-змагальні мережі з підтримкою дифузії зберігають одноетапну генерацію, вирішуючи традиційні проблеми нестабільності генеративно-змагальної мережі через дифузійні механізми. Емпіричні дані демонструють, що ці гібридні підходи послідовно перевершують чисті дифузійні або змагальні методи за показниками якості й ефективності. Подано опис виявлених значних прогалин у дослідженнях, зокрема щодо теоретичного об'єднання функцій утрат змагального та дифузійних підходів, проблем масштабування для високороздільної генерації й розширення за межі синтезу зображень на інші модальності. Проведений у роботі аналіз закладає основу для розуміння того, як можна викори-*

стати комплементарні переваги дифузійних і змагальних технік для розроблення ефективніших генеративних фреймворків, що поєднують покриття розподілу з покращеним реалізмом.

Ключові слова: *дифузійні моделі, генеративно-змагальні мережі, генеративні моделі, стабільність навчання, ефективність семплювання, гібридні архітектури, глибоке навчання.*

Introduction

Integrating adversarial objectives into diffusion models has significantly advanced the generation of high-fidelity samples while preserving the theoretical guarantees of score-based approaches. Diffusion models have emerged as powerful generative frameworks that learn to reverse a predefined noise process through score matching, ensuring convergence to the true data distribution [1]. Despite their theoretical soundness, these models often produce overly smooth or blurred outputs when sampling with limited denoising steps or in compressed latent spaces. Integrating adversarial losses overcomes this fundamental limitation by introducing a discriminator that distinguishes between real and generated samples, effectively compelling the generator to capture fine-grained details that pure score matching might miss [2].

Adversarial-diffusion hybrids can be broadly categorized into several architectural approaches, each with distinct advantages. Distillation-based methods represent a new category in which adversarial training enhances the process of compressing a large "teacher" model into a more efficient "student". For instance, Adversarial Diffusion Distillation (ADD) optimizes both score-distillation and adversarial losses for a few-step diffusion process to preserve sharp details that might otherwise be lost during compression [3]. Building on this foundation, progressive schemes such as SDXL-lightning implement sequential adversarial refinement across multiple distillation stages, systematically improving the student's output fidelity [4].

Beyond distillation, researchers have developed techniques that embed adversarial components directly within the diffusion training process. Constant Acceleration Flow introduces an innovative flow discriminator that evaluates whether intermediate denoising trajectories align with an accelerated process, thereby enforcing more realistic dynamics and faster convergence [5]. Similarly, Structure-Guided Adversarial Training employs a manifold discriminator operating in latent space to ensure that learned score fields maintain consistency with the underlying data geometry [6].

The purpose of this paper is to provide a comprehensive analysis of emerging adversarial-diffusion hybrid architectures, examining their respective strengths and limitations in terms of sample quality, generation speed, and training stability. We aim to identify critical research gaps and promising directions for future work, particularly regarding the theoretical unification of adversarial and score-matching objectives and their potential applications to complex multimodal datasets.

1. Background

1.1.Diffusion Models

Diffusion models represent a class of generative approaches that learn data distributions by reversing a gradual noise addition process through score matching techniques. These models operate by adding Gaussian noise to data across multiple timesteps, then training a neural network to estimate the gradient of the log data density at each noise level [1]. The generative process begins from pure noise and progressively denoises the sample by following the estimated score function, typically implemented via discretized stochastic differential equation (SDE) or ordinary differential equation (ODE). The training objective minimizes the error between network predictions and the actual noise added to the data, providing theoretical guarantees on distribution matching and ensuring robust mode coverage. Despite these advantages, diffusion models often require numerous sampling steps during generation, resulting in computational inefficiency and potentially blurred outputs when using fewer steps [7]. Recent research has focused on accelerating sampling procedures while maintaining generation quality through techniques such as distillation and improved ODE solvers [8].

1.2.Adversarial Training

Adversarial training frameworks employ a competitive learning paradigm between generator and discriminator networks to produce high-fidelity samples. In the canonical generative adversarial network (GAN) architecture, a generator transforms random noise into synthetic samples while a discriminator learns to distinguish between real and generated data [9]. The generator optimizes an adversarial loss designed to maximize the probability of the discriminator misclassifying generated samples as real, thereby encouraging the production of increasingly realistic outputs. This adversarial dynamic enables GANs to capture fine-grained perceptual details and high-frequency components that enhance visual quality beyond what traditional likelihood-based models achieve. However, despite their ability to generate sharp, realistic images, adversarial training suffers from fundamental limitations, including training instability, mode collapse, and sensitivity to hyperparameter selection [10]. These challenges have motivated extensive research into stabilization techniques, such as spectral normalization, and alternative objective functions, such as the Wasserstein distance [11].

2. Adversarially-Refined Diffusion Models

Integrating adversarial training techniques with diffusion-based generation significantly improves sample fidelity and sampling efficiency in generative modeling. Adversarially-refined diffusion models utilize

discriminator networks to critique diffusion outputs, thereby creating a training dynamic where the diffusion model learns to produce increasingly realistic samples. This section examines two distinct approaches: distillation-based hybrids that compress diffusion models into faster generators and in-loop adversarial methods that directly incorporate adversarial feedback during diffusion training.

2.1.Distillation-based Hybrids

Knowledge distillation combined with adversarial supervision enables the compression of diffusion models while maintaining high-quality outputs. Distillation-based hybrids employ a teacher-student paradigm where a pre-trained diffusion model (teacher) transfers its capabilities to a simplified student model and adds the ability of accelerated generation [5]. The student receives dual supervision through a distillation loss matching the teacher's denoising behavior and an adversarial loss pushing outputs toward the real data distribution. This adversarial guidance effectively addresses the quality degradation typically observed when aggressively compressing diffusion models to fewer steps.

Adversarial Diffusion Distillation (ADD) demonstrates how score-based distillation and GAN-style training can be effectively combined to train one-step generative models [3]. In ADD, a student network distills knowledge from a large diffusion model like Stable Diffusion while simultaneously learning to fool a discriminator judging the realism of generated images. The discriminator provides crucial feedback that forces the student to capture fine details often missed by pure distillation approaches. This combination enables ADD to achieve remarkable image fidelity even when generating samples in just 1–4 denoising steps, effectively combining GAN-like speed with diffusion model quality.

Latent Adversarial Diffusion Distillation (LADD) extends this approach by applying adversarial distillation in latent space to enable high-resolution generation. LADD operates in the compressed latent domain of an autoencoder but applies adversarial loss to decoded outputs, ensuring the preservation of high-frequency details. By shifting adversarial distillation to latent feature space, LADD dramatically reduces computational requirements for large-resolution synthesis compared to pixel-space ADD. The latent-domain discriminator provides effective feedback on both global structure and local texture, allowing models like SD3-Turbo to match state-of-the-art image quality using only four denoising steps [12].

SDXL-Lightning implements a progressive multi-stage distillation process with adversarial guidance at each stage. Rather than directly distilling to a one-step model, SDXL-Lightning creates a sequence of intermediate models with progressively fewer sampling steps. Each distillation stage employs adversarial loss to maintain image quality while increasing generation speed. This staged approach preserves diversity in earlier phases while focusing on detail refinement in later stages, producing a one-step generator that achieves high image quality at 1024×1024 resolution without the fidelity loss typical of aggressive distillation [4].

Diffusion2GAN takes a different approach by framing distillation as a supervised image translation problem. This method generates paired examples from a diffusion teacher (random noise vectors and corresponding images) and trains a GAN to replicate this mapping in a single step. The adversarial loss ensures the GAN's outputs match both the diffusion model's results and real images. Using pre-trained discriminators and multiscale adversarial feedback, Diffusion2GAN stabilizes training and successfully captures the complex distribution learned by the diffusion model, creating a one-step generator with diffusion-level quality and GAN-like interactivity [13].

2.2.In-loop Adversarial Diffusion

Incorporating adversarial loss directly into the diffusion training process enhances model capabilities without requiring separate distillation. In-loop adversarial diffusion methods train diffusion models from scratch with discriminators guiding the learning process. Unlike distillation approaches, the discriminator evaluates intermediate denoising states or sample sets during training, transforming diffusion optimization into a minimax game similar to GANs. This direct integration helps enforce global realism and structural consistency that standard diffusion losses might not capture.

Constant Acceleration Flow (CAF) employs discriminators to guide ODE-based diffusion trajectories during training. CAF models the diffusion sampling with an acceleration term and trains this acceleration predictor using adversarial objectives. The discriminator distinguishes between learned model trajectories and ground truth diffusion trajectories from training data, effectively correcting the generation path when it deviates from realistic patterns. This continuous adversarial guidance helps CAF avoid flow-crossing issues and quality degradation observed in other fast-generation techniques, resulting in more accurate few-step sampling [5].

Structure-Guided Adversarial Training (SADM) uses specialized discriminators to enforce structural relationships present in real data. Rather than evaluating individual samples, SADM's structure discriminator assesses the organization of multiple samples or latent representation consistency. This approach applies adversarial loss to the diffusion model's score field in latent space, forcing the model to produce denoising directions that align with the real data distribution structure. By adding manifold-level supervision to traditional instance-level diffusion loss, SADM significantly improves generation fidelity and diversity, achieving better FID scores on challenging benchmarks like high-resolution ImageNet [6].

Adversarially-refined diffusion represents a powerful synthesis of diffusion models' diversity with GANs' realism and efficiency. By strategically incorporating adversarial feedback through distillation or in-loop

training, these approaches overcome the traditional limitations of both frameworks, delivering high-fidelity, high-speed generative models that represent the current state-of-the-art in image synthesis.

3. Diffusion-Assisted GANs

Diffusion-assisted GANs integrate the denoising diffusion process within GAN frameworks to address the inherent instability of adversarial training. These hybrid models improve sample quality and training stability by combining structured noise processes with GAN discriminator feedback. The integration occurs through two primary mechanisms: noise-based regularization of GAN training dynamics and diffusion processes operating in compressed latent spaces. Empirical results demonstrate that this combination of diffusion and adversarial learning produces more realistic outputs while mitigating typical GAN training issues, such as mode collapse and convergence difficulties.

3.1. Noise-Smoothing GANs

Noise-smoothing GANs use diffusion processes to inject controlled noise during training, effectively regularizing adversarial learning dynamics. Diffusion-GAN uses this approach by applying a forward diffusion chain to both real and generated images, progressively adding Gaussian noise. A timestep-dependent discriminator then distinguishes between diffused real and diffused generated images at multiple noise levels. This mechanism provides smoother, more consistent feedback to the generator rather than emphasizing high-frequency details. The generator optimizes its parameters by backpropagating gradients through the diffusion process, learning to create images that remain realistic under significant noise perturbations. This technique stabilizes training and improves both image quality and data efficiency compared to conventional GANs [14].

Adversarial Noise Scheduling represents another noise-smoothing strategy that optimizes the diffusion noise schedule itself through adversarial training. Rather than using fixed schedules like linear or cosine decay, this method trains a noise schedule generator alongside the diffusion model. The schedule adjusts to maximize sample realism as determined by an adversarial criterion. The diffusion process learns optimal noise injection or removal at each step based on discriminator feedback comparing generated outputs with real images. By treating the noise schedule as a learnable component in the adversarial framework, the model adapts its denoising trajectory to produce outputs that better deceive the discriminator. This adversarially optimized scheduling leads to smoother denoising and higher-fidelity samples while sometimes improving sampling efficiency compared to hand-crafted schedules [15].

3.2. Latent-Space Hybrids

Latent-space hybrid methods perform diffusion in lower-dimensional latent spaces while applying adversarial losses either on output images or latent features. Latent Denoising Diffusion GAN (LDDGAN) utilizes a pre-trained autoencoder to compress images into compact latent codes, where a diffusion model then generates new samples. Since denoising operates on smaller latent representations, the model samples significantly faster than pixel-space diffusion while maintaining distribution coverage. After decoding latent samples back to images, a GAN discriminator evaluates them against real images, enforcing an adversarial loss that enhances realism. This division of labor allows the diffusion model to handle coarse structure and diversity in latent space while the discriminator refines fine details in image space. LDDGAN achieves faster sampling without sacrificing image quality, reporting improved fidelity and diversity compared to standard diffusion models while maintaining competitive generation speed [16].

LaDiffGAN represents an alternative latent-space hybrid that employs diffusion-based denoising as supervision for GAN training in latent space. Unlike LDDGAN, LaDiffGAN does not perform diffusion during inference; instead, it uses diffusion noise during training to guide the GAN's learning process. The generator operates on latent codes that are randomly perturbed with Gaussian noise of varying magnitudes during training. The model learns to denoise corrupted latent inputs while producing corresponding output images. A special latent diffusion GAN loss aligns generated image features with real image features under noise perturbations, effectively teaching the generator to reverse diffusion steps in latent space. Simultaneously, a standard adversarial loss ensures visual realism by comparing generated and real images. This dual training signal, combining diffusion-style denoising in latent space with image-level adversarial loss, stabilizes the GAN and improves its ability to model complex distributions. By learning to recover clean representations from noisy latent inputs, the generator gains diffusion-like strengths, such as noise robustness and mode coverage, while maintaining the crisp details characteristic of adversarial training. In unsupervised image-to-image translation tasks, LaDiffGAN outperforms conventional GAN baselines and approaches or exceeds pure diffusion models' visual quality while maintaining efficient one-step generation [17].

Diffusion-Assisted GANs demonstrate that incorporating diffusion processes into GANs through adversarial objectives helps to address fundamental challenges in GAN training. Whether smoothing discriminator learning with noise or using diffusion as a guiding mechanism in latent space, these techniques effectively combine strengths from both approaches. The integration of adversarial loss with diffusion stabilizes training dynamics and enhances sample diversity and fidelity. This promising direction suggests that further exploration of adversarially guided diffusion could continue advancing generative modeling capabilities.

4. Comparative Analysis

Hybrid adversarial-diffusion models fall into three distinct categories, each balancing fidelity, speed, and complexity differently. These include distillation-based hybrids that compress diffusion models into fast samplers using adversarial feedback, in-loop adversarial diffusion approaches that integrate discriminators during diffusion training, and diffusion-assisted GANs that incorporate diffusion processes into GAN frameworks. All three categories aim to combine diffusion models' diversity and stability with GANs' sharpness and speed through different architectural and algorithmic trade-offs.

Generation speed versus image fidelity represents a key comparison dimension across hybrid approaches. Distillation-based hybrids prioritize speed by compressing generation to as few as 1-4 denoising steps while maintaining high fidelity. Adversarial Diffusion Distillation achieves comparable quality to its diffusion teacher in just four steps [12], while SDXL-Lightning demonstrates state-of-the-art 1024×1024 image synthesis in 1–8 steps through progressive distillation of Stable Diffusion XL [4]. The adversarial component in these methods significantly enhances visual sharpness in low-step regimes. In contrast, in-loop adversarial approaches vary in focus: Structure-Guided Adversarial Training prioritizes fidelity by enforcing manifold-level realism at standard diffusion sampling speeds [6], and Constant Acceleration Flow emphasizes speed through optimized diffusion dynamics [5]. Diffusion-assisted GANs offer single-pass generation speed while using diffusion-based techniques to enhance image quality. Diffusion-GAN maintains one-step generation but produces more realistic images than standard GANs through diffusion-based noise scheduling [14]. For example, LaDiffGAN operates in one step with improved stylization quality by incorporating latent diffusion guidance [17]. All hybrid approaches leverage adversarial losses to enhance fidelity when generation steps are limited, though distillation and GAN-based methods specifically target minimal-step synthesis for speed.

Training stability versus methodological complexity presents another important consideration. Adversarial objectives introduce minimax dynamics that can destabilize training, requiring different stabilization strategies across approaches. Distillation-based hybrids benefit from pre-trained teacher models that provide strong baseline signals via score or feature distillation, anchoring training even as discriminators fine-tune output distributions. This architecture, especially with progressive distillation, has proven relatively stable. For instance, SDXL-Lightning gradually reduces diffusion steps while carefully designing discriminators to prevent divergence [4]. The trade-off appears in added complexity: training requires optimizing a student model against both teacher and discriminator, incurring substantial computational costs and hyperparameter tuning. In-loop adversarial diffusion methods eliminate the external teacher, simplifying dependencies but potentially complicating training. SADM directly integrates a discriminator into diffusion training to align with real data manifold structures, achieving state-of-the-art fidelity but requiring novel “structure” discriminators and careful training strategies to ensure convergence [6]. Diffusion-assisted GANs address traditional GAN stability challenges through diffusion-based mechanisms. Diffusion-GAN stabilizes discriminator training by gradually increasing noise-effectively creating a curriculum that prevents discriminator overfitting [14]. LaDiffGAN employs pre-trained diffusion models in latent space as supervisory signals, reducing mode collapse and improving convergence [17]. Each hybrid type must navigate adversarial training instability: some leverage existing diffusion models for stability, others modify training procedures through structural losses or noise schedules.

Application suitability varies significantly across hybrid categories depending on deployment priorities. Distillation-based hybrids excel when pre-trained diffusion models are available and low-latency deployment is required. Methods like ADD or SDXL-Lightning demand intensive one-time training but subsequently provide near-instant high-quality generation, making them ideal for real-time synthesis or on-device applications [4; 12]. For the maximum generative quality or adaptation to new domains from scratch, in-loop adversarial diffusion approaches may be the most appropriate. SADM offers superior image quality and cross-domain generalization when fidelity is critical and additional training complexity is acceptable [6]. If the goal is to reduce sampling steps without external teacher models, techniques like CAF provide an efficient path by integrating acceleration directly into diffusion training [5]. Diffusion-assisted GANs offer advantages when instantaneous generation is essential or in situations when classical diffusion models underperform. For tasks like image-to-image translation or small-data generation, GANs with diffusion guidance can outperform pure approaches. LaDiffGAN demonstrates superior stylization while avoiding overfitting on limited data [17], and Diffusion-GAN provides sharper outputs than standard GANs with improved stability [14]. These approaches are particularly fit for interactive applications or video synthesis where generation speed is critical and slight diversity compromises are acceptable.

Table 1

Comparison of hybrid adversarial-diffusion approaches across key criteria

Approach Category	Sample Quality	Generation Speed	Training Stability	Resource Efficiency
Distillation-Based Hybrids (nADD, SDXL-Lightning)	Very high fidelity (near teacher model quality even with few steps)	Very fast sampling (1–4 diffusion steps)	Relatively stable training (teacher guidance with adversarial fine-tuning)	Heavy training cost (requires large pre-trained model and discriminator); light inference load (few steps)
In-Loop Adversarial Diffusion (SADM, CAF)	High fidelity (improved generative quality; SADM achieves state-of-the-art)	Moderate speed (standard diffusion steps; faster if using accelerated flows like CAF)	Moderate stability (minimax training requires careful design; no external teacher)	High training cost (training from scratch with added objectives); inference cost varies (CAF reduces steps, others do not)
Diffusion-Assisted GANs (Diffusion-GAN, LaDiffGAN)	High fidelity for GAN standards (sharper details, better coverage than vanilla GANs)	Instant generation (single-step GAN-like inference)	Improved stability vs. standard GANs (noise or diffusion guidance stabilizes training)	Complex training (GAN + diffusion components); very efficient inference (generator forward pass only)

In summary, combining adversarial loss with diffusion yields diverse hybrid techniques with distinct trade-offs. Distillation-based hybrids deliver diffusion-quality images with drastically reduced sampling time but require complex distillation processes. In-loop adversarial diffusion methods integrate advantages during training without external teachers but demand careful design for stability. Diffusion-assisted GANs achieve the fastest generation and use diffusion to enhance quality and stability, though inheriting GAN training complexities. The selection among these approaches depends on specific priorities: maximum quality, minimum latency, training simplicity, or performance under limited data. Table \ref{tab:comparison} summarizes how these three categories compare across key evaluation criteria.

5. Conclusions and Future Directions

Adversarial-diffusion hybrid models substantially improve generative modeling by leveraging complementary advantages of both approaches [6; 12]. These hybrid architectures achieve better image quality, faster sampling speed, and more robust training dynamics than either method alone. Empirical evidence consistently shows that adversarial components contribute to finer details and enhanced realism, while diffusion elements provide structural coherence and prevent mode collapse. For instance, Sauer et al. demonstrated efficient one-step generation through adversarial distillation without sacrificing quality [3; 12], and Yang et al. reported record-breaking image fidelity by incorporating discriminators into diffusion training pipelines [6].

Current hybrid frameworks fall into three distinct categories with unique integration strategies. Distillation-based approaches compress multi-step diffusion into single-step generators by using pre-trained diffusion models as teachers and incorporating adversarial losses to preserve details [12]. In-loop integration methods embed discriminators directly within the diffusion process, often using noise-conditioned discriminators to guide each denoising step. Diffusion-assisted GANs instead enhance traditional GAN architectures with diffusion components, employing score-based generation for initialization or guidance. Despite methodological differences, all these approaches effectively combine diffusion's noise removal process with the adversarial training mechanism.

The emergence of these hybrid models reveals that diffusion and adversarial objectives serve as complementary rather than competing paradigms. Diffusion models excel at stable density estimation and comprehensive mode coverage, while adversarial training provides powerful feedback for photorealistic details and adaptive learning. The fusion of these properties enables adversarial-diffusion hybrids to achieve results previously unattainable, effectively bridging likelihood-based and adversarial approaches. This convergence suggests a broader trend toward unified generative frameworks that leverage multiple learning objectives simultaneously.

Despite empirical successes, the theoretical foundation for combining adversarial and score-based training remains undeveloped. Most hybrid architectures rely on experimental validation rather than rigorous mathematical justification. The formal reconciliation between diffusion objectives (typically optimizing likelihood or score matching) and GAN objectives (minimizing implicit divergences through minimax games) requires further investigation. Understanding how these different training signals interact (affecting convergence properties, mode coverage, and sample quality) would enable a more principled design of hybrid models. Such theoretical frameworks could guide the development of specialized loss functions that ensure stability when combining adversarial and diffusion criteria.

Balancing image quality, generation speed, and training stability presents a significant challenge, particularly at scale. Current hybrid approaches involve inevitable trade-offs: reducing diffusion steps improves speed but may compromise quality, while optimizing for maximum fidelity often requires more complex discriminators that complicate training. These challenges intensify when scaling to higher resolutions or more complex data modalities like video and 3D content. Early research demonstrates promising results for real-time high-resolution image generation [4] and short video sequences, but generating extended videos or detailed 3D environments at comparable quality levels remains unexplored. Advancing these capabilities will require innovations in both model architecture and training methodologies, which may simultaneously achieve fidelity, speed, and stability.

Extending adversarial-diffusion frameworks beyond image generation represents a promising research direction. Current hybrid models predominantly focus on image synthesis, leaving their applicability to other domains largely unexplored. Adapting these techniques to text or audio generation involves unique challenges due to different data structures and evaluation metrics. While diffusion models have shown success in text-to-image and audio synthesis tasks, integrating adversarial components for these modalities may require fundamentally different approaches. Multimodal generation presents another opportunity, potentially enabling hybrid models that simultaneously generate correlated content across multiple modalities, using diffusion for cross-modal consistency and adversarial feedback for realism enhancement. Addressing these challenges could improve generative performance across diverse domains, combining the best aspects of diffusion and adversarial approaches to create content with superior fidelity, diversity, and generation speed.

References

1. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840–6851. https://doi.org/10.48550/arXiv.2006.11239.

2. Chen, J., Cai, H., Chen, J., Xie, E., Yang, S., Tang, H., ... & Han, S. (2024). Deep compression autoencoder for efficient high-resolution diffusion models. arXiv preprint arXiv:2410.10733. https://doi.org/10.48550/arXiv.2410.10733.

3. Sauer, A., Lorenz, D., Blattmann, A., & Rombach, R. (2023). Adversarial diffusion distillation. arXiv. arXiv preprint arXiv:2311.17042. https://doi.org/10.48550/arXiv.2311.17042.

4. Lin, S., Wang, A., & Yang, X. (2024). Sdxl-lightning: Progressive adversarial diffusion distillation. arXiv preprint arXiv:2402.13929. https://doi.org/10.48550/arXiv.2402.13929.

5. Park, D., Lee, S., Kim, S., Lee, T., Hong, Y., & Kim, H. J. (2024). Constant Acceleration Flow. arXiv preprint arXiv:2411.00322. https://doi.org/10.48550/arXiv.2411.00322.

6. Yang, L., Qian, H., Zhang, Z., Liu, J., & Cui, B. (2024). Structure-guided adversarial training of diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7256–7266). https://doi.org/10.48550/arXiv.2402.17563.

7. Nichol, A. Q., & Dhariwal, P. (2021, July). Improved denoising diffusion probabilistic models. In International conference on machine learning (pp. 8162–8171). PMLR. https://doi.org/10.48550/arXiv.2102.09672.

8. Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456. https://doi.org/10.48550/arXiv.2011.13456.

9. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27. https://doi.org/10.48550/arXiv.1406.2661.

10. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214–223). PMLR. https://doi.org/10.48550/arXiv.1701.07875.

11. Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957. https://doi.org/10.48550/arXiv.1802.05957.

12. Sauer, A., Boesel, F., Dockhorn, T., Blattmann, A., Esser, P., & Rombach, R. (2024, December). Fast high-resolution image synthesis with latent adversarial diffusion distillation. In SIGGRAPH Asia 2024 Conference Papers (pp. 1–11). http://dx.doi.org/10.1145/3680528.3687625.

13. Kang, M., Zhang, R., Barnes, C., Paris, S., Kwak, S., Park, J., ... & Park, T. (2024, September). Distilling diffusion models into conditional gans. In European Conference on Computer Vision (pp. 428–447). Cham: Springer Nature Switzerland. https://doi.org/10.48550/arXiv.2405.05967.

14. Wang, Z., Zheng, H., He, P., Chen, W., & Zhou, M. (2022). Diffusion-gan: Training gans with diffusion. arXiv preprint arXiv:2206.02262. https://doi.org/10.48550/arXiv.2206.02262.

15. Hang, T., Gu, S., Geng, X., & Guo, B. (2024). Improved noise schedule for diffusion training. arXiv preprint arXiv:2407.03297. https://doi.org/10.48550/arXiv.2407.03297.

16. Trinh, L. T., & Hamagami, T. (2024). Latent Denoising Diffusion GAN: Faster sampling, Higher image quality. IEEE Access. http://dx.doi.org/10.1109/ACCESS.2024.3406535.

17. Liu, X., Zeng, B., Gao, S., Li, S., Feng, Y., Li, H., ... & Zhang, B. (2024). Ladiffgan: Training gans with diffusion supervision in latent spaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1115–1125). https://doi.org/10.1109/CVPRW63382.2024.00118.

Дата першого надходження рукопису до видання: 23.05.2025
Дата прийнятого до друку рукопису після рецензування: 20.06.2025
Дата публікації: 25.07.2025

УДК 621.3.037.37
DOI https://doi.org/10.36994/2788-5518-2025-01-09-11

ПЕРЕТВОРЕННЯ БІНОМІАЛЬНИХ ЧИСЕЛ У ДВІЙКОВІ

Борисенко О. А., *д.т.н., професор, Сумський державний університет, Суми, Україна. https://orcid.org/0000-0001-7466-9135, 5352008@ukr.net*

Хацько А. О., *аспірант, Сумський державний університет, Суми, Україна. https://orcid.org/0009-0004-6642-2991, andrey5135vk@gmail.com*

Анотація. *У статті пропонується підхід до перетворення біноміальних чисел у двійкові. Запропоновано ефективний метод, що базується на поетапному декодуванні з одночасним застосуванням принципу повного перебору можливих комбінацій. Суть методу полягає в поступовому наближенні до результату шляхом синхронного аналізу значень біноміальних і двійкових чисел. Зокрема, перебір біноміальних чисел здійснюється в напрямі зменшення їх значень, тоді як відповідні двійкові еквіваленти формуються в напрямі зростання, доки біноміальне представлення не досягне свого мінімального порогового значення. Як наслідок, кінцевим результатом такого поетапного перетворення є точно підібране двійкове число, що відповідає заданому біноміальному значенню. Такий алгоритм перетворення вирізняється своєю логічною прозорістю, обчислювальною простотою та зручністю для практичної реалізації в цифрових системах.*

Особливу увагу в роботі приділено надійності методу. Завдяки внутрішнім властивостям біноміальних чисел, алгоритм демонструє високу стійкість до зовнішніх завад і шумів, що можуть виникати в процесі передачі або зберігання інформації. Однією з вагомих переваг запропонованого підходу є його здатність до самокорекції: навіть за наявності незначних помилок у вхідних даних метод забезпечує точне відновлення двійкового еквівалента, що істотно підвищує загальну надійність системи.

Варто зазначити, що запропонований алгоритм не демонструє надзвичайно високої швидкодії порівняно з деякими іншими методами, однак це компенсується його простотою, технологічною доступністю й низькими вимогами до обчислювальних ресурсів. Така комбінація характеристик робить його надзвичайно зручним для вбудованих систем і пристроїв з обмеженим апаратним забезпеченням. Алгоритм легко інтегрується в наявні цифрові платформи, де перетворення біноміальних чисел у двійкові є важливим складником інформаційної обробки.

Запропонований метод є надійним, ефективним і технологічно доцільним рішенням для широкого спектра застосувань у сфері цифрових комунікацій, обробки сигналів, криптографії та інших галузях, де критично важливими є точність, стійкість до помилок і безперебійна робота в умовах дії зовнішніх завад. Завдяки поєднанню простоти реалізації, адаптивності та здатності до корекції, цей підхід має високий потенціал для практичного впровадження.

Ключові слова: *двійкові числа, біноміальні числа, перебір, завадостійкість, біноміальні коефіцієнти.*

CONVERSION OF BINOMIAL NUMBERS INTO BINARY NUMBERS

Oleksiy Borysenko, *D.Sc. (Tech.), Professor, Sumy State University, Sumy, Ukraine. https://orcid.org/0000-0001-7466-9135, 5352008@ukr.net*

Andrii Khatsko, *PhD Student, Sumy State University, Sumy, Ukraine. https://orcid.org/0009-0004-6642-2991, andrey5135vk@gmail.com*

Abstract. *The article proposes an approach to converting binomial numbers into binary form. An efficient method is introduced, based on step-by-step decoding combined with the simultaneous application of the principle of exhaustive enumeration of possible combinations. The core idea of the method involves gradually approximating the result through synchronized analysis of binomial and binary values. Specifically, the binomial numbers are iterated in decreasing order, while the corresponding binary equivalents are generated in increasing order, until the binomial representation reaches its minimum threshold. As a result, the final outcome of this step-by-step transformation is a precisely matched binary number corresponding to the given binomial value. This conversion algorithm is characterized by logical clarity, computational simplicity, and ease of practical implementation in digital systems.*

Particular attention in the study is given to the reliability of the method. Due to the inherent properties of binomial numbers, the algorithm exhibits high resilience to external interferences and noise that may occur during the transmission or storage of information. One of the significant advantages of the proposed approach is its self-correcting capability: even in the presence of minor errors in the input data, the method

ensures accurate reconstruction of the binary equivalent, which significantly enhances the overall reliability of the system.

It should be noted that the proposed algorithm does not demonstrate exceptionally high speed compared to some existing methods. However, this is more than compensated by its simplicity, technological accessibility, and low computational resource requirements. This combination of characteristics makes it highly suitable for embedded systems and devices with limited hardware capabilities. The algorithm can be easily integrated into existing digital platforms where binomial-to-binary conversion is an essential part of information processing.

The proposed method offers a reliable, efficient, and technologically feasible solution for a wide range of applications in digital communications, signal processing, cryptography, and other fields where accuracy, error resistance, and stable operation under external interference are critically important. Owing to its simplicity, adaptability, and error-correction capabilities, this approach holds significant potential for practical implementation.

Key words: *binary numbers, binomial numbers, enumeration, noise resistance, binomial coefficients.*

Вступ

Більшість сучасних цифрових систем обробки й передачі інформації будуються на основі двійкових чисел, які належать до природної двійкової системи числення. Вона залишається найпростішою завдяки своїй нульовій складності, однак двійкові числа без додаткового кодування не забезпечують можливості виявлення помилок. У зв'язку з цим розроблені складніші позиційні системи числення, зокрема біноміальні, які дають змогу підвищити завадостійкість цифрових систем.

Біноміальні числа, які генеруються в межах біноміальної системи числення, вирізняються високою завадостійкістю завдяки своїй структурі. Однак для практичної реалізації цієї системи часто виникає потреба в перетворенні біноміальних чисел у двійкові, що є важливим для багатьох спеціалізованих цифрових пристроїв, які вимагають підвищеної точності й стійкості до помилок.

Перетворення біноміальних чисел у двійкові, запропоноване в попередніх дослідженнях [1; 2], має значну складність у реалізації, особливо в апаратних рішеннях. Це може знижувати завадостійкість під час обробки даних. Із цієї причини на практиці частіше застосовуються простіші методи, такі як перебір комбінацій, що хоч і поступаються у швидкодії, але забезпечують стабільні результати й легкість інтеграції в системи передачі інформації.

Біноміальні системи числення, зокрема, мають особливе значення для побудови складних комбінаторних структур, які використовуються для кодування, створення рівноважних кодів та інших композицій. Однак для їх ефективного застосування необхідні ефективні перетворювачі кодів із біноміальних чисел у двійкові, що робить розвиток таких алгоритмів актуальним завданням у сучасній цифровій обробці інформації

Існують дві основні задачі перетворення, пов'язані з біноміальними числами, для яких переборні алгоритми є ефективними рішеннями, забезпечуючи високу простоту й надійність. З одного боку, це задача перетворення двійкових чисел у біноміальні, а з іншого – зворотне перетворення біноміальних чисел у двійкові. Оскільки біноміальні числа генеруються за допомогою двійкових біноміальних систем числення, вони мають здатність перебирати, генерувати й нумерувати комбінаторні коди, забезпечуючи перешкодостійкість, що робить їх корисними для передачі, зберігання й обробки інформації в цифрових системах.

Виконання досліджень

Основним завданням роботи є розроблення переборного алгоритму для ефективного перетворення біноміальних чисел у двійкові. Це особливо актуально для практичних систем, де після перетворення двійкових чисел у біноміальні часто виникає потреба виконати зворотне перетворення для коректної роботи цифрових пристроїв. Алгоритм повинен забезпечувати мінімальну складність і надійність, що є критично важливим для стабільної роботи цифрових систем.

Поняття двійкових чисел. Двійкові числа – це числа, які представлені у двійковій системі числення, де використовуються тільки дві цифри: 0 і 1. Двійкова система є основою для роботи сучасних комп'ютерів і цифрових пристроїв. Кожна цифра у двійковому числі називається біт, і вона має вагу, яка визначається позицією цього біта в числі.

Поняття біноміальних чисел. Існує багато різновидностей біноміальних чисел [1]. У робот розглядаються найбільш прості двійкові біноміальні числа.

Двійковими біноміальними числами називаються послідовності розрядів цифр 0 і 1 з вагами у вигляді біноміальних коефіцієнтів, які дають змогу перетворювати їх у номери природних чисел і зворотно.

Нерівномірні біноміальні числа. Існують нерівномірні й отримані на їх основі рівномірні біноміальні числа [1]. Нерівномірним біноміальним числом називається число з двійковим алфавітом довжини n, у якому кількість одиниць не перевищує значення k, а кількість нулів n – k [1]. Вони мають різну довжину й тому можуть ефективно використовуватися для зберігання та передачі інформації. У цьому випадку біноміальні числа меншої довжини кодують більш імовірні повідомлення, а довші –

менш імовірні. Другим їх завданням є те, що вони створюють рівномірні біноміальні числа, які можуть використовуватися для побудови завадостійких цифрових пристроїв [1].

Діапазон нерівномірних і, відповідно. рівномірних біноміальних чисел визначається біноміальним коефіцієнтом:

$$P = C_{n+1}^{k} = C_{n+1}^{n-k} = \frac{(n+1)!}{k!(n-k+1)!}. \quad (1)$$

У таблиці 1 наведені величини діапазонів P нерівномірних біноміальних чисел для n = 1, 2, …, 20 і k = 2.

Таблиця 1

Діапазони біноміальних чисел P

n	P	n	P	n	P	n	P
1	1	6	21	11	66	16	136
2	3	7	28	12	78	17	153
3	6	8	36	13	91	18	171
4	10	9	45	14	105	19	190
5	15	10	55	15	120	20	210

Алгоритм побудови нерівномірних біноміальних чисел у зростаючому порядку з будь-якого біноміального числа складається з таких кроків:

1. В останній розряд справа нерівномірного біноміального числа, у якому знаходиться 0, записується 1 і підраховується кількість одиниць у ньому.

2. Якщо кількість одиниць у біноміальному числі менша за k, то справа від отриманої одиниці записується нуль і далі відбувається повернення до кроку 1.

3. Якщо кількість одиниць у біноміальному числі дорівнює k, а нулів перед ними немає, то – кінець.

4. Якщо кількість одиниць у біноміальному числі буде дорівнювати k і перед останніми з них буде стояти хоча б один нуль, то цей 0 повинен бути замінений на 1, а в молодші стосовно нього справа розряди записуються нулі до того часу, поки їх кількість у числі не стане дорівнювати $n – k$.

5. Повернення до кроку 1.

У таблиці 2 для n = 4 наведені 10 можливих нерівномірних біноміальних чисел зліва з кількістю одиниць k = 3, а справа – з k = 2.

Таблиця 2

Нерівномірні біноміальні числа з n = 4 і k = 3,2

№	Бін. числа	№	Бін. числа	№	Бін. числа	№	Бін. числа
0	00	5	1010	0	000	5	011
1	010	6	1011	1	0010	6	1000
2	0110	7	1100	2	0011	7	1001
3	0111	8	1101	3	0100	8	101
4	100	9	111	4	0101	9	11

Рівномірні біноміальні числа. Окрім нерівномірних біноміальних чисел, при побудові біноміальних цифрових пристроїв використовуються рівномірні біноміальні числа. Вони, на відміну від нерівномірних чисел, мають постійну довжину (див. таблицю 3), яка містить n розрядів. Їх особливістю є те, що кількість нулів у них перед останньою одиницею справа не може бути більшою $n – k$, що призводить до їх завадостійкості [1]. Тобто побудовані на основі рівномірних біноміальних чисел цифрові пристрої чи програми будуть завадостійкі.

Рішення задачі. Кількісний еквівалент числа n – розрядної двійкової k-біноміальної системи числення:

$$A_i = (a_{j-1},\ a_{j-2},...,a_0)$$

$$i = 0,1,...,P-1$$

визначається числовою функцією:

$$A_i = a_{j-1}C_{n-1}^{k-q_j} + ... + a_l C_{n-j+l}^{k-q_{l+1}} + ... + a_0 C_{n-j}^{k-q_1} \quad (2)$$

Вона має дві системи обмежень:

або

$$\begin{cases} q_0 = k \\ k \le j \le n-1 \\ a_0 = 1 \end{cases}$$

або

$$\begin{cases} n-k = j-q_0 \\ 0 \le q_0 \le k-1 \\ a_0 = 0, \end{cases}$$

де q_0 – кількість одиниць;
j – кількість розрядів;
l = 0, 1, ..., j –1 – порядковий номер розряду;

Сума одиничних значень цифр:

$$q_{l+1} = \sum_{\gamma=l+1}^{j-1} a_\gamma,$$

максимальне число при цьому:

$$1C_{n-1}^{k-q_j} + 1C_{n-2}^{k-q_j-1} + \ldots + 1C_{n-j}^{k-q_1} = C_n^k - 1 = P - 1$$

Відповідно, кількість біноміальних чисел:

$$P = C_n^k$$

Як випливає з формули (1), ці числа нерівномірні, тобто можуть відрізнятися своєю довжиною. Щоб отримати відповідний біноміальному числу двійковий номер, треба підставити це число у формулу (2).

У таблиці 1 ці числа з їх кількісними еквівалентами і двійковими номерами показані для k = 4, n = 6. Їх кількість, відповідно, дорівнює:

$$P = C_n^k = C_6^4 = 15.$$

Таблиця 3

Нерівномірні біноміальні числа та їх кількісні еквіваленти

№	Біноміальне число	Кількісний еквівалент	№
0	00	$0 \cdot C_5^4 + 0 \cdot C_4^4$	0000
1	010	$0 \cdot C_5^4 + 1 \cdot C_4^4 + 0 \cdot C_3^3$	0001
2	0110	$0 \cdot C_5^4 + 1 \cdot C_4^4 + 1 \cdot C_3^3 + 0 \cdot C_2^2$	0010
3	01110	$0 \cdot C_5^4 + 1 \cdot C_4^4 + 1 \cdot C_3^3 + 1 \cdot C_2^2 + 0 \cdot C_1^1$	0011
4	01111	$0 \cdot C_5^4 + 1 \cdot C_4^4 + 1 \cdot C_3^3 + 1 \cdot C_2^2 + 1 \cdot C_1^1$	0100
5	100	$1 \cdot C_5^4 + 0 \cdot C_4^3 + 0 \cdot C_3^3$	0101
6	1010	$1 \cdot C_5^4 + 0 \cdot C_4^3 + 1 \cdot C_3^3 + 0 \cdot C_2^2$	0110
7	10110	$1 \cdot C_5^4 + 0 \cdot C_4^3 + 1 \cdot C_3^3 + 1 \cdot C_2^2 + 0 \cdot C_1^1$	0111
8	10111	$1 \cdot C_5^4 + 0 \cdot C_4^3 + 1 \cdot C_3^3 + 1 \cdot C_2^2 + 1 \cdot C_1^1$	1000
9	1100	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 0 \cdot C_3^2 + 0 \cdot C_2^2$	1001
10	11010	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 0 \cdot C_3^2 + 1 \cdot C_2^2 + 0 \cdot C_1^1$	1010
11	11011	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 0 \cdot C_3^2 + 1 \cdot C_2^2 + 1 \cdot C_1^1$	1011
12	11100	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 1 \cdot C_3^2 + 0 \cdot C_2^1 + 0 \cdot C_1^1$	1100
13	11101	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 1 \cdot C_3^2 + 0 \cdot C_2^1 + 1 \cdot C_1^1$	1101
14	1111	$1 \cdot C_5^4 + 1 \cdot C_4^3 + 1 \cdot C_3^2 + 1 \cdot C_2^1$	1111

Відповідно до (2), алгоритм перетворення нерівномірних біноміальних чисел у двійкові числа складається з такого:

1. З їх підстановки у формулу 1.
2. Обчислення біноміальних коефіцієнтів у розрядах, де стоять 1.
3. Підсумовування отриманих значень біноміальних коефіцієнтів.

Результат буде номером нерівномірного біноміального числа.

Доповнюючи в таблиці 3 розряди справа нулями, можемо отримати рівномірні біноміальні числа, як це показано в таблиці 4. Їх кількісний еквівалент залишиться незмінним.

Таблиця 4

Рівномірні біноміальні числа та їх кількісні еквіваленти

№	Біноміальне число	Рівн. біном. число	Кількісний еквівалент	№
0	00	00000	$0\cdot C_5^4+0\cdot C_4^4$	0000
1	010	01000	$0\cdot C_5^4+1\cdot C_4^4+0\cdot C_3^3$	0001
2	0110	01100	$0\cdot C_5^4+1\cdot C_4^4+1\cdot C_3^3+0\cdot C_2^2$	0010
3	01110	01110	$0\cdot C_5^4+1\cdot C_4^4+1\cdot C_3^3+1\cdot C_2^2+0\cdot C_1^1$	0011
4	01111	01111	$0\cdot C_5^4+1\cdot C_4^4+1\cdot C_3^3+1\cdot C_2^2+1\cdot C_1^1$	0100
5	100	10000	$1\cdot C_5^4+0\cdot C_4^3+0\cdot C_3^3$	0101
6	1010	10100	$1\cdot C_5^4+0\cdot C_4^3+1\cdot C_3^3+0\cdot C_2^2$	0110
7	10110	10110	$1\cdot C_5^4+0\cdot C_4^3+1\cdot C_3^3+1\cdot C_2^2+0\cdot C_1^1$	0111
8	10111	10111	$1\cdot C_5^4+0\cdot C_4^3+1\cdot C_3^3+1\cdot C_2^2+1\cdot C_1^1$	1000
9	1100	11000	$1\cdot C_5^4+1\cdot C_4^3+0\cdot C_3^2+0\cdot C_2^2$	1001
10	11010	11010	$1\cdot C_5^4+1\cdot C_4^3+0\cdot C_3^2+1\cdot C_2^2+0\cdot C_1^1$	1010
11	11011	11011	$1\cdot C_5^4+1\cdot C_4^3+0\cdot C_3^2+1\cdot C_2^2+1\cdot C_1^1$	1011
12	11100	11100	$1\cdot C_5^4+1\cdot C_4^3+1\cdot C_3^2+0\cdot C_2^1+0\cdot C_1^1$	1100
13	11101	11101	$1\cdot C_5^4+1\cdot C_4^3+1\cdot C_3^2+0\cdot C_2^1+1\cdot C_1^1$	1101
14	1111	11110	$1\cdot C_5^4+1\cdot C_4^3+1\cdot C_3^2+1\cdot C_2^1$	1111

Особливістю нерівномірних біноміальних чисел є те, що вони можуть бути використані для швидкої передачі інформації. Однак вони не можуть виявляти помилки, які виникають при їх передачі. Цю задачу успішно виконують рівномірні біноміальні числа. Ознакою помилки в них буде поява в числі кількості одиниць більшої за k або поява перед останньою 1 зліва більше n – k – 1. Таких помилкових біноміальних чисел для n = 5 буде 17. Таким числом, наприклад, буде число 10010 або число 11111.

Методи перетворення біноміальних чисел у двійкові. На практиці для використання двійкових чисел потрібні перетворювачі біноміальних чисел у двійкові, оскільки цифрові системи, як правило, працюють із двійковими значеннями. Загальний вигляд процесу перетворення біноміального числа у двійкове число зображено на рис. 1.

Рис. 1. Перетворення біноміальних чисел у двійкові

У ньому біноміальне число подається на перетворювач кодів, який залежно від обраного алгоритму виконує перетворення з різною швидкодією.

Найбільш швидкодіючим буде табличний метод перетворення, за якого в пам'яті перетворювача зберігається готовий результат, а вхідне біноміальне число виступає як адреса. За цією адресою можна отримати відповідне двійкове число. Однак недоліком цього алгоритму є потреба у великій ємності пам'яті, що обмежує кількість біноміальних чисел, які можуть зберігатися.

Менш швидкодіючим, але з достатньою швидкодією й без великих вимог до пам'яті є алгоритм, який використовує числову функцію для біноміальних чисел. Проте його реалізація може бути досить складною, особливо в схемному вигляді:

$$C(n,k) = \frac{n!}{k!(n-k)!}.$$

Тобто алгоритм отримує значення n і k, обчислює відповідне значення біноміального коефіцієнта, а потім перетворює його у двійкову форму. Це дає змогу уникнути використання великої пам'яті, оскільки результат обчислюється безпосередньо під час запиту й не потрібно зберігати таблицю всіх можливих результатів.

Однак реалізація цього методу є складнішою, особливо якщо алгоритм має бути апаратно реалізованим, наприклад, у цифровій схемі. Потрібно враховувати точність розрахунків факторіалів і правильно обробляти великі числа, що може ускладнити розробку й вплинути на загальну швидкодію.

Найменш швидкодіючим, але простим і надійним методом є переборний алгоритм, який використовує програмно або апаратно реалізований біноміальний лічильник. Цей лічильник підсумовує одиниці, а разом із ним працює двійковий лічильник, який зменшує своє значення. Система порівняння перевіряє, чи досягнув двійковий лічильник нульового стану.

Процес роботи переборного алгоритму:

Ініціалізація – біноміальний лічильник починається зі значення, яке потрібно перетворити, а двійковий лічильник ініціалізується на 0.

Перебір – біноміальний лічильник послідовно підсумовує одиниці, тоді як двійковий лічильник зменшує своє значення.

Порівняння – процес триває, поки двійковий лічильник не досягне нульового значення, після чого схема порівняння видає сигнал.

Виведення результату – коли двійковий лічильник переходить у нульовий стан, значення двійкового лічильника буде дорівнювати результату перетворення біноміального числа.

Висновок

Таким чином, надані в статті переборні алгоритми перетворення біноміальних чисел у двійкові є найбільш простими й надійними. Незважаючи на певний недолік у вигляді малої швидкодії, для багатьох практичних застосувань ця швидкодія є цілком достатньою, що робить ці алгоритми ефективними в таких випадках. Ефективність перетворення зростає завдяки врахуванню завадостійкості біноміальних чисел, які за своєю природою володіють властивістю стійкості до помилок, що дає їм змогу забезпечувати надійність навіть в умовах перешкод.

Література

1. Борисенко О. А. Дискретна математика : підручник. Суми, 2019. 255 с.
2. Бардачов Ю., Соколова Н., Ходаков В. Дискретна математика. Київ : Вища шк., 2002. 287 с.
3. Бондаренко М., Білоус Н., Руткас А. Комп'ютерна дискретна математика. Київ : Компанія СМІТ, 2004. 480 с.
4. Капітонова Ю., Кривий С., Летичевський О. Основи дискретної математики. Київ : Наук. думка, 2002. 580 с.

References

1. Borysenko, O. A. (2019) Dyskretna matematyka : pidruchnyk [Discrete Mathematics], 255 s. [in Ukrainian].
2. Bardachov, Yu., Sokolova, N., Khodakov, V. (2002) Dyskretna matematyka [Discrete Mathematic]. 287 s. [in Ukrainian].
3. Bondarenko, M., Bilous, N., Rutkas, A. (2004) Komp'iuterna dyskretna matematyka [Computer Discrete Mathematics]. 480 s. [in Ukrainian].
4. Kapitonova, Yu., Kryvyi, S., Letychevskyi, O. (2002) Osnovy dyskretnoi matematyky [Fundamentals of Discrete Mathematics]. 580 s. [in Ukrainian].

Дата першого надходження рукопису до видання: 15.05.2025
Дата прийнятого до друку рукопису після рецензування: 10.06.2025
Дата публікації: 25.07.2025

UDC 004.738.5
DOI https://doi.org/10.36994/2788-5518-2025-01-09-12

INVESTIGATING THE PERFORMANCE IMPACT OF LAZY LOADING IN WEB APPLICATIONS

Borovskova Ye. A., *master's degree, Software Engineer, AppsFlyer Ltd, Kyiv, Ukraine. https://orcid.org/0009-0008-5481-8090*

Abstract. *The study's relevance is due to the growing performance requirements of modern web applications that process significant amounts of data and provide a high user experience. Under these conditions, lazy loading is considered an effective approach to optimise resource utilisation, reduce response times and reduce the load on network and computing resources. At the same time, it has been established that, despite the obvious advantages, there are limitations associated with implementing this technology that require a detailed analysis.*

The study's purpose is to assess the impact of lazy loading on key performance indicators of web applications, identify its advantages and limitations, and develop recommendations for the effective implementation of this technology.

The study applied the methodology of experimental modelling of web applications in different scenarios. Two versions of the application were developed: one without lazy loading and the other with its implementation. The experiments were conducted in networks with different bandwidths (4G, 3G), and performance was evaluated based on indicators such as page load time, RAM consumption, and network resources.

The study found that lazy loading significantly reduces loading time (up to 50% on a 4G network) and reduces the amount of data transferred by almost 3.5 times. This significantly improves the user experience and optimises the use of server resources. At the same time, a number of limitations have been identified, such as dependence on the stability of the network connection, complexity of implementation, and potential problems with SEO indexing of dynamic content.

The conclusions emphasise the need to consider the application's specifics and users' needs when implementing lazy loading. In particular, it is important to ensure compatibility with search engines, adapt to weak network conditions, and consider accessibility requirements for users with disabilities.

Prospects for further research are identified in the development of approaches to optimise SEO indexing of dynamic content, ensure accessibility for all user categories, and integrate lazy loading into systems that handle critical data. Another interesting area is the analysis of device power consumption when using this technology.

Key words: *lazy loading; web applications; performance; optimization; dynamic loading; loading time; memory consumption; network resources; SEO; accessibility.*

ДОСЛІДЖЕННЯ ВПЛИВУ LAZY LOADING НА ПРОДУКТИВНІСТЬ ВЕБ-ЗАСТОСУНКІВ

Євгенія Боровскова, *магістр, інженер програмного забезпечення, AppsFlyer Ltd, Київ, Україна. https://orcid.org/0009-0008-5481-8090*

Анотація. *Актуальність дослідження зумовлена все більшими вимогами до продуктивності сучасних вебзастосунків, які опрацьовують значні обсяги даних і забезпечують високий рівень користувацького досвіду. У цих умовах використання lazy loading розглядається як дієвий підхід до оптимізації завантаження ресурсів, скорочення часу відгуку та зниження навантаження на мережеві й обчислювальні ресурси. Водночас установлено, що, попри очевидні переваги, упровадження цієї технології має низку обмежень, які потребують детального аналізу.*

Метою роботи є оцінювання впливу lazy loading на ключові показники продуктивності вебзастосунків, визначення його переваг та обмежень, а також розроблення рекомендацій щодо ефективного впровадження цієї технології.

Дослідження проводилося із застосуванням методології експериментального моделювання роботи вебзастосунків у різних сценаріях. Створено дві версії застосунку: одну – без використання lazy loading, а іншу – з його впровадженням. Експерименти проводили за умов мереж із різною пропускною здатністю (4G, 3G), а продуктивність оцінювали на основі таких показників, як час завантаження сторінок, використання оперативної пам'яті та споживання мережевих ресурсів.

У результаті дослідження встановлено, що lazy loading забезпечує значне скорочення часу завантаження сторінок (до 50% у мережі 4G) і зменшення обсягу переданих даних майже в 3,5 раза.

Це значно покращує користувацький досвід, оптимізує використання серверних ресурсів. Водночас виявлено низку обмежень, серед яких – залежність від стабільності мережевого з'єднання, складність реалізації та потенційні труднощі із SEO-індексацією динамічного контенту.

У висновках наголошено на важливості врахування специфіки застосунку й потреб користувачів під час упровадження lazy loading. Зокрема, значущим є забезпечення сумісності з пошуковими системами, адаптація до умов слабких мереж і врахування вимог доступності для користувачів з обмеженими можливостями.

Перспективи подальших досліджень включають розроблення підходів для оптимізації SEO-індексації динамічного контенту, забезпечення доступності для всіх категорій користувачів та інтеграції lazy loading у системи, що працюють із критичними даними. Також важливим напрямом є аналіз енергоспоживання пристроїв під час використання цієї технології.

Ключові слова: *lazy loading; вебзастосунки; продуктивність; оптимізація; динамічне завантаження; час завантаження; споживання пам'яті; мережеві ресурси; SEO; доступність.*

Introduction

Web applications that serve many users face the challenge of efficiently managing resources, lowering page loading speeds, and providing a quality user experience. As the amount of data loaded initially increases, difficulties arise due to long response times, network congestion, and low performance of users' devices. One of the most promising approaches to solving these problems is using lazy loading technology, which allows you to optimize the downloading process by delaying its loading.

This problem is of great scientific importance, as it requires research into the impact of different loading strategies on the performance of web applications, considering the specifics of modern frameworks, browsers, and network protocols. The practical significance lies in increasing the efficiency of web applications, reducing the load on servers, and improving the user experience, which is critical for developers, online service providers, and end users.

The topic of researching the impact of lazy loading on web application performance is actively developing. Many scientific works focus on analyzing the principles of this technology, evaluating its effectiveness, and practical approaches to implementation.

The works of R.-M. Bâra, C. A. Boiangiu, and C. Tudose investigate the impact of lazy loading on key performance indicators of web applications, including page load time and memory usage. The authors consider implementation scenarios and offer recommendations for their optimization [1]. Similar aspects are analyzed by S. Turcotte, Gokhale, and F. Tip, who studied how lazy loading can increase the responsiveness of real-time web applications by reducing delays [2].

In their work, E. Mjelde and A. L. Opdahl investigate methods for reducing the loading time of web applications on different devices, focusing on devices with limited resources. Their analysis covers lazy loading as one of the key elements of web application performance [3]. S. K. Shivakumar takes a closer look at the optimization of mobile web applications, particularly the impact of lazy loading on saving network resources and reducing data volume [4].

The study by F. Yan, Z. Xu, and their colleagues aims to optimise front-end development performance. The authors propose a solution to improve the efficiency of dynamic web applications using lazy loading [5]. Wickham M. focuses on the practical aspects of implementing lazy loading for images, which is critical for reducing the loading time of large pages [6].

I. Shvedov investigates the feasibility of using lazy loading in web applications, emphasising its role in improving speed and reducing computing resources [7]. I. Khlysta and A. Bondarchuk consider the problem of partial content updating in SPA applications, using lazy loading as a key approach [8].

A. V. Antonenko and his colleagues describe innovative methods of implementing this technology in their work. The authors analyze the peculiarities of displaying information in web browsers and the role of lazy loading in reducing server load [9].

M. Hajian studies the use of lazy loading in the performance of Angular applications, focusing on improving the performance of large dynamic systems [10]. O. Kostenko and V. Furashev investigate the regulatory aspects of web space development, including technological challenges and their impact on users [11].

K. Chyzhmar and his co-authors analyse information security in the context of web application development, paying attention to the role of optimisation, in particular lazy loading [12]. F. Al-Hawari, in his work, considers design patterns for data management in web systems, highlighting lazy loading as an element of big data management [13].

In his book about progressive web applications, S. Domes demonstrates practical cases of using lazy loading in React applications to achieve loading speed and interactivity [14].

Thus, the analyzed studies demonstrate significant interest in implementing lazy loading as a key tool for optimizing web applications.

Despite studying the principles of lazy loading and its impact on performance, several key aspects still need to be solved. The conditions for using this technology in real-world web applications with complex

architectures require further research, especially in cases with high server loads or poor network quality (for example, in 3G networks). A limited number of studies analyze in detail how lazy loading affects the power consumption of mobile devices or performance in multimodule systems.

Special attention is drawn to the problems of SEO indexing dynamic content. Search engines are often unable to fully index pages with delayed loading, which can negatively affect websites' visibility in search results. Approaches to ensuring the accessibility of dynamic content for users with disabilities are also insufficiently developed, which is critical in terms of meeting modern standards of inclusiveness.

The proposed research will contribute to solving these problems. Experimental analysis will allow us to assess the impact of lazy loading in complex practical scenarios, particularly in conditions of low network speeds. It will also develop recommendations for optimizing SEO indexing of dynamic content and ensuring accessibility, which will help adapt the technology to modern web development needs. This will not only broaden the understanding of lazy loading capabilities but also help eliminate existing limitations for its effective implementation.

The article aims to investigate the impact of lazy loading technology on web application performance, identify its advantages and limitations, and provide practical recommendations for implementation to optimize the user experience and resource efficiency.

Objectives of the article:

1. To investigate the principles of lazy loading and assess its impact on the performance of web applications, including page load time, memory, and network resource consumption.
2. To conduct an experimental study of the implementation of lazy loading in various web application scenarios to identify the main advantages and limitations of its use in practical conditions.
3. To develop recommendations for the effective implementation of lazy loading in web applications based on the analysis of the results obtained and the identified limitations.

Research results

Lazy loading is an important approach in web development that allows you to optimize application performance by loading resources only when needed. The main principle of this technology is to reduce the initial load on the system by deferring the loading of modules, images, videos, or other components that are needed only at a specific moment of the user's work. This approach aims to improve page loading speeds, reduce network resource use, and optimize RAM consumption. In web applications that often contain large data or have a complex structure, lazy loading is a key tool for ensuring high performance (Table 1).

In today's environment, lazy loading is widely used to solve several problems related to the performance of web applications. For example, multi-page websites like e-commerce platforms avoid loading all pages simultaneously. Only those pages that the user accesses are loaded during the session. Also, delayed loading significantly reduces system response time on media platforms with a lot of video content.

Using attributes such as loading='lazy' allows the browser to control the loading process automatically. For example, in static websites, images or iframes are loaded only when they are in the user's visible area. This significantly reduces the consumption of network resources, especially on mobile devices.

Table 1

Basic principles of lazy loading in web applications

How it works	Explanation
Deferred loading of modules	Modules are loaded dynamically only when the user goes to a specific section or page.
Upload images on demand	Images are loaded only when the user scrolls to the appropriate area.
Using special attributes	The loading='lazy' HTML attribute allows the browser to automatically implement delayed image loading.
Dynamically connect scripts and styles	The download is performed using JavaScript depending on the user's actions.

In frameworks such as Angular, the lazy loading principle is implemented through modular connection mechanisms. This means that instead of loading the entire application on the first visit, only those modules that correspond to the functionality selected by the user are loaded. In React, similar functionality is achieved using libraries such as React-lazy and Suspense, which simplify the implementation of deferred component loading. As a result, the user gets faster access to content, load times are reduced, and developers ensure more efficient use of server resources. This approach is ideal for optimizing multifunctional platforms and websites that target a global audience.

As an approach to optimizing web applications, lazy loading affects key performance indicators, including page load time, RAM consumption, and network resource usage. This method reduces the load on the system by delaying the loading of content or modules only when needed. As a result, resources are optimized, the user experience is improved, and load times are reduced. Page load time is a critical metric because it affects the first impression of the user, as well as the overall performance of the application.

Using lazy loading allows you to download only the necessary data, significantly reducing the amount of traffic, especially in conditions of weak network connections (Table 2).

Today, lazy loading is proving to be effective in various industries, such as e-commerce, education, media, and entertainment. For example, in e-commerce, stores that use lazy loading of product images ensure that the first screen is displayed quickly, even if the catalog contains thousands of items. The user sees relevant content without delay, which increases conversion and reduces bounce rates [2]. In educational platforms that include interactive tasks or multimedia content, lazy loading allows you to load only active modules. This provides quick access to learning materials, reducing memory and time consumption. For example, video tutorials in eLearning platforms load only when the corresponding module is opened.

In media applications that focus on streaming video or displaying galleries, delayed loading allows you to optimize your traffic consumption. This is critical for mobile users with limited data. It allows consumers to get quick access to the content they need, avoiding network or device overload.

Table 2

Impact of lazy loading on key web application performance indicators

Performance indicator	**Explanation of the impact without Lazy Loading**	**Explanation of the impact with Lazy Loading**
Page load time	All resources are loaded simultaneously, which leads to longer loading times, especially with large amounts of content and multimedia data.	Only the necessary elements are loaded at the time of use, which significantly reduces loading time, especially for the first screen and on weak devices.
Amount of data transferred	The need to load all resources (images, videos, modules) regardless of their relevance to the current user session.	Only the data that is needed at a particular moment is transmitted, reducing server load and traffic consumption.
RAM usage	High memory requirements due to the simultaneous loading of all modules, which can cause delays or crashes on low performance devices.	Reduced load due to gradual data loading, which optimises memory usage and reduces the risk of overloading devices.
Network performance	Significant delays due to the transfer of large amounts of data in a single session, which is especially problematic for low-speed networks or limited mobile connection resources.	Optimised loading reduces latency, improving the user experience even on low-bandwidth networks.
Power consumption of devices	Increased power consumption due to high processor and memory requirements, which reduces the battery life of mobile devices.	Rational energy consumption by reducing the amount of computing and optimising the use of resources.

Lazy loading is becoming an important tool in improving the performance of web applications, allowing not only to adapt the operation of applications to the needs of users, but also to ensure the efficient use of computing and network resources. Its implementation is strategically important for modern developments focused on interactivity and a high level of user experience [1].

An experimental study of the implementation of lazy loading in web applications was conducted to assess its impact on key performance indicators in various scenarios. The main goal of the experiment was to analyze the loading time, RAM, and network resource efficiency in a web application environment, particularly in the example of multi-page websites built on the Angular framework.

The experimental conditions included simulation of a web application with 10 pages, where the probability of visiting all pages during one session is low. For comparison, two versions of the application were created: without using lazy loading and with its implementation. The metrics used were the initial amount of data downloaded, the time to first page display, and resource consumption under different network conditions (4G and 3G).

The use of lazy loading in this experiment was implemented through the mechanism of dynamic module loading using Angular. The code for routing was as follows: export const routes: Routes = [{ path: 'user', loadChildren: () => import('./pages/user/user.module').then(m => m.UserModule) }];

This approach allows you to download only those modules the user needs at a particular moment, avoiding system overload (Table 3).

The analysis of the results showed that lazy loading significantly reduces the time it takes to load the first screen and optimizes RAM consumption. For example, in a 4G environment, an application with lazy loading achieved a load time of 1 second, while without this approach, the time was about 3 seconds. On a 3G network, the results were even more impressive, showing a reduction in time from 5 to 1.5 seconds. In addition, the amount of downloaded data was reduced by almost 3.5 times, which helped reduce the server and network infrastructure load.

The experiment's practical results prove that lazy loading is an effective tool for optimising multi-page web applications. It provides faster access to content for users with different levels of access to resources, increasing the system's overall efficiency and adaptability. This approach is recommended for use in applications targeting a global audience, especially in a mobile environment with limited capabilities [10].

Table 3

Comparison of web application performance with and without lazy loading

Scenario of use	No Lazy Loading	With Lazy Loading
The initial volume of data	~2.5 MB	~700 KB
Download time (4G)	~3 seconds	~1 second
Download time (3G)	~5 seconds	~1.5 seconds
RAM usage	High due to simultaneous loading	Low due to the step-by-step loading process
User experience	Low due to long waiting times	High due to quick access

Lazy loading is a key web application optimization tool that provides significant performance benefits. However, its implementation is accompanied by several limitations that must be considered when implementing it in practice. This technology helps to improve the efficiency of web applications, especially multi-page platforms, by dynamically loading modules, resources, or content only when needed. This reduces the server load and optimizes the speed of displaying the first screens.

Among the main limitations of lazy loading is the increased complexity of implementation. To integrate the technology, developers must adapt the application architecture, ensuring the correct routing and management of dynamic module loading [6]. This can increase development time and costs, especially in large-scale projects with many components. There may also be difficulties in ensuring technology compatibility with different frameworks or third-party libraries, which requires additional resources for testing and debugging.

Another important limitation is the dependence on the stability of the network connection. Although lazy loading optimizes traffic consumption, real-time resource loading can create delays if the user's connection is slow or unstable. This is especially true for multimedia content, such as videos or large images, where delays can negatively impact the user experience.

Another limitation is the risk of deteriorating SEO (search engine optimization) performance [14]. Since most search engines index the content of pages during the first load, lazy loading can result in some dynamic content remaining unavailable for indexing. This can affect your site's search engine rankings, especially if the content is key to driving traffic.

You should also consider potential accessibility issues for users with disabilities. For example, dynamic loading may not consider specific requirements such as keyboard navigation or screen readers. This can create additional barriers for such users, reducing the accessibility of the web application.

Effective implementation of lazy loading in real-world web projects requires careful planning and adherence to several practical recommendations to ensure maximum performance and user experience. One key aspect is a preliminary analysis of the web application architecture to determine the most resource-intensive modules or components that should be loaded dynamically. This requires testing with performance monitoring tools, such as Lighthouse or WebPageTest, to identify bottlenecks in the content loading process.

An important step is to properly organize the routing. In multi-page applications, modules should be structured in such a way that only the necessary components are loaded according to the route selected by the user. This can be implemented through modern frameworks, such as Angular or React, which support dynamic import mechanisms. In particular, it is useful for Angular to use a modular loading mechanism, where each module is connected only when the corresponding route is activated. This approach minimizes the initial amount of uploaded data and reduces the server load.

To optimize multimedia content, such as images or videos, it is recommended to use built-in browser attributes, such as loading='lazy'` which allow you to automate the process of delayed loading. This is especially effective in cases where a significant part of the page contains large files that can affect the first rendering time. Additionally, for interactive elements such as videos or iframes, you need to take into account their impact on the loading time and ensure that they load only after user interaction.

Particular attention should be paid to optimisation for mobile devices that have limited resources. In this case, it is necessary to test the performance of lazy loading on low-speed networks, such as 3G, to ensure stable and fast loading even under difficult connection conditions. It is also recommended that backup loading mechanisms be implemented in the event of a failure of the main scenario [3].

An additional factor in successful implementation is ensuring compatibility with search engines. To do this, it is necessary to adhere to standards that guarantee the correct indexing of dynamic content. One of these approaches is the use of server-side rendering (SSR) [11], which ensures that the underlying content is loaded on the server, making it available to search engines even before lazy loading is activated.

To summarise, the effective implementation of lazy loading depends on the balance between performance, user experience, and technical capabilities of the application. It requires taking into account the architectural features of the project, testing in real conditions, and integrating modern performance monitoring tools. With the right approach, lazy loading becomes a powerful tool for optimizing web applications, ensuring their adaptability to various use cases.

Conclusions

The study of lazy loading's impact on web application performance has revealed its significant advantages, including optimizing page loading time, reducing the number of resources used, and improving the user experience. This technology is particularly effective for multi-page applications and applications that operate with large amounts of multimedia content, allowing for adaptability to different network environments. Experiments have confirmed that delayed loading reduces the amount of data required for the initial download and reduces the load on servers.

At the same time, several problems have been identified that limit the implementation of lazy loading in practical conditions. The main ones are the increased complexity of the technology implementation, dependence on the stability of the network connection, and potential risks of deterioration of SEO optimization due to dynamic content. In addition, the article identifies accessibility issues for users with disabilities that arise due to dynamic loading of resources.

Recommendations for implementing lazy loading in real projects include a thorough analysis of the application architecture, using modern frameworks for dynamic module connection, and adapting the technology to the conditions of low-speed networks. An important component is to ensure compatibility with search engines using techniques such as server-side rendering. To minimize risks, comprehensive testing at all stages of development is recommended.

Prospects for further research include a deeper analysis of methods for ensuring accessibility for all categories of users, the development of adaptive approaches to SEO indexing of dynamic content, and the study of the cost-effectiveness of implementing lazy loading in projects of different scales and specifics. Particular attention should be paid to integrating this technology into systems that handle critical data and analyzing its impact on the power consumption of modern mobile devices.

Bibliography

1. Bâra R.-M., Boiangiu C. A., Tudose C. Analysing the performance impacts of lazy loading in web applications. *Journal of Information Systems & Operations Management.* 2024. Vol. 18, № 1. P. 1–15. URL: https://web.rau.ro/websites/jisom/Vol.18%20No.1%20-%202024/JISOM%2018.1.pdf#page=9 (date of access: 13.12.2024).

2. Turcotte S., Gokhale, Tip F. Increasing the Responsiveness of Web Applications by Introducing Lazy Loading. *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, Luxembourg, Luxembourg, 2023. P. 459–470. DOI: https://doi.org/10.1109/ASE56229.2023.00192 (date of access: 13.12.2024).

3. Mjelde E., Opdahl A. L. Load-Time Reduction Techniques for Device-Agnostic Web Sites. *Journal of Web Engineering.* 2017. Vol. 16, № 3–4. P. 311–346. URL: https://journals.riverpublishers.com/index.php/JWE/article/view/3291 (date of access: 13.12.2024).

4. Shivakumar S. K. Mobile Web Performance Optimization. *Modern Web Performance Optimization: Methods, Tools, and Patterns to Speed Up Digital Platforms.* Apress, Berkeley, CA. 2020. P. 79–103. DOI: https://doi.org/10.1007/978-1-4842-6528-4_4 (date of access: 13.12.2024).

5. Yan F., Xu Z., Zhong Y. S., HaiBei Z., Ge C. Y. Research on Performance Optimization Scheme for Web Front-End and Its Practice. *Liang Q., Wang W., Liu X., Na Z., Li X., Zhang B. (eds) Communications, Signal Processing, and Systems. Lecture Notes in Electrical Engineering.* Springer, Singapore, 2021. Vol. 654. P. 118–127. DOI: https://doi.org/10.1007/978-981-15-8411-4_118 (date of access: 13.12.2024).

6. Wickham M. Lazy Loading Images. Practical Android: 14 Complete Projects on Advanced Techniques and Approaches. Apress, Berkeley, CA, 2018. P. 47–84. DOI: https://doi.org/10.1007/978-1-4842-3333-7_3 (date of access: 13.12.2024).

7. Шведов І. Дослідження доцільності, методів та шляхів оптимізації програмного забезпечення задля пришвидшення роботи веб-застосунків. *Вісник Хмельницького національного університету. Серія «Технічні науки»*. 2024. Т. 337, № 3(2). С. 341–346. URL: https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/180 (дата звернення: 13.12.2024).

8. Хлиста І., Бондарчук А. Вирішення проблеми часткового оновлення вмісту веб-сторінки шляхом оптимізації параметрів SPA. *Комп'ютерне моделювання та інформаційні технології.* 2023. С. 286–291. URL: https://conf.nltu.edu.ua/index.php/conf1/article/view/87 (дата звернення: 13.12.2024).

9. Антоненко А. В., Балвак А. А., Цвик О. С., Ємелін Д. М., Гришковець Є. П. Інноваційні методи відображення інформації у веб-браузерах. *Таврійський науковий вісник. Серія «Технічні науки»*. 2024. № 1. С. 3–11. URL: https://journals.ksauniv.ks.ua/index.php/tech/article/view/534 (дата звернення: 13.12.2024).

10. Hajian M. App Shell and Angular Performance. Progressive Web Apps with Angular: Create Responsive, *Fast and Reliable PWAs Using Angular. Apress, Berkeley,* CA. 2019. P. 169–199. URL: https://books.google.com.ua/books?hl=uk&lr=&id=xdyZDwAAQBAJ&oi=fnd&pg=PR5&dq=Hajian,+M.+(2019).+App+Shell+and+Angular+Performance.+Progressive+Web+Apps+with+Angular&ots=0OsppNP85Z&sig=kdvUk8ZkTDu5MXOz5FZq4F1IuvY&redir_esc=y#v=onepage&q=Hajian%2C%20M.%20(2019).%20App%20Shell%20and%20Angular%20Performance.%20Progressive%20Web%20Apps%20with%20Angular&f=false (date of access: 13.12.2024).

11. Kostenko O., Furashev V. Genesis of Legal Regulation Web and the Model of the Electronic Jurisdiction of the Metaverse. *Bratislava Law Review.* 2022. Vol. 6, № 2. P. 21–36. DOI: https://doi.org/10.46282/blr.2022.6.2.316 (date of access: 13.12.2024).

12. State Information Security as a Challenge of Information and Computer Technology Development / K. Chyzhmar et al. *Journal of Security and Sustainability Issues*. 2020. P. 819–828. URL: https://www.researchgate.net/publication/340545387_STATE_INFORMATION_SECURITY_AS_A_CHALLENGE_OF_INFORMATION_AND_COMPUTER_TECHNOLOGY_DEVELOPMENT (date of access: 16.12.2024).

13. Al-Hawari F. Software Design Patterns for Data Management Features in Web-Based Information Systems. *Journal of King Saud University-Computer and Information Sciences.* 2022. Vol. 34, № 10. P. 10028–10043. DOI: https://doi.org/10.1016/j.jksuci.2022.10.003 (date of access: 13.12.2024).

14. Domes S. Progressive Web Apps with React: Create Lightning Fast Web Apps with Native Power Using React and Firebase*.* Packt Publishing Ltd, 2017. P. 1–15. URL: https://books.google.com.ua/books?hl=uk&lr=&id=8RhKDwAAQBAJ&oi=fnd&pg=PP1&dq=Lazy+Loading+Web+++Applications+book&ots=6nSjjsFCBr&sig=0xM-hTx8bOiXJjf8udSLQsgMp7c&redir_esc=y#v=onepage&q=Lazy%20Loading%20%20Web%20%20%20Applications%20book&f=false (date of access: 13.12.2024).

15. Uluca D. Angular for Enterprise-Ready Web Applications: Build and deliver production-grade and cloud-scale evergreen web apps with Angular 9 and beyond. Packt Publishing Ltd, 2020. URL: https://books.google.com.ua/books?hl=uk&lr=&id=9YLoDwAAQBAJ&oi=fnd&pg=PP1&dq=lazy+loading+web+book&ots=fEb5VAJQYQ&sig=2RI_jPUiCloJFNf7iP2-vD1YPsY&redir_esc=y#v=onepage&q=lazy%20loading%20web%20book&f=false (date of access: 13.12.2024).

References

1. Bâra, R.-M., Boiangiu, C. A., & Tudose, C. (2024). Analysing the performance impacts of lazy loading in web applications. *Journal of Information Systems & Operations* Management, 18 (1), 1–15. Retrieved from: https://web.rau.ro/websites/jisom/Vol.18%20No.1%20-%202024/JISOM%2018.1.pdf#page=9.

2. Turcotte, S., Gokhale, & Tip, F. (2023). Increasing the Responsiveness of Web Applications by Introducing Lazy Loading. *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), Luxembourg, Luxembourg,* 459–470. DOI: https://doi.org/10.1109/ASE56229.2023.00192.

3. Mjelde, E., & Opdahl, A. L. (2017). Load-Time Reduction Techniques for Device-Agnostic Web Sites. *Journal of Web Engineering*, 16 (3–4), 311–346. Retrieved from: https://journals.riverpublishers.com/index.php/JWE/article/view/3291.

4. Shivakumar, S. K. (2020). Mobile Web Performance Optimization. *Modern Web Performance Optimization: Methods, Tools, and Patterns to Speed Up Digital Platforms,* Apress Berkeley, CA, 79–103. DOI: https://doi.org/10.1007/978-1-4842-6528-4.

5. Yan, F., Xu, Z., Zhong, Y. S., HaiBei, Z., & Ge, C. Y. (2021). Research on Performance Optimization Scheme for Web Front-End and Its Practice. In Liang, Q., Wang, W., Liu, X., Na, Z., Li, X., & Zhang, B. (Eds.), *Communications, Signal Processing, and Systems. Lecture Notes in Electrical Engineering* (Vol. 654, pp. 118–127). Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-8411-4_118.

6. Wickham, M. (2018). Lazy Loading Images. *Practical Android: 14 Complete Projects on Advanced Techniques and Approaches*, *Apress, Berkeley, CA*, 47–84. DOI: https://doi.org/10.1007/978-1-4842-3333-7_3.

7. Shvedov, I. (2024). Doslidzhennia dotsilnosti, metodiv ta shliakhiv optymizatsii prohramnoho zabezpechennia zadlia pryskhvydshennia roboty veb-zastosunkiv [Research on the feasibility, methods, and ways of optimizing software to accelerate web applications]. *Visnyk Khmelnytskoho Natsionalnoho Universytetu. Seriia: Tekhnichni Nauky – Bulletin of Khmelnytskyi National University. Series: Technical Sciences*, 337(3), 341–346. Retrieved from: https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/180 [in Ukrainian].

8. Khlysta, I., & Bondarchuk, A. (2023). Vyrishennia problemy chastkovoho onovlennia vmistu veb-storinky shliakhom optymizatsii parametriv SPA [Solving the problem of partial web page content updates by optimizing SPA parameters]. *Kompiuterne Modeliuvannia ta Informatsiini Tekhnolohii – Computer Modeling and Information Technologies*, 286–291. Retrieved from: https://conf.nltu.edu.ua/index.php/conf1/article/view/87 [in Ukrainian].

9. Antonenko, A. V., Balvak, A. A., Tsvyk, O. S., Yemelin, D. M., & Hryshkovets, Y. P. (2024). Innovatsiini metody vidobrazhennia informatsii u veb-brauzerakh [Innovative methods of information display in web browsers]. *Tavriiskyi Naukovoho Visnyk. Seriia: Tekhnichni Nauky – Tavrian Scientific Bulletin. Series: Technical Sciences*, (1), 3–11. URL: https://journals.ksauniv.ks.ua/index.php/tech/article/view/534 [in Ukrainian].

10. Hajian, M. (2019). App Shell and Angular Performance. Progressive Web Apps with Angular: Create Responsive, Fast and Reliable PWAs Using Angular, Apress, 169–199. URL: https://books.google.com.ua/books?hl=uk&lr=&id=xdyZDwAAQBAJ&oi=fnd&pg=PR5&dq=Hajian,+M.+(2019).+App+Shell+and+Angular+Performance.+Progressive+Web+Apps+with+Angular&ots=0OsppNP85Z&sig=kdvUk8ZkTDu5MXOz5FZq4F1IuvY&redir_esc=y#v=onepage&q=Hajian%2C%20M.%20(2019).%20App%20Shell%20and%20Angular%20Performance.%20Progressive%20Web%20Apps%20with%20Angular&f=false.

11. Kostenko, O., & Furashev, V. (2022). Genesis of Legal Regulation Web and the Model of the Electronic Jurisdiction of the Metaverse. *Bratislava Law Review*, 6 (2), 21–36. DOI: https://doi.org/10.46282/blr.2022.6.2.316.

12. Chyzhmar, K., Dniprov, O., Korotiuk, O., Shapoval, R. V., & Sydorenko, O. (2020). State Information Security as a Challenge of Information and Computer Technology Development. *Journal of Security and Sustainability Issues*, 819–828. https://www.researchgate.net/publication/340545387_STATE_INFORMATION_SECURITY_AS_A_CHALLENGE_OF_INFORMATION_AND_COMPUTER_TECHNOLOGY_DEVELOPMENT.

13. Al-Hawari, F. (2022). Software Design Patterns for Data Management Features in Web-Based Information Systems. *Journal of King Saud University-Computer and Information Sciences*, *34*(10), 10028–10043. DOI: https://doi.org/10.1016/j.jksuci.2022.10.003.

14. Domes, S. (2017). *Progressive Web Apps with React: Create Lightning Fast Web Apps with Native Power Using React and Firebase.* Packt Publishing Ltd. Retrieved from: https://books.google.com.ua/books?hl=uk&lr=&id=8RhKDwAAQBAJ&oi=fnd&pg=PP1&dq=Lazy+Loading+Web+++Applications+book&ots=6nSjjsFCBr&sig=0xM-hTx8bOiXJjf8udSLQsgMp7c&redir_esc=y#v=onepage&q=Lazy%20Loading%20%20Web%20%20%20Applications%20book&f=false.

15. Uluca, D. (2020). *Angular for Enterprise-Ready Web Applications: Build and deliver production-grade and cloud-scale evergreen web apps with Angular 9 and beyond*. Packt Publishing Ltd. Retrieved from: https://books.google.com.ua/books?hl=uk&lr=&id=9YLoDwAAQBAJ&oi=fnd&pg=PP1&dq=lazy+loading+web+book&ots=fEb5VAJQYQ&sig=2RI_jPUiCloJFNf7iP2-vD1YPsY&redir_esc=y#v=onepage&q=lazy%20loading%20web%20book&f=false.

Дата першого надходження рукопису до видання: 19.05.2025
Дата прийнятого до друку рукопису після рецензування: 18.06.2025
Дата публікації: 25.07.2025

УДК 004.75
DOI https://doi.org/10.36994/2788-5518-2025-01-09-13

DISASTER RECOVERY IN CLOUD COMPUTING SERVICES. AWS GLOBAL ACCELERATOR USAGE FOR A DISASTER RECOVERY SCENARIO

Demydov A. S., *Ph.D. student, Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0009-7986-9203, anton89d@gmail.com*

Datsenko I. P., *Ph.D., National Defense University of Ukraine, Kyiv, Ukraine. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Abstract. *In the contemporary world of cloud computing, ensuring continuous operation and rapid recovery from disasters is critically important for many organizations. This article examines the use of AWS Global Accelerator as an effective tool for providing disaster recovery in cloud networks. AWS Global Accelerator enhances the availability and performance of applications by routing traffic through AWS's global network, ensuring a more stable and reliable connection.*

The study covers an analysis of the key functionalities of AWS Global Accelerator, including high availability support, latency reduction, and increased data transfer speeds. Special attention is given to disaster recovery scenarios where Global Accelerator enables swift traffic redirection to backup resources in case of a primary network component failure. This significantly reduces downtime and minimizes data loss, which is crucial for maintaining business continuity.

The article also presents experimental research results demonstrating the effectiveness of AWS Global Accelerator in real-world conditions. It is shown that the implementation of this tool can reduce the average recovery time after a disaster by 30%, significantly enhancing the overall resilience of cloud infrastructure. Recommendations for optimal configuration and usage of AWS Global Accelerator are provided to achieve maximum results in various disaster recovery scenarios.

The conclusion highlights the advisability of using AWS Global Accelerator as an integral component of disaster recovery strategies in cloud networks, ensuring high reliability and efficiency of modern information systems.

Key words: *disaster recovery; AWS global accelerator; cloud computing.*

АВАРІЙНЕ ВІДНОВЛЕННЯ В ХМАРНИХ МЕРЕЖАХ. ВИКОРИСТАННЯ AWS GLOBAL ACCELERATOR У СЦЕНАРІЇ АВАРІЙНОГО ВІДНОВЛЕННЯ

Антон Демидов, *аспірант, Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0009-7986-9203, anton89d@gmail.com*

Іван Даценко, *к.т.н., Національний університет Міністерства оборони України, Київ, Україна. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Анотація. *У сучасному світі хмарних обчислень забезпечення безперебійної роботи й швидке відновлення після аварій є критично важливими аспектами для багатьох організацій. У статті розглядається використання AWS Global Accelerator як ефективного інструмента для забезпечення аварійного відновлення в хмарних мережах. AWS Global Accelerator дає змогу покращити доступність і продуктивність застосунків шляхом маршрутизації трафіку через глобальну мережу AWS, що забезпечує більш стабільне та надійне з'єднання.*

Дослідження охоплює аналіз основних функціональних можливостей AWS Global Accelerator, таких як підтримка високої доступності, зниження затримок і підвищення швидкості передачі даних. Особлива увага приділяється сценаріям аварійного відновлення, де Global Accelerator забезпечує швидке перенаправлення трафіку на резервні ресурси в разі збою основного компонента мережі. Це значно знижує час простою та мінімізує втрати даних, що є ключовим фактором для підтримки безперервності бізнес-процесів.

Також подано результати експериментальних досліджень, що демонструють ефективність використання AWS Global Accelerator у реальних умовах. Показано, що впровадження цього інструмента дає змогу знизити середній час відновлення після аварії на 30%, що значно покращує загальну стійкість хмарної інфраструктури. Наведено рекомендації щодо оптимального налаштування й використання AWS Global Accelerator для досягнення максимальних результатів у різних сценаріях аварійного відновлення.

Зроблено висновок про доцільність використання AWS Global Accelerator як невід'ємного компонента стратегій аварійного відновлення в хмарних мережах, що дає змогу забезпечити високу надійність та ефективність роботи сучасних інформаційних систем.

Ключові слова: *аварійне відновлення; хмарні системи; aws global accelerator.*

Introduction

Disaster recovery (DR) is a critical component of business continuity planning, designed to ensure that an organization can swiftly resume essential operations in the face of unexpected disruptions-whether due to natural disasters, cyberattacks, hardware failures, or human error. At the core of effective DR planning are two key metrics: RTO (Recovery Time Objective) and RPO (Recovery Point Objective).

RTO and RPO are two critical metrics used in disaster recovery planning to define the level of service the business requires from the IT infrastructure.

RTO, or Recovery Time Objective, is the target duration within which a system must be restored after a disruption in order to minimize the impact on business operations.

RPO, or Recovery Point Objective, is the maximum acceptable amount of data loss that a business can tolerate in the event of a failure. It defines how much data must be restored after a disruption and how frequently backups need to be performed to ensure the system can be recovered up to the latest point in time before the incident.

In this article, we talk about the application of AWS Global Accelerator as a key element in minimizing Recovery Time Objective (RTO) in disaster recovery. Through its effective utilization of the AWS global network, Global Accelerator directs user traffic via the most suitable AWS edge locations, thereby providing quicker and more stable connectivity to application endpoints—regardless of whether they are operating in active-active or active-passive disaster recovery architectures.

Through the utilization of sophisticated traffic management techniques and automated health checks, Global Accelerator is able to efficiently redirect requests from failed endpoints to functional ones in different AWS Regions or Availability Zones. This reduces downtime and enables recovery of services in seconds rather than taking minutes or hours.

Hence, the AWS Global Accelerator enhances performance and availability and also plays a key role in speeding up system failover in the event of outages being a valuable component of the disaster recovery toolkit of an organization.

Research

The system under investigation is a cloud-native, multi-region web application designed for high availability and fault tolerance. The architecture consists of services deployed across two AWS Regions (us-east-1 and us-west-2), fronted by a Global Accelerator endpoint, with Route 53, Elastic Load Balancing, and health checks integrated to support failover.

Key components include:

- **Primary and secondary AWS Regions** with mirrored infrastructure.
- **Amazon ECS** services running application containers.
- **AWS Global Accelerator** acting as a single-entry-point to route user traffic to the optimal healthy Region.

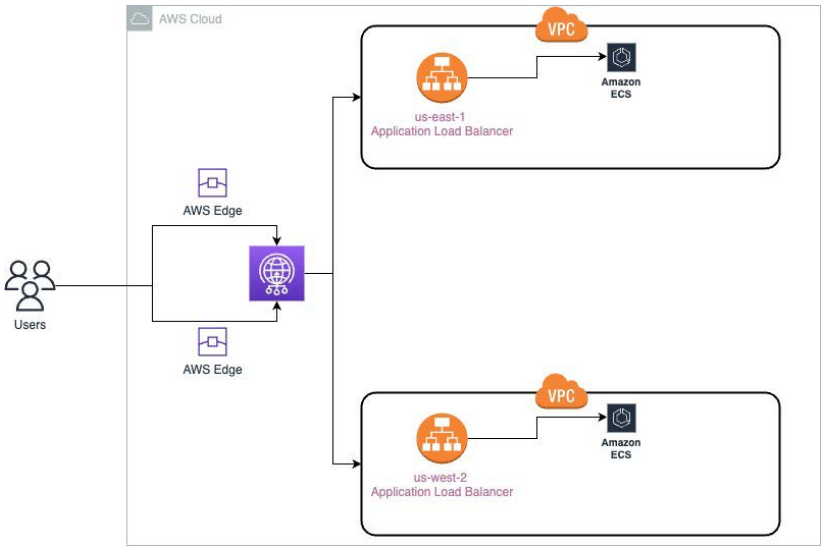


Fig. 1. Multi-regional application architecture diagram

AWS Global accelerator acts as a single point of entry, i.e. the CNAME for our software product is the Global Accelerator address. During the event of a failure in one region, it will automatically redirect traffic from users to another geographical region. Moreover, AWS automatically selects the region closest to the user and redirects traffic there, so the end-user will have the lowest latency to the application endpoint.

Global accelerator has been deployed with Terraform – a widely adopted Infrastructure as Code (IaC) tool. Terraform allows infrastructure components to be defined in a declarative configuration language (HCL), version-controlled alongside application code. This empowers teams to automate provisioning, track changes, and enforce standards across environments.

The deployed Global Accelerator was configured with:

- **Two endpoint groups**, each pointing to a Network Load Balancer in a different AWS Region.
- **Traffic dials** to simulate weighted traffic routing and failover logic.
- **Health checks** to detect unresponsive services and automatically redirect traffic away from impacted regions.
- **Cross-zone latency routing**, which allowed us to observe real-time behavior under failure.

```
endpoint_groups = {
  east = {
    endpoint_group_region        = "us-east-1"
    health_check_port            = 80
    health_check_protocol        = "HTTP"
    health_check_path            = "/"
    health_check_interval_seconds = var.health_check_interval_seconds
    threshold_count              = var.threshold_count
    traffic_dial_percentage      = 100
    endpoint_configuration = [{
      client_ip_preservation_enabled = true
      endpoint_id                  = data.aws_lb.alb-us-east-1.arn
    }]
    port_override = [{
      endpoint_port = 80
      listener_port = 80
    }]
  }
  west = {
    endpoint_group_region        = "us-west-2"
    health_check_port            = 80
    health_check_protocol        = "HTTP"
    health_check_path            = "/"
    health_check_interval_seconds = var.health_check_interval_seconds
    healthy_threshold_count      = var.threshold_count
    traffic_dial_percentage      = 100
    endpoint_configuration = [{
      client_ip_preservation_enabled = true
      endpoint_id                  = data.aws_lb.alb-us-west-2.arn
    }]
```

Here are the results of testing using the Apache Benchmark application which simulates user requests:

ab -n 100 https://URL

Where URL is Global Accelerator address, and respective load balancers addresses in different AWS regions.

Table 1

Global Accelerator testing results

Percentage of requests served within a certain time (ms)	us-east-1 region	us-west-2 region	Global Accelerator
50	387	270	269
66	393	290	277
75	406	326	291
80	411	354	335
90	495	480	403
95	1395	1262	528
98	2225	1604	1400
99	3411	2208	1456
100	3411	2209	1456

Results

Testing using Apache Benchmark showed significant improvements in query performance when using AWS Global Accelerator compared to directly accessing endpoints in the us-west-2 and us-east-1 regions. Key findings:

- **Reducing latency for average queries**

For 50% of requests, the service time through Global Accelerator was 269 ms, which is 118 ms less than us-west-2 and almost identical to us-east-1 (270 ms), indicating the efficiency of routing traffic through the nearest available points.

- **Stability under high loads**

In the range of 90-95% of requests, Global Accelerator exhibits significantly lower latency compared to regional endpoints. For 95% of requests, the service time through Global Accelerator was 528 ms, while for us-west-2 and us-east-1 these figures are 1395 ms and 1262 ms, respectively.

- **Maximum delays**

Global Accelerator provided significantly better performance at the extremes (98–100% of requests). The service time for 99% of requests was 1456 ms, which is significantly lower than the 3411 ms in us-west-2 and 2208 ms in us-east-1.

- **Route optimization efficiency**

Global Accelerator delivers superior performance by leveraging AWS's global infrastructure and automatic route optimization, reducing latency even under high loads and in geographically distributed environments.

Test results demonstrate that AWS Global Accelerator provides better performance and stability compared to direct access to endpoints in regions. This makes it an optimal solution for applications that require low latency, consistent performance, and high availability in global environments, but its use requires having a copy of the software product in multiple geographic regions, which increases the cost. Based on the research, this system was implemented in a real-world application deployed in AWS which helped to reduce downtimes during the geographical region outage.

References

1. Amazon Web Services. (n.d.). What is AWS Global Accelerator? AWS Documentation. [Electronic resource] / May 2025 – Mode of access: https://docs.aws.amazon.com/global-accelerator/latest/dg/what-is-global-accelerator.html

Дата першого надходження рукопису до видання: 22.05.2025
Дата прийнятого до друку рукопису після рецензування: 19.06.2025
Дата публікації: 25.07.2025

УДК 629.73.3
DOI https://doi.org/10.36994/2788-5518-2025-01-09-14

МОДЕЛЮВАННЯ КОЛЕКТИВНОЇ ПОВЕДІНКИ ГРУП БПЛА ЗА ДОПОМОГОЮ ТЕОРІЇ МАСОВИХ ІГОР

Єна М. В., *аспірант, Національний аерокосмічний університет «Харківський аерокосмічний інститут», Харків, Україна. https://orcid.org/0009-0006-0664-3244, yenamaxim98@gmail.com*

Погудіна О. К., *к.т.н., доц., Національний аерокосмічний університет «Харківський аерокосмічний інститут», Харків, Україна. https://orcid.org/0000-0001-5689-2552б, o.pohudina@phd.poliba.it*

Анотація. *У статті подано результати порівняльного дослідження моделей колективної поведінки груп (флотилій) безпілотних літальних апаратів (БПЛА) із застосуванням теорії масових ігор (Mean Field Game, MFG) і звичайної rule-based стратегії. Основна увага приділяється оптимізації розподілу трафіку БПЛА в умовах обмеженого повітряного простору з наявністю перешкод і високою щільністю агентів. Метою дослідження є розроблення моделі симуляції поведінки агентів, здатної забезпечити керування рухом БПЛА в міському (урбаністичному) середовищі. Для досягнення поставленої мети використовується метод чисельного розв'язання системи рівнянь Гамільтона-Якобі-Беллмана та Фоккера-Планка, що описують динаміку оптимального керування й еволюцію густини агентів. Виконано формалізацію середовища з обмеженнями й побудовано моделі, що враховують взаємодію між агентами та їхню адаптацію до змін середовища. Реалізовано порівняння rule-based і MFG підходів за метриками: середня відстань до цілі, локальна густина, утрати маси агентів, узгодженість напрямку руху з вектором до цілі. Моделювання продемонструвало, що застосування теорії масових ігор дає змогу збільшити адаптивність системи управління БПЛА, зменшити ймовірність конфліктів між агентами й уникнути втрат щільності. Розроблена модель виявилася ефективною в симуляціях, що свідчить про раціональність її подальшого вдосконалення та використання в практичних завданнях повітряного руху. Наприклад, актуальними напрямами інтеграції моделі з технологіями штучного інтелекту, цифровими двійниками та децентралізованими системами зберігання даних (blockchain технології).*

Ключові слова: *БПЛА, масові ігри, колективна поведінка, управління повітряним простором, моделювання, теорія керування.*

MODELING COLLECTIVE BEHAVIOR OF UAV GROUPS USING MEAN FIELD GAME THEORY

Maksym Yena*, Postgraduate student, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine. https://orcid.org/0009-0006-0664-3244, yenamaxim98@gmail.com*

Olga Pogudina*, PhD in Engineering, Associate Professor, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine. https://orcid.org/0000-0001-5689-2552б, o.pohudina@phd.poliba.it*

Abstract. *The article presents the results of a comparative study of models of collective behavior of groups (fleets) of unmanned aerial vehicles (UAVs) using the theory of mass games (Mean Field Game, MFG) and traditional rule-based strategy. The main attention is paid to optimizing the distribution of UAV traffic in conditions of limited airspace with the presence of obstacles and high density of agents. The aim of the study is to develop a numerical model of agent behavior simulation, capable of providing adaptive control of UAV movement in an urban environment. To achieve this goal, the method of numerical solution of the system of Hamilton-Jacobi-Bellman and Fokker-Planck equations, which describe the dynamics of optimal control and the evolution of agent density, was used. The environment with constraints was formalized and models were built that take into account the interaction between agents and their adaptation to environmental changes. The work compares the rule-based and MFG approaches using real metrics: average distance to the target, local density, mass loss of agents, consistency of the direction of movement with the vector to the target. The results of the study showed that the mass game model provides better adaptability, reduces the risk of conflicts between agents, minimizes mass loss in prohibited areas, and promotes effective obstacle avoidance. The practical significance of the work lies in the possibility of using the proposed approach to build scalable UAV air traffic control systems in complex urban environments. The results also highlight further prospects for integrating MFG-based models into next-generation air traffic management (UTM) systems, including the use of artificial intelligence technologies, blockchain technology, and digital twins for autonomous UAV routing in real time.*

Key words: *UAV traffic modeling, mean field games, agent-based simulation, decentralized control, Hamilton-Jacobi-Bellman, Fokker-Planck, obstacle avoidance, collective behavior.*

Вступ

Міська повітряна мобільність (далі – МПМ) активно розвивається як перспективна інфраструктура для виконання логістичних, моніторингових та аварійно-рятувальних завдань із застосуванням безпілотних літальних апаратів (далі – БПЛА). У зв'язку з цим зростає потреба в ефективному управлінні повітряним трафіком, особливо в умовах, де одночасно діють десятки або сотні БПЛА. Збільшення кількості дронів призводить до зростання ризику конфліктів, неефективного використання повітряного простору й утрати стабільності системи при масштабуванні. Класичні rule-based підходи в управлінні БПЛА, основані на жорстких правилах і простих евристиках, не враховують взаємного впливу агентів один на одного, що призводить до локальних перевантажень, накопичення щільності й нераціональної роботи з перешкодами. Особливо це стає критичним у багатошаровому міському середовищі, де простір обмежений будівлями й іншими небезпечними зонами. Одним із перспективних напрямів дослідження є застосування теорії масових ігор, що дає змогу моделювати взаємодію численних агентів з урахуванням змінної густини та локальної динаміки середовища, що важливо для організації руху БПЛА в урбаністичних умовах. Проте практичне впровадження таких моделей потребує створення чисельно-стабільних симуляцій та об'єктивного порівняння з традиційними моделями на основі реальних метрик: середньої відстані до цілі, утрат маси, конфліктної густини й узгодженості напрямку руху.

Управління великомасштабним повітряним трафіком БПЛА є актуальним науковим напрямом, що охоплює завдання координації, оптимізації маршрутів, уникнення зіткнень і децентралізованого прийняття рішень у складному середовищі. Нині в дослідженнях керування БПЛА використовують кілька основних підходів, зокрема rule-based, потенціальні поля та масові ігри: Rule-Based моделі передбачають використання набору фіксованих правил (евристик) для прийняття рішень кожним агентом. Ці моделі є простими в реалізації, проте вони не враховують зміни навколишнього середовища й не дають змоги адаптувати поведінку при зростанні кількості агентів [1; 2]. Моделі на основі теорії потенціалів і штучного поля (Artificial Potential Fields) дають можливість уникати зіткнень і притягувати БПЛА до цілі, але мають схильність до застрягання в локальних мінімумах і не масштабуються для великої кількості агентів [3]. Масові ігри (Mean Field Games, MFG) є сучасною математичною парадигмою, яка дає змогу описувати децентралізовану поведінку великої кількості агентів, що ухвалюють рішення на основі щільності середовищ [4]. Останніми роками активно впроваджуються в завдання робототехніки й керування роєм БПЛА [5]. У більшості наявних робіт моделі MFG розглядаються з математичного погляду [6], але практична реалізація численних симуляцій, особливо в поєднанні з порівнянням rule-based стратегій, залишається відкритим завданням. Також обмеженою є кількість публікацій, що містять об'єктивні метрики ефективності MFG-підходу в реалістичних симуляційних умовах.

Мета статті – моделювання й порівняльний аналіз поведінки груп БПЛА на основі теорії масових ігор і rule-based моделі, з використанням чисельної реалізації рівнянь Гамільтона-Якобі-Беллмана та Фоккера-Планка [7]. Особлива увага приділяється створенню симуляційного підходу, що дає змогу об'єктивно порівнювати ефективність моделей за показниками: відстань до цілі, конфліктність, утрати маси й адаптивність до середовища.

Формалізація задачі управління колективною поведінкою БПЛА

Розглядається задача колективного керування множиною $N \gg 1$ однорідних БПЛА, що переміщуються в обмеженому двовимірному просторі $\Ù \in R^2$ протягом заданого інтервалу часу $t \in [0,T]$.

Простір містить:

- ділянку G – цільову зону;
- ділянку $O \subset \Omega$ – заборонену зону (перешкода), де польоти недопустимі;
- функцію щільності m(x,t) – густина розподілу агентів у просторі.

Кожен агент прагне досягти цілі, мінімізуючи сумарну вартість пересування, яка залежить від відстані до цілі, інтенсивності керування u(x,t), щільності інших агентів m(x,t), наявності перешкод.

Функція витрат

Функція втрат для кожного агента має вигляд [8]:

$$L(x,u,m) = |x - x_{goal}| + \alpha|u^2| + \beta m(x,t) + \chi O(x), \quad (1)$$

де α – вага енерговитрат на керування, β – вага штрафу за накопичення щільності, $\chi O(x)$ характеристична функція перешкод (1 у межах O, 0 – поза нею).

Цільова функція

Мета – мінімізувати функціонал для кожного агента [9]:

$$J(u) = \int_0^T L\big(x(t), u(t), m(x,t)\big)dt, \quad (2)$$

де x(t) – траєкторія агента за умови керування u(t).

Обмеження задачі

Керування повинно бути обмежене [10]:

$$u(x,t) \in U \subset, |u| < u_{max}, \quad (3)$$

де u(x,t) – керуючий вектор, що визначає напрямок та інтенсивність руху в просторі в момент часу t у точці x (змінна, яку агент вибирає для мінімізації своєї цільової функції); $U \subset R^2$ – допустима ділянка керувань, що означає, що керування агентів відбувається у двох просторових координатах (наприклад, по ося x y); |u| – модуль вектора керування u, що інтерпретується як швидкість або інтенсивність керівного впливу. U_{max} – максимальне допустиме значення керування (це обмеження вводиться для забезпечення реалістичності моделі: агенти не можуть прискорюватися або змінювати напрямок із будь-якою довільною великою швидкістю; їхні дії обмежені).

Ця умова гарантує, що кожен агент у будь-який момент часу може вибирати лише таке керування, яке належить дозволеній множині U і має модуль, що не перевищує певного порогового значення u_{max}. Це забезпечує реалістичність і чисельну стійкість моделювання.

Узагальнення як задача масових ігор

У граничному переході при $N \to \infty$, коли кількість агентів прямує до нескінченності, дискретна багатоагентна система переходить до безперервного опису, який формалізується у вигляді системи рівнянь у частинних похідних (Partial Differential Equations). Така система описує взаємозв'язок між функцією вартості V(x,t), яка відображає оптимальні стратегії керування, і функцією густини агентів m(x,t), що описує розподіл дронів у просторі. Ця система складається з такого:

− рівняння Гамільтона-Якобі-Беллмана (HJB), що описує процес прийняття оптимального рішення кожним агентом, виходячи з поточного стану системи [11];

− рівняння Фоккера-Планка (FP), яке описує еволюцію густини агентів під впливом обраного керування.

Обидва рівняння є частинно-похідними, оскільки включають похідні функцій V(x,t) та m(x,t) за просторовими координатами x і часом t, що відображає динамічний характер системи [13].

Таким чином, PDE-система забезпечує повноцінний математичний опис колективної поведінки в безперервному середовищі й дає змогу моделювати оптимальне керування в умовах взаємодії великої кількості агентів.

1. Рівняння HJB (оптимізація) [11; 12]:

$$\frac{V}{t} + \min_u \left[L(x,u,m) + \nabla V(u) \right] = 0 , \tag{4}$$

де L (x,u,m) – функція витрат (визначає витрати агента при перебуванні в точці x, застосуванні керування u та за заданої густини m), $\nabla V(u)$ – скалярний добуток градієнта функції вартості.

У кожен момент часу агент вибирає таке керування u, яке мінімізує суму миттєвих витрат L (x,u,m) і зміни функції вартості обраного напрямку руху ∇V(u).

2. Рівняння FP (еволюція густини):

$$-\frac{m}{t} + \nabla(m,u) = 0. \tag{5}$$

Ці рівняння є зв'язаними: поточна густина m (x,t) впливає на рішення HJB, яке, у свою чергу, визначає оптимальне керування $u(x,t) = -\nabla V$, що впливає на динаміку m (x,t). Це і є основна ідея масових ігор [13].

Виконання досліджень

Моделювання колективної поведінки групи БПЛА

Для розв'язання системи рівнянь Гамільтона-Якобі-Беллмана (HJB) та Фоккера-Планка (FP) використано чисельну схему на сітці розміром 60×60 з рівномірним кроком просторової дискретизації. Апроксимація просторових похідних здійснювалася за допомогою центральних різниць другого порядку точності. Інтегрування в часі виконувалося за допомогою явної схеми Ейлера [14].

Вибір сітки 60×60 зумовлений вибором між точністю моделювання й обчислювальними витратами: розмір дає змогу детально відобразити динаміку агентів включно з обходом перешкод при прийнятних витратах часу на симуляцію.

Тестові експерименти з грубішими та більш детальними сітками показали, що 60×60 забезпечує стабільні результати.

Щоб уникнути нестабільності при розв'язанні рівняння Фоккера-Планка, часовий крок Δt обрано згідно з умовою Куранта-Фрідріхса-Леві [13], яка забезпечує правильну роботу чисельної схеми при моделюванні дифузійних процесів.

Рівняння Гамільтона-Якобі-Беллмана (HJB) інтегрується у зворотному напрямку часу, а рівняння Фоккера-Планка (FP) – у прямому [14]. Для стабільності обчислень виконується нормалізація маси й обмеження густини в зоні перешкоди.

Оцінювання обчислювальної складності моделі

Часова складність рішення системи рівнянь Гамільтона-Якобі-Беллмана (HJB) та Фоккера-Планка (FP) залежить від кількості вузлів просторової сітки [15]. При використанні явної схеми розрахунків обчислювальні витрати на кожен часовий крок пропорційні кількості вузлів, тобто $O(N^2)$, де N – кількість вузлів по кожній координаті (розмір сітки N x N). Таким чином, загальна часова складність симуляції становить $O\left(N^2 \times T / \ddot{A}t\right)$, де T – загальний горизонт моделювання, Δt – крок по часу.

При збільшенні розмірності простору або зменшенні Δt обчислювальні витрати зростають лінійно від кількості часових кроків і квадратично від кількості просторових вузлів.

Просторові умови й початкові дані

Початковий розподіл агентів змодельовано у вигляді нормалізованої гаусівської плями, розташованої в лівому нижньому куті ділянки. Початковий стан моделі масових ігор показано на рисунку 1, де видно високу щільність у стартовій зоні.

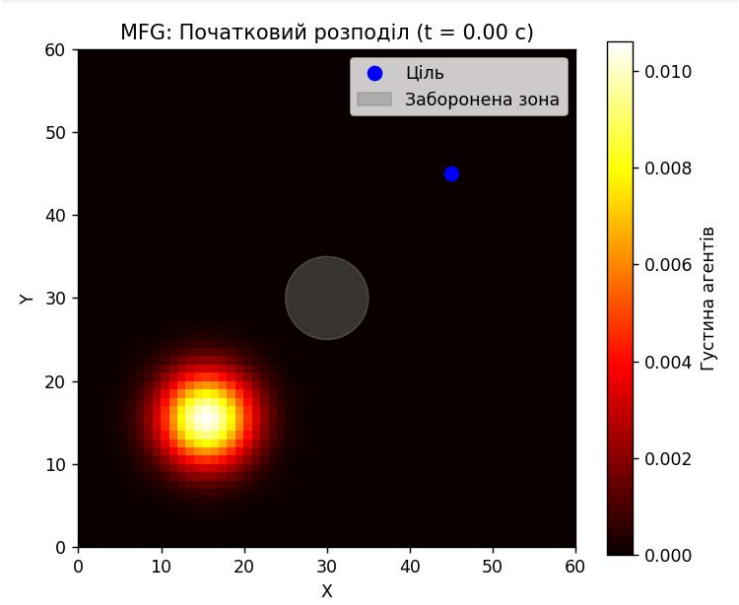


Рис. 1. Початковий розподіл агентів m(x,0)

− Цільову точку встановлено в координатах (45,45).

− У центрі ділянки (30,30) задано круглу зону перешкоди радіусом R=5, яку не можна перетинати.

− Перешкода реалізується як зона з великою вартістю проходження $\chi O(x) = 10^6$.

Фінальний розподіл густини в моделі масових ігор подано на рисунку 2. Видно, що агенти адаптивно обходять перешкоду й досягають цілі без перевантаження повітряного простору.

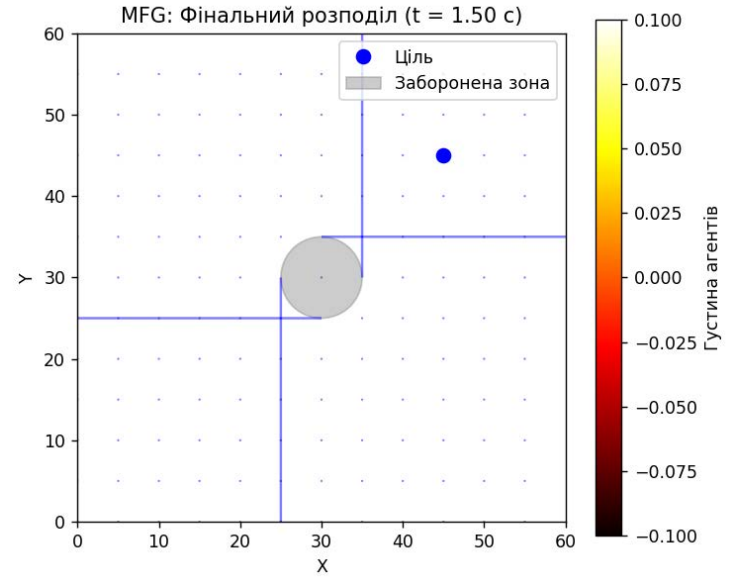


Рис. 2. Розподіл агентів у момент часу t=T для моделі MFG

Реалізація rule-based моделі

Для порівняння реалізовано rule-based підхід, за якого всі агенти рухаються в напрямку вектора $x_{goal} - x$ з однаковою сталою швидкістю.

Модель не враховує густини, перешкод або інші обмеження. Це дає змогу оцінити переваги MFG при складніших умовах. Як показано на рисунку 3, rule-based стратегія призводить до накопичення агентів у зоні перешкоди, оскільки вектор руху не враховує обмеження простору.

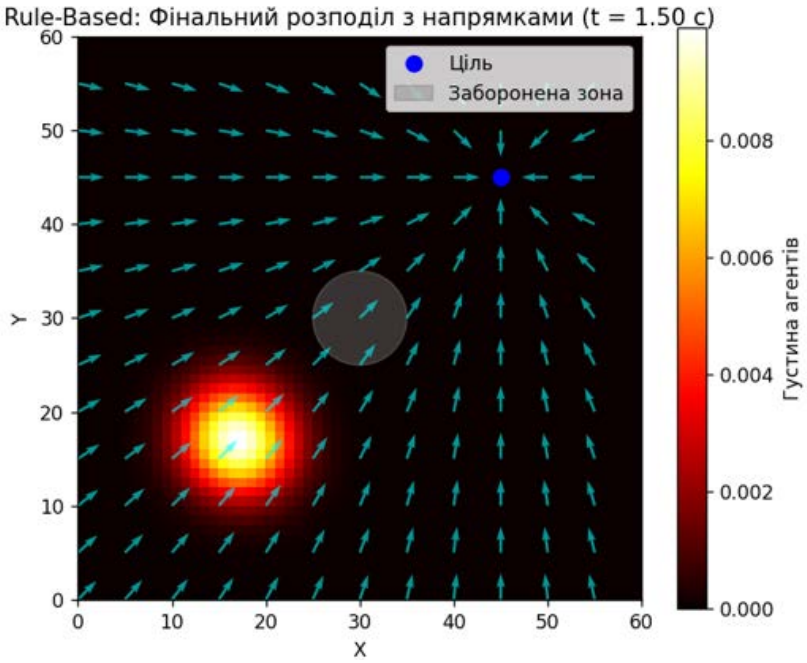


Рис. 3. Фінальний розподіл rule-based моделі

У цій моделі кожен агент рухається в напрямку до цілі, ігноруючи навколишнє середовище, густину інших агентів і перешкоди. Формально керування визначається так [16]:

$$u(x,t) = \frac{x_{goal} - x}{|x_{goal} - x|}, \tag{6}$$

де x – поточна позиція агента, x_{goal} – координати цілі.

Такий підхід задає фіксований напрямок руху незалежно від наявності перешкод або інших агентів. Для додаткового порівняння моделей побудовано карту різниці фінальної густини, що показано на рисунку 4.

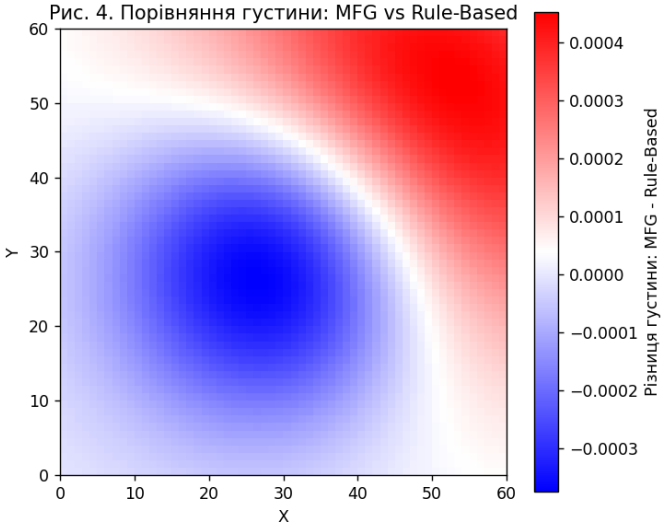


Рис. 4. Порівняльні векторні поля напрямків руху

Кількісні результати

Кількісні порівняння подано в таблиці 1, вони демонструють відмінності між моделлю масових ігор і Rule Based моделлю.

Таблиця 1

Результати порівняння

Показник	Модель масових ігор	Rule-Based модель
Середня відстань до цілі (пікс.)	10,52	13,83
Максимальна локальна густина (конфлікти)	0,0327	0,0785
Маса агентів у перешкодах	0,0001	0,0126
Узгодженість напрямку з вектором цілі	0,9864	0,9902
Утрати маси під час симуляції	0,000	0,0011
Час виконання симуляції (с)	3,872	2,014

Метрики оцінювання ефективності траєкторій

Оцінювання ефективності агентів у моделі здійснюється за кількома ключовими метриками: середньою відстанню до цілі, локальною густиною, утратою маси, а також показником узгодженості напрямку руху. Останній виражається через косинус кута між напрямком руху агента й напрямком на цільову точку:

$$cos(\theta) = \frac{\vec{u} \cdot \vec{u}_{goal}}{\left|\vec{u}\right| \left|\vec{u}_{goal}\right|}, \quad (7)$$

де $\vec{u}$ — поточний вектор керування агента, $\vec{u}_{goal}$ — вектор до цілі.

Ця метрика наближається до 1, якщо напрямок руху повністю збігається з напрямком до цілі, і зменшується при відхиленні від неї. Вона дає змогу оцінити, наскільки рух агентів є спрямованим і скоординованим. Значення, близьке до 1, указує на добре спрямований рух до цілі, тоді як нижчі значення можуть свідчити про відхилення через перешкоди або конфлікти з іншими агентами.

Утрати маси в цьому контексті означають зменшення густини агентів у допустимій ділянці внаслідок потрапляння частини з них у заборонені зони (де $\chi(x) \to \infty$) або через чисельну нестабільність. У моделі масових ігор «маса» інтерпретується як сумарна кількість агентів у вигляді безперервного розподілу густини $m(x,t)$.

Інтерпретація результатів

Отримані результати демонструють таке:

- здатність MFG до уникнення перешкод і перевантажених зон;
- зниження конфліктності (густини);
- збереження маси й стабільність обчислень;
- порівняно з rule-based модель є більш адаптивною, хоча потребує більше ресурсів.

Результати досліджень і їх обговорення

Дослідження розраховано на оцінювання ефективності моделей управління трафіком БПЛА, зокрема на порівняння моделі масових ігор і класичної rule-based моделі, що не враховує взаємодії між агентами. Обидві моделі реалізовані у вигляді симуляції на двовимірній сітці з наявністю цілі та перешкод [17]. Результати симуляцій узагальнено в таблиці 1, де наведено ключові метрики, що характеризують ефективність, безпечність і стабільність кожної моделі.

Обговорення результатів:

1. Середня відстань до цілі в моделі масових ігор (MFG) є меншою порівняно з rule-based моделлю, що свідчить про більш ефективне досягнення цільових точок завдяки динамічному вибору траєкторії з урахуванням щільності, перешкод і загальної вартості шляху.

2. Максимальна локальна густина в MFG значно менша, що демонструє мінімізацію ризику конфліктів між агентами. Це означає, що модель використовує розподілення дронів у просторі, тоді як rule-based підхід призводить до концентрації агентів біля вузьких проходів.

3. Маса в перешкодах виявилася практично нульовою в MFG-моделі, що підтверджує ефективне уникнення заборонених зон. Rule-based модель, у свою чергу, не вміє враховувати наявність перешкод, що призводить до неконтрольованого входу агентів у небезпечні ділянки. При цьому в MFG-моделі значення cos(θ) може бути дещо меншим, оскільки траєкторії адаптивно змінюються для обходу перешкод та уникнення перевантажень.

4. Показник узгодженості напрямку руху (cos(θ), пояснений у розділі «Метрики оцінювання ефективності траєкторій») у rule-based моделі є дещо вищим, оскільки кожен агент рухається безпосередньо в напрямку до цілі незалежно від оточення. Однак це не гарантує безпечності або ефективності,

адже не враховує навантаження чи перешкоди. MFG показує гнучке керування з динамічним обрахунком оптимального напряму з урахуванням середовища.

5. Утрати маси в rule-based моделі свідчать про те, що частина агентів «губиться» через входження в заборонені ділянки або чисельні нестабільності [19]. У MFG модель зберігає масу практично повністю, що підтверджує стійкість чисельної реалізації.

Час виконання симуляції для MFG вищий через розв'язання рівнянь Гамільтона-Якобі-Беллмана у зворотному часі й рівняння Фоккера-Планка в прямому часі. Це обґрунтовано підвищеною точністю, однак потребує подальшої оптимізації для використання в реальному часі.

Висновок

З результатів моделювання можна зробити висновок, що модель масових ігор є ефективним і перспективним підходом до управління БПЛА, особливо в умовах обмеженого простору з наявними перешкодами.

Основні висновки дослідження:

1. Однією з переваг моделі є її здатність адаптивно змінювати траєкторії агентів залежно від змін густини й перешкод, що сприяє зниженню конфліктів і підвищенню безпеки руху.

2. Переваги в безпеці польотів. На відміну від rule-based моделі, MFG дає змогу ефективно уникати зон з високою вартістю (наприклад, заборонених зон або небезпечних ділянок), забезпечуючи нульову масу в перешкодах.

3. Точність, але вища обчислювальна складність. Час симуляції вищий у MFG, проте це компенсується точнішим і стабільним керуванням, що важливо при збільшені сценаріїв.

Ключові результати та їхнє значення:

1. Адаптивне керування трафіком у реальному часі. У результаті створення симуляційної моделі підтверджено, що алгоритм на основі MFG здатний оперативно формувати напрямок руху агентів залежно від поточної густини й перешкод у середовищі. Це дає змогу уникнути перевантаження маршруту й забезпечити безпечну маршрутизацію без потреби в людському втручанні.

2. Покращене уникнення конфліктів і перевантажень. Завдяки врахуванню колективної густини у функції вартості, MFG модель забезпечує розосередження агентів, що суттєво знижує максимальну локальну густину та, як наслідок, ризики зіткнень у зоні з високим трафіком.

3. Ефективне обходження заборонених зон. На відміну від rule-based моделі, яка не враховує обмеження, MFG забезпечує повну відмову від входження в зони з високою вартістю (перешкоди), що дає змогу використовувати її в критичних середовищах (наприклад, у зонах обмеженого доступу або поблизу інфраструктури).

4. Оптимізація часу й маршрутів з урахуванням взаємодії. Алгоритм на базі MFG мінімізує середню відстань до цілі та дає агентам змогу приймати рішення не ізольовано, а на основі загального стану простору, забезпечуючи кооперативне планування з дотриманням принципів мінімізації витрат.

5. Наукова новизна й перспективність підходу. Уперше у вітчизняних дослідженнях реалізовано порівняльну чисельну модель з використанням рівнянь Гамільтона-Якобі-Беллмана (HJB) та Фоккера-Планка (FP) у контексті моделювання БПЛА. Це створює основу для розширення моделі до 3D, мультигрупового й енергетично-обмеженого моделювання.

Обмеження моделі

Попри високі результати, запропонована модель має низку обмежень. По-перше, симуляція виконується у 2D-просторі без урахування висоти, що спрощує реальні аеродинамічні умови для БПЛА. По-друге, модель не враховує фізичних обмежень апаратів (максимальна швидкість, інерція, енергоспоживання), а також не моделює канали зв'язку чи дії в умовах утрат сенсорної інформації. Але для стратегічного рівня аналізу динаміки така спрощена модель дає змогу виявити ключові закономірності й оцінити потенціал децентралізованого керування. У подальших дослідженнях передбачається розширення до 3D-моделі з урахуванням реальних обмежень польоту, енергетичних ресурсів і комунікаційних факторів, що дасть змогу значно підвищити прикладну цінність розробки.

Перспективи застосування: результати дослідження демонструють, що модель масових ігор є перспективною для практичного застосування в задачах керування трафіком дронів, в умовах міського середовища з високою щільністю й складністю навколишнього простору.

Запропонована модель має перспективи:

1. Керування повітряними коридорами в містах. MFG може забезпечити гнучку маршрутизацію в реальному часі, уникаючи ділянок перевантаження й перешкод, що робить її придатною для автоматизованих систем UTM (UAV Traffic Management).

2. Інтеграція в розумну інфраструктуру міст. Модель може бути застосована в цифрових двійниках повітряного простору для симуляцій, передбачення навантаження й прийняття рішень щодо динамічного закриття або відкриття повітряних коридорів.

3. Основа для автономних колективних місій. Завдяки можливості колективного планування без централізованого управління, модель може бути використана для координації груп дронів, що виконують спільні завдання (пошук, доставка, моніторинг тощо).

4. Розширення на 3D-сценарії й енергетичні обмеження. У наступних дослідженнях модель мож-

на адаптувати до тривимірного простору, де додатково враховуватимуться енергоспоживання, вітрові умови та багаторівневе планування.

5. Інтеграція зі штучним інтелектом. Поєднання MFG з методами глибокого навчання або посиленого навчання (Reinforcement Learning) може забезпечити розумні динамічні стратегії для автономного керування великими мережами БПЛА.

Література

1. Wang G., Yao W., Zhang X., Li Z. A mean-field game control for large-scale swarm formation flight in dense environments. *Sensors.* 2022. Vol. 22, № 14. P. 5437. DOI: https://doi.org/10.3390/s22145437.
2. Chen D., Qi Q., Zhuang Z., Wang J. Mean field deep reinforcement learning for fair and efficient UAV control. *IEEE Internet of Things Journal.* 2020. Vol. 7, № 9. P. 8472–8485. DOI: https://doi.org/10.1109/JIOT.2020.3008299.
3. Chow J. Y. J. Dynamic UAV-based traffic monitoring under uncertainty as a stochastic arc-inventory routing policy. *International Journal of Transportation Science and Technology.* 2016. Vol. 5, № 3. P. 167–185. DOI: https://doi.org/10.1016/j.ijtst.2016.11.002.
4. Altaher M., Elmougy S. A new mean-field game approach for closed-loop strategic swarm maneuvering. *Proceedings of the International Telecommunications Conference (ITC-Egypt).* 2023. P. 1–6. DOI: https://doi.org/10.1109/ITC-Egypt58155.2023.10206155.
5. Schulmeyer N. P., Fala N. An agent-based modelling framework for assessing the impact of UAV on the air traffic system. *AIAA SciTech 2024 Forum.* 2024. DOI: https://doi.org/10.2514/6.2024-3889.
6. Izadi Moud H., Shojaei A., Flood I., Zhang X. Monte Carlo based risk analysis of unmanned aerial vehicle flights over construction job sites. *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO).* 2018. P. 451–458. DOI: https://doi.org/10.5220/0006868804510458.
7. Wang C.-H. J., Deng C., Low K. H. Parametric study of structured UTM separation recommendations with physics-based Monte Carlo distribution for collision risk model. *Drones.* 2023. Vol. 7, № 6. P. 345–358. DOI: https://doi.org/10.3390/drones7060345.
8. Tantardini M., Ieva F., Tajoli L., Piccardi C. Comparing methods for comparing networks. *Scientific Reports.* 2019. Vol. 9. P. 17557. DOI: https://doi.org/10.1038/s41598-019-53708-y.
9. Lin A. T., Fung S. W., Li W., Osher S. J. Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games. *Proceedings of the National Academy of Sciences (PNAS).* 2021. Vol. 118, № 31. e2024713118. DOI: https://doi.org/10.1073/pnas.2024713118.
10. De Paola A., Trovato V., Angeli D., Strbac G. A Mean Field Game Approach for Distributed Control of Thermostatic Loads Acting in Simultaneous Energy-Frequency Response Markets. *IEEE Transactions on Smart Grid.* 2019. Vol. 10, № 6. P. 5987–5999. DOI: https://doi.org/10.1109/TSG.2019.2895247.
11. Zhang J., Xu W., Gao H., Pan M., Han Z., Zhang P. Codebook-Based Beam Tracking for Conformal Array-Enabled UAV mmWave Networks. *IEEE Internet of Things Journal.* 2021. Vol. 8, № 1. P. 244–261. DOI: https://doi.org/10.1109/JIOT.2020.3005394.
12. Gao H., Lin A. T., Banez R. A., та ін. Belief and Opinion Evolution in Social Networks: A High-Dimensional Mean Field Game Approach. *Proceedings of ICC 2021 – IEEE International Conference on Communications.* 2021. DOI: https://doi.org/10.1109/ICC42927.2021.9500884.
13. Lasry J.-M., Lions P.-L. Mean field games. *Japanese Journal of Mathematics.* 2007. Vol. 2. P. 229–260. DOI: https://doi.org/10.1007/s11537-007-0657-8.
14. Lin A. T., Fung S. W., Li W., Osher S. APAC-Net: Alternating the Population and Agent Control via Two Neural Networks to Solve High-Dimensional Stochastic Mean Field Games. February 2020. DOI: https://doi.org/10.13140/RG.2.2.22346.52165.
15. Han Z., Niyato D., Saad W., Başar T., Hjørungnes A. Game Theory in Wireless and Communication Networks: Theory, Models, and Applications. Cambridge : Cambridge University Press, 2011. DOI: https://doi.org/10.1017/CBO9780511895043.
16. Achdou Y., Laurière M. Mean Field Games and Applications: Numerical Aspects. *Mean Field Games. Lecture Notes in Mathematics.* Springer, Cham, 2020. Vol. 2281. DOI: https://doi.org/10.1007/978-3-030-59837-2_4.
17. Caines P. E. Mean Field Games. Encyclopedia of Systems and Control. Springer, Cham, 2021. DOI: https://doi.org/10.1007/978-3-030-44184-5_30.
18. Park S., Kim H. T., Lee S., Joo H., Kim H. Survey on Anti-Drone Systems: Components, Designs, and Challenges. *IEEE Access.* 2021. Vol. 9. P. 42635–42659. DOI: https://doi.org/10.1109/ACCESS.2021.3065926.
19. Kim H., Park J., Bennis M., Kim S.-L. Massive UAV-to-Ground Communication and its Stable Movement Control: A Mean-Field Approach. *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* 2018. DOI: https://doi.org/10.1109/SPAWC.2018.8445906.

References

1. Wang, G., Yao, W., Zhang, X., & Li, Z. (2022). A mean-field game control for large-scale swarm formation flight in dense environments. *Sensors*, 22(14), 5437. https://doi.org/10.3390/s22145437.
2. Chen, D., Qi, Q., Zhuang, Z., & Wang, J. (2020). Mean field deep reinforcement learning for fair and efficient UAV control. *IEEE Internet of Things Journal*, 7(9), 8472–8485. https://doi.org/10.1109/JIOT.2020.3008299.
3. Chow, J. Y. J. (2016). Dynamic UAV-based traffic monitoring under uncertainty as a stochastic arc-inventory routing policy. *International Journal of Transportation Science and Technology*, 5(3), 167–185. https://doi.org/10.1016/j.ijtst.2016.11.002.
4. Altaher, M., & Elmougy, S. (2023). A new mean-field game approach for closed-loop strategic swarm maneuvering. In *Proceedings of the International Telecommunications Conference (ITC-Egypt)* (pp. 1–6). https://doi.org/10.1109/ITC-Egypt58155.2023.10206155.
5. Schulmeyer, N. P., & Fala, N. (2024). An agent-based modelling framework for assessing the impact of UAV on the air traffic system. In *AIAA Scitech 2024 Forum*. https://doi.org/10.2514/6.2024-3889.

6. Izadi Moud, H., Shojaei, A., Flood, I., & Zhang, X. (2018). Monte Carlo based risk analysis of unmanned aerial vehicle flights over construction job sites. In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO)* (pp. 451–458). https://doi.org/10.5220/0006868804510458.

7. Wang, C.-H. J., Deng, C., & Low, K. H. (2023). Parametric study of structured UTM separation recommendations with physics-based Monte Carlo distribution for collision risk model. *Drones*, 7(6), 345–358. https://doi.org/10.3390/drones7060345.

8. Tantardini, M., Ieva, F., Tajoli, L., & Piccardi, C. (2019). Comparing methods for comparing networks. Scientific Reports, 9, 17557. https://doi.org/10.1038/s41598-019-53708-y.

9. Lin, A. T., Fung, S. W., Li, W., & Osher, S. J. (2021). Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games. Proceedings of the National Academy of Sciences, 118(31), e2024713118. https://doi.org/10.1073/pnas.2024713118.

10. De Paola, A., Trovato, V., Angeli, D., & Strbac, G. (2019). A Mean Field Game Approach for Distributed Control of Thermostatic Loads Acting in Simultaneous Energy-Frequency Response Markets. IEEE Transactions on Smart Grid, 10(6), 5987–5999. https://doi.org/10.1109/TSG.2019.2895247.

11. Zhang, J., Xu, W., Gao, H., Pan, M., Han, Z., & Zhang, P. (2021). Codebook-Based Beam Tracking for Conformal Array-Enabled UAV mmWave Networks. IEEE Internet of Things Journal, 8(1), 244–261. https://doi.org/10.1109/JIOT.2020.3005394.

12. Gao, H., Lin, A. T., Banez, R. A., et al. (2021). Belief and Opinion Evolution in Social Networks: A High-Dimensional Mean Field Game Approach. Proceedings of ICC 2021 – IEEE International Conference on Communications. https://doi.org/10.1109/ICC42927.2021.9500884.

13. Lasry, J. M., & Lions, P. L. (2007). Mean field games. Japanese Journal of Mathematics, 2, 229–260. https://doi.org/10.1007/s11537-007-0657-8.

14. Lin, A. T., Fung, S. W., Li, W., & Osher, S. (2020). APAC-Net: Alternating the Population and Agent Control via Two Neural Networks to Solve High-Dimensional Stochastic Mean Field Games. https://doi.org/10.13140/RG.2.2.22346.52165.

15. Han, Z., Niyato, D., Saad, W., Başar, T., & Hjørungnes, A. (2011). Game Theory in Wireless and Communication Networks: Theory, Models, and Applications. Cambridge University Press. https://doi.org/10.1017/CBO9780511895043.

16. Achdou, Y., & Laurière, M. (2020). Mean Field Games and Applications: Numerical Aspects. In P. Cardaliaguet & A. Porretta (Eds.), Mean Field Games (Lecture Notes in Mathematics, Vol. 2281). Springer. https://doi.org/10.1007/978-3-030-59837-2_4.

17. Caines, P. E. (2021). Mean Field Games. In J. Baillieul & T. Samad (Eds.), Encyclopedia of Systems and Control. Springer. https://doi.org/10.1007/978-3-030-44184-5_30.

18. Park, S., Kim, H. T., Lee, S., Joo, H., & Kim, H. (2021). Survey on Anti-Drone Systems: Components, Designs, and Challenges. IEEE Access, 9, 42635–42659. https://doi.org/10.1109/ACCESS.2021.3065926.

19. Kim, H., Park, J., Bennis, M., & Kim, S.-L. (2018). Massive UAV-to-Ground Communication and its Stable Movement Control: A Mean-Field Approach. In 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE. https://doi.org/10.1109/SPAWC.2018.8445906.

Дата першого надходження рукопису до видання: 09.05.2025
Дата прийнятого до друку рукопису після рецензування: 16.06.2025
Дата публікації: 25.07.2025

УДК 004.4
DOI https://doi.org/10.36994/2788-5518-2025-01-09-15

ЗАСТОСУВАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРОГНОЗУВАННЯ НЕРІВНОМІРНОСТІ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ ПРИВАТНИМИ ДОМОГОСПОДАРСТВАМИ

Ксенич А. І., *к.т.н., Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. https://orcid.org/0009-0007-6569-951X, andrii.ksenych@nung.edu.ua*

Анотація. *У статті розглянуто актуальну проблему переходу на відновлювані джерела енергії в Україні, зокрема, в умовах військового стану, коли централізована енергосистема зазнає значного навантаження й пошкоджень. В умовах енергетичної нестабільності все більше приватних домогосподарств установлюють автономні сонячні електростанції (АСЕС) як джерело безперебійного живлення. Однак генерація електроенергії АСЕС, що має переважно денний профіль, не завжди відповідає нерівномірному добовому споживанню енергії споживачами, що породжує проблему узгодження попиту і пропозиції. Метою дослідження є розроблення підходу до прогнозування добової нерівномірності споживання електроенергії приватним домогосподарством з використанням інструментів штучного інтелекту, зокрема штучних нейронних мереж (ШНМ), реалізованих на основі бібліотеки Keras. Здійснено комплексний аналіз закономірностей електроспоживання на основі статистичних даних реального приватного домогосподарства за певний період. Розглянуто три варіанти класифікації добової нерівномірності, для кожного з них проведено математичне моделювання та розрахунок середньоквадратичної похибки прогнозування, що дало змогу оцінити точність обраних моделей. Результати показали, що найменша похибка прогнозу досягається при аналізі закономірностей у розрізі окремих днів тижня, про що свідчить наявність чітко вираженої поведінкової моделі споживання. У процесі дослідження обрано оптимальну архітектуру нейромережі, методи її навчання й параметри моделі, які забезпечують високу точність прогнозування. Запропонований підхід дає змогу формувати обґрунтовані прогнози щодо споживання електроенергії, що може бути використано як для підвищення енергоефективності домогосподарств, так і для оптимізації роботи автономних енергосистем і систем накопичення енергії. Результати дослідження мають практичну цінність для власників АСЕС, розробників систем «розумного дому» й учасників енергетичного ринку, що працюють у сфері децентралізованого енергопостачання.*

Ключові слова: *штучні нейронні мережі, Keras, автономні сонячні електростанції, електроспоживання.*

USING ARTIFICIAL NEURAL NETWORKS TO PREDICT ELECTRICITY CONSUMPTION BY PRIVATE HOUSEHOLDS

Andrii Ksenych, *Ph.D., Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. https://orcid.org/0009-0007-6569-951X, andrii.ksenych@nung.edu.ua*

Abstract. *The article considers the current problem of switching to renewable energy sources in Ukraine, in particular under martial law, when the centralized power system is experiencing significant load and damage. In conditions of energy instability, more and more private households are installing autonomous solar power plants (ASPPs) as a source of uninterrupted power supply. However, the generation of electricity by the ACES, which has a predominantly daily profile, does not always correspond to the uneven daily energy consumption by consumers, which creates the problem of matching supply and demand. The aim of the study is to develop an approach to forecasting the daily unevenness of electricity consumption by a private household using artificial intelligence tools, in particular artificial neural networks (ANNs), implemented on the basis of the Keras library. The article provides a comprehensive analysis of electricity consumption patterns based on statistical data from a real private household for a certain period. Three options for classifying daily unevenness are considered, and for each of them mathematical modeling and calculation of the mean square error of prediction are carried out, which made it possible to assess the accuracy of the selected models. The results showed that the smallest forecast error is achieved when analyzing patterns in the context of individual days of the week, as evidenced by the presence of a clearly expressed behavioral consumption model. In the process of research, the optimal architecture of the neural network, its training methods and model parameters were selected, which ensure high accuracy of prediction. The proposed approach allows to form reasonable forecasts of electricity consumption, which can be used both to increase the energy efficiency of households and to optimize the operation of autonomous power systems and energy storage systems. The results of the research have practical value for owners of ACES, developers of "smart home" systems and energy market participants working in the field of decentralized energy supply.*

Key words: *artificial neural networks, Keras, autonomous solar power plants, electricity consumption.*

Вступ

У більшості країн світу в останнє десятиліття спостерігається чіткий тренд енергетичної трансформації в бік зеленої генерації. Наприклад, у Європі впровадження відновлюваних джерел енергії (далі – ВДЕ), таких як сонячна енергія, є основою для досягнення цілей Європейського зеленого курсу, зокрема скорочення викидів парникових газів на 55% до 2030 року й досягнення кліматичної нейтральності до 2050 року [1]. Україна також слідує трендам розвитку сонячної генерації та інших ВДЕ, і це набуває особливого значення в умовах війни з росією, яка систематично знищує елементи критичної енергетичної інфраструктури. У цьому контексті все більше українців встановлюють автономні сонячні електростанції (далі – АСЕС) для своїх домогосподарств.

Порівнюючи переваги АСЕС з тими, що працюють за «зеленим тарифом», можна виділити низку основних:

– незалежність від енергомережі. СЕС за «зеленим тарифом» працюють лише за умови підключення до центральної мережі. У разі відключень електроенергії через обстріли чи аварії ці системи автоматично припиняють генерацію. АСЕС з акумуляторами забезпечує безперервну подачу енергії;

– акумулювання енергії для власних потреб. АСЕС оснащені системами зберігання енергії (акумуляторами), що дає змогу накопичувати енергію, вироблену вдень, і використовувати її вночі чи під час відсутності сонця;

– стійкість до війни та криз. В умовах знищення об'єктів енергетичної інфраструктури держава може обмежувати викуп електроенергії за «зеленим тарифом». АСЕС дає змогу уникнути залежності від таких обмежень;

– екологічна й енергетична безпека. Використання автономних СЕС зменшує навантаження на національну енергосистему, що сприяє стабілізації та безпеці її роботи в умовах криз.

Поряд із суттєвими перевагами автономні СЕС мають і низку недоліків. Зокрема, така електростанція генерує електроенергію нерівномірно, залежно від часу доби, погодних умов і пори року. Основна ж проблема полягає в невідповідності між генерацією енергії та її споживанням домогосподарством. Наведемо два основні виклики, з якими стикаються власники автономних СЕС:

– надлишок електроенергії вдень. У денний час, коли сонячна генерація є найвищою, виробляється значна кількість електроенергії, часто більше, ніж потрібно для поточного споживання домогосподарства. Якщо в цей час акумуляторні батареї заповнені, надлишок енергії не може бути ефективно використаний або збережений, що призводить до її втрати;

– недостатня генерація в похмурну погоду чи в нічний час. У похмурну погоду обсяг генерації електроенергії є кратно менший, ніж у сонячну, а в нічний час генерація відсутня повністю, тому домогосподарство змушене використовувати енергію, накопичену в акумуляторах протягом дня. Якщо батареї розряджені, може виникнути дефіцит енергії, особливо в умовах підвищеного вечірнього споживання (опалення, освітлення, побутові прилади).

Усунути ці недоліки можна шляхом підвищення ємності систем зберігання електроенергії в поєднані з використанням систем розумного будинку (Smart Home) для оптимізації споживання електроенергії та її оптимального перерозподілу протягом дня.

Виконання досліджень

Огляд доступних систем розумних будинків, таких як Home Assistant та OpenHAB, показав, що оптимізація споживання електроенергії реалізується користувачем шляхом прописування певних правил та умов, за якими енергосистема будинку змінює свій режим роботи [2; 3]. При цьому ці правила закладаються, виходячи з аналізу обсягу й нерівномірності споживання електроенергії протягом дня, тижня чи місяця. Таким чином, чим достовірніше можна спрогнозувати обсяги споживання та генерації електроенергії в розрізі дня чи тижня, тим ефективніше її можна буде використати й оптимально перерозподілити в зазначеному часовому проміжку.

Аналізуючи часові закономірності, які впливають на величину споживання електроенергії типовим домогосподарством, можемо виділити ключові:

– добова нерівномірність зумовлена закономірностями поведінки споживачів протягом доби (ранкові та вечірні піки споживання, які чергуються з денними й нічними спадами);

– тижнева нерівномірність. Зазвичай у будні дні сумарне споживання є меншим, оскільки більшість мешканців працює поза домом. У вихідні дні загальний рівень споживання є вищим у зв'язку з більшою активністю використання побутових приладів удома;

– місячна нерівномірність. Тут насамперед варто виділити сезонність – зимою споживання є більшим через опалення і триваліший період роботи освітлення, літом є зростання споживання через кондиціювання повітря, а весна/осінь має найнижче середнє споживання, оскільки менша потреба в кліматичному обладнанні. Також на зміну патернів споживання впливає наявність святкових днів і канікул.

Існує й безліч інших факторів, які впливають на обсяг і нерівномірність споживання енергії приватним будинком. Їх вплив і врахування шляхом математичного моделювання є недоцільним, оскільки часто вони мають випадковий характер і не піддаються моделюванню з достатньою вірогідністю.

У роботі пропонується використання нейронних мереж для прогнозування величини споживання електроенергії окремим домогосподарством.

Шляхом досліджень установлено, що середнє годинне споживання електроенергії протягом дня, залежно від місяця, тижня, дня й години, може бути прогнозовано з високою достовірністю нейронною мережею, яка має конфігурацію, представлену на рисунку 1.

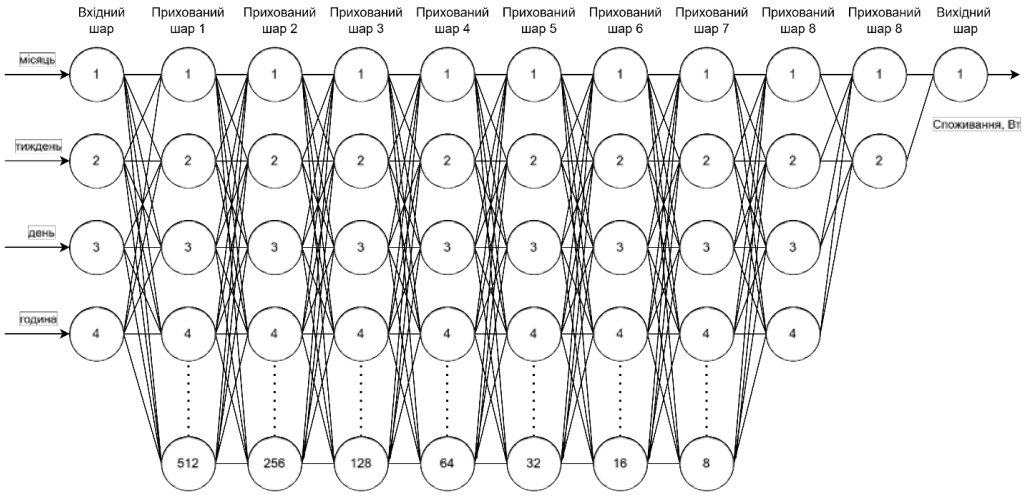


Рис. 1. Конфігурація нейронної мережі для прогнозування добової нерівномірності споживання електроенергії протягом місяця

Нейронна мережа має такі параметри:
– кількість шарів – 11 (вхідний, вихідний, 8 прихованих);
– функція активації для прихованих шарів – relu;
– алгоритм навчання – метод зворотного поширення помилки;
– кількість епох (циклів) навчання – 50000.

Для реалізації та навчання цієї моделі вибрано бібліотеку Keras, оскільки це високорівнева бібліотека для машинного навчання, яка працює поверх бібліотеки TensorFlow [4]. Вона надає зручний і зрозумілий інтерфейс для створення та тренування нейронних мереж.

Апробація моделі проведена на основі даних споживання реального домогосподарства, обладнаного автономною СЕС і розташованого в м. Івано-Франківську, Україна. За основу взято споживання електроенергії протягом грудня місяця 2024 р.

Закономірність споживання визначено шляхом аналізу й математичного моделювання для трьох варіантів нерівномірності:
– добова нерівномірність порівняно з усередненою добовою протягом місяця;
– добова нерівномірність у розрізі усередненої для певного дня тижня;
– добова нерівномірність у розрізі усередненої для будніх днів (понеділок-п'ятниця) і вихідних (субота-неділя) днів тижня.

Відхилення прогнозування добової нерівномірності визначено шляхом обчислення середньоквадратичної похибки за такою формулою:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i')^2, \quad (1)$$

де n – кількість точок, шт; Y_i – реальне годинне споживання, Вт/год; Y_i' – усереднене споживання, Вт/год.

За результатами дослідження встановлено, що для наведеного домогосподарства найменша квадратична похибка відповідає добовій нерівномірності для певного дня тижня, оскільки вона є найбільш вираженою протягом досліджуваного місяця. Тому саме цей варіант усереднених величин енерговитратності будинку в подальшому використаний для навчання штучної нейронної мережі й наведений графічно на рисунку 2.

Шляхом моделювання та навчання отримано нейронну мережу, яка з високою достовірність прогнозує середньогодинне споживання електроенергії для досліджуваного домогосподарства. Як приклад, на рисунку 3 подано результати порівняння отриманих значень для одного дня – понеділка.

Як видно з рисунка 3, прогнозоване значення з високою достовірністю збігається з реальним споживанням на всьому часовому проміжку доби, оскільки середньоквадратичне відхилення для цього варіанта становить менше ніж 5 (Вт·год)^2. Аналогічна картина й для інших днів тижня. Це свідчить про те, що запропонована архітектура нейромережі, методи її навчання й параметри моделі забезпечують високу точність прогнозування добової нерівномірності споживання електроенергії приватним домогосподарством.

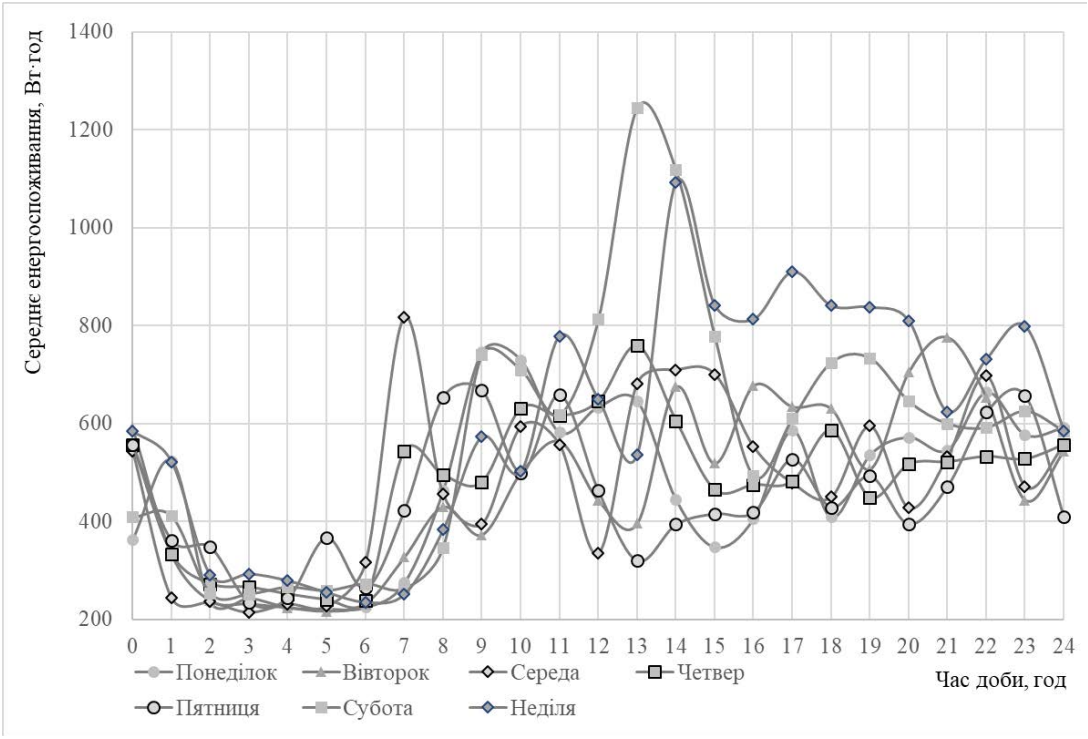


Рис. 2. Результати усередненого енергоспоживання приватного будинку протягом доби в розрізі днів тижня

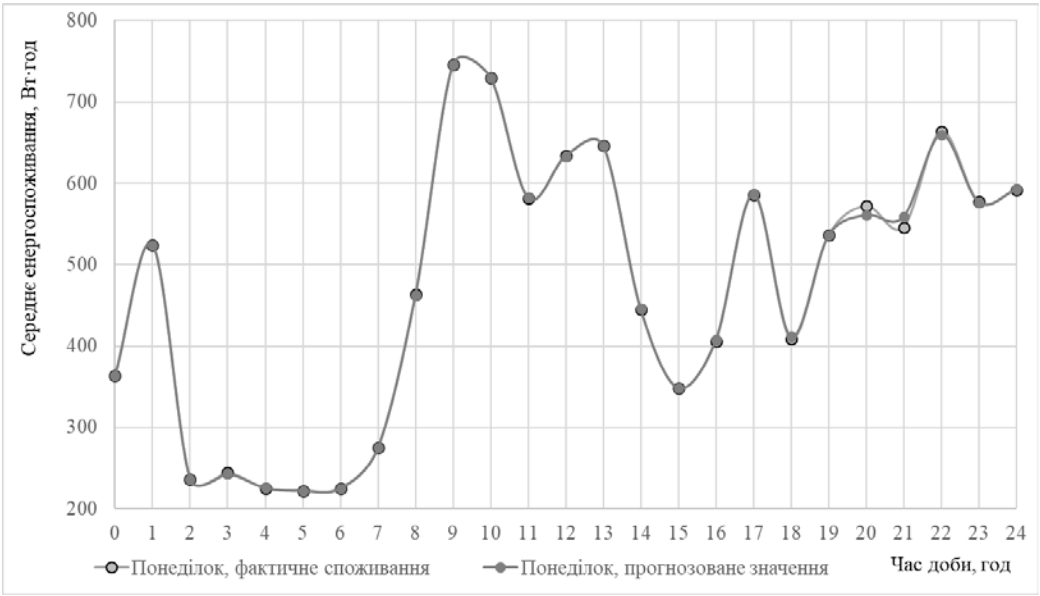


Рис. 3. Результати порівняння фактичних і прогнозованих нейромережею значень середненого енергоспоживання будинку для дня тижня – понеділка

Висновок

У ході дослідження розроблено й апробовано підхід до прогнозування добової нерівномірності енергоспоживання приватного домогосподарства з використанням штучних нейронних мереж. Зокрема, установлено конфігурацію, методи та параметри навчання штучної нейронної мережі, яка з високою достовірністю прогнозує усереднене споживання електроенергії приватним домогосподарством на основі статистичних даних споживання за минулий час. Найвища точність такого прогнозування досягнута при класифікації споживання в розрізі днів тижня, що свідчить про вплив споживчих

звичок домогосподарств і необхідність урахування цього чинника при розробці моделей. Отримані результати підтвердили ефективність застосування інструментів штучного інтелекту, зокрема бібліотеки Keras, для моделювання складних поведінкових закономірностей енергоспоживання.

Запропонований підхід має практичну значущість в умовах енергетичної нестабільності, зумовленої військовими діями в Україні, і дає змогу більш ефективно інтегрувати автономні сонячні електростанції в енергетичну систему домогосподарств. Достовірне прогнозування споживання дає можливість краще оптимізувати роботу систем накопичення енергії, підвищити рівень енергоефективності й забезпечити безперебійне енергопостачання в критичних умовах.

Література

1. Учасники проектів Вікімедіа. Європейський зелений курс – Вікіпедія. *Вікіпедія*. URL: https://uk.wikipedia.org/wiki/ Європейський_зелений_курс (дата звернення: 30.01.2025).
2. Documentation. *Home Assistant*. URL: https://www.home-assistant.io/docs/ (дата звернення: 30.01.2025).
3. Introduction. *openHAB*. URL: https://www.openhab.org/docs/ (дата звернення: 30.01.2025).
4. Keras documentation: Keras 3 API documentation. *Keras: Deep Learning for humans*. URL: https://keras.io/api/ (дата звернення: 30.01.2025).

References

1. Wikimedia project participants. European Green Deal – Wikipedia. Wikipedia. URL: https://en.wikipedia.org/wiki/European_Green_Deal (accessed: 01.30.2025).
2. Documentation. Home Assistant. URL: https://www.home-assistant.io/docs/ (accessed: 01.30.2025).
3. Introduction. openHAB. URL: https://www.openhab.org/docs/ (accessed: 01.30.2025).
4. Keras documentation: Keras 3 API documentation. Keras: Deep Learning for humans. URL: https://keras.io/api/ (accessed: 01.30.2025).

Дата першого надходження рукопису до видання: 14.05.2025
Дата прийнятого до друку рукопису після рецензування: 10.06.2025
Дата публікації: 25.07.2025

УДК 004.9:681.3
DOI https://doi.org/10.36994/2788-5518-2025-01-09-16

APPLICATION OF MONTE CARLO METHODS FOR MODELING AND PREDICTING USER BEHAVIOR ON SOCIAL MEDIA

Mysiuk I. V., *PhD student, Ivan Franko National University of Lviv, Lviv, Ukraine. https://orcid.org/0000-0002-3641-4518, iruna.musyk8a@gmail.com*

Mysiuk R. V., *Ph.D, Ivan Franko National University of Lviv, Lviv, Ukraine. https://orcid.org/0000-0002-7843-7646, mysyukr@ukr.net*

Abstract. *Understanding and predicting user behavior on social media is challenging due to the dynamic nature of users' online interactions. Monte Carlo methods are one of the methods for modeling user engagement, content virality, sentiment evolution and personalized recommendations in dynamic environments.*

The article investigates the application of Monte Carlo methods for modeling and predicting user behavior in social networks. Several variants of this approach are considered, including the quasi-Monte Carlo, the classical Monte Carlo method, and the Metropolis-Hastings methods, which demonstrated better efficiency for the selected data set due to more uniform coverage of user behavior scenarios, taking into account the selected parameters of user activities. Convergences by Monte Carlo methods with a sample of 10000 values are also displayed, and based on the main selected parameters for modeling user behavior, namely likes, shares and comments, the effectiveness of the selected methods with such a data set with many parameters is studied.

Based on the data obtained, a basic predictive analytics system was developed, which can be further developed through integration into recommendation systems to increase content personalization and improve the analysis of user activities in social networks. In addition, the VADER lexical analyzer was used to assess the emotional tone of users and their virality on social networks was investigated. The results obtained allowed us to display the distribution of likes, comments and expansions in social networks and to carry out predictive analytics, which can be integrated into recommender systems based on the results of modeling. The process of modeling virality and expansion of content is also shown and the principle of operation of the program is schematically presented. The results of the performance evaluation confirm the feasibility of using a combination of quasi-Monte Carlo methods and show their effectiveness and accuracy of prediction in the analysis of user activity for a comprehensive study of behavior in social networks. Virality prediction reflects the spread of viral content.

Key words: *Monte Carlo simulation, social media analytics, user behavior modeling, machine learning, predictive analytics.*

ЗАСТОСУВАННЯ МЕТОДІВ МОНТЕ-КАРЛО ДЛЯ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ПОВЕДІНКИ КОРИСТУВАЧІВ У СОЦІАЛЬНИХ МЕРЕЖАХ

Ірина Мисюк, *аспірантка, Львівський національний університет імені Івана Франка, Львів, Україна. https://orcid.org/0000-0002-3641-4518, iruna.musyk8a@gmail.com*

Роман Мисюк, *д.ф., Львівський національний університет імені Івана Франка, Львів, Україна. https://orcid.org/0000-0002-7843-7646, mysyukr@ukr.net*

Анотація. *Розуміння та прогнозування поведінки користувачів у соціальних мережах є складним завданням через динамічний характер онлайн-взаємодій користувачів. Методи Монте-Карло є одним із методів моделювання залученості користувачів, вірусності контенту, еволюції настроїв і персоналізованих рекомендацій у динамічних середовищах.*

У статті досліджено застосування методів Монте-Карло для моделювання та прогнозування поведінки користувачів у соціальних мережах. Розглянуто декілька варіантів цього підходу, серед яких – квазі-Монте-Карло, класичний метод Монте-Карло й Метрополіс-Гастингс методи продемонстрували кращу ефективність для обраного набору даних завдяки рівномірнішому охопленню сценаріїв поведінки користувачів, ураховуючи обрані параметри їхніх активностей. Також відображено збіжності методами Монте-Карло з вибіркою у 10000 значень і на основі основних обраних параметрів для моделювання поведінки користувачів, а саме лайків, поширень і коментарів, досліджено ефективність обраних методів при такому наборі даних з багатьма параметрами.

Базуючись на отриманих даних, розробили базову систему прогнозної аналітики, яка може отримати подальший розвиток завдяки інтеграції в рекомендаційні системи для підвищення персо-

налізації контенту й покращення аналізу активностей користувачів у соціальних мережах. Крім того, застосовано лексичний аналізатор VADER для оцінювання емоційного тону користувачів і досліджено їх вірусність у соціальних мережах. Отримані результати дали змогу відобразити розподіл лайків, коментарів і поширень у соціальних мережах. Також відображено процес моделювання вірусності й розповсюдження контенту, схематично подано принцип роботи програми. Результати оцінок ефективності підтверджують доцільність використання комбінації квазі-Монте-Карло методів і показують їх ефективність, точність прогнозування при аналізі активностей користувачів для комплексного дослідження поведінки в соціальних мережах. Прогнозування вірусності відображає поширення вірусного контенту.

Ключові слова: *моделювання методом Монте-Карло, аналітика соціальних мереж, моделювання поведінки користувачів, машинне навчання, прогнозна аналітика.*

Introduction

Social media platforms have become almost the main method of communication in the development of digital communication, influencing public opinion, marketing strategies and online interactions, etc. Modeling user behavior using digital platforms is crucial for businesses, researchers and policymakers looking to predict engagement trends, optimize content delivery and identify anomalies such as misinformation or bot activity and match users with better content.

However, user interactions, such as likes, shares, comments, and their changing moods, are inherently uncertain and depend on various external factors, making it difficult to make more accurate predictions when analyzing social media or displaying incompatible content on digital platforms. The Monte Carlo method is widely used in probabilistic modeling and analysis of results and offers a powerful approach to modeling behavior of social media users. By generating multiple results based on probabilistic distributions, Monte Carlo modeling allows you to estimate, for example, the probability of different patterns of interaction, the spread of viral content, and changes in user sentiment over time. This method will help eliminate uncertainty in the dynamics of social media, making it significant in forecasting and analyzing data, recommendation systems, and fraud detection. This article explores the application of the method Monte Carlo for modeling and predicting user behavior in social networks, as well as highlighting the advantages and problems of using probabilistic modeling in the dynamic environment of social networks. By using the Monte Carlo method, it is possible to improve decision-making in the platform's content strategy, advertising, and moderation, ultimately improving user experience and online security. Predictive analytics and modeling allows you to optimize your posting time for maximum user engagement for social media marketers and analysts.

Social media user behavior modeling research mostly focuses on machine learning, statistical modeling, and relationship analysis in graph structures. However, a small number of studies have used probabilistic methods, including Monte Carlo methods, for some cases of predicting patterns of interaction, dissemination of information, and mood dynamics in social networks. In general, the areas studied in this area can be divided into the use of probabilistic models for the analysis of social networks, Monte Carlo methods in predicting social networks, the study of user behavior and network dynamics and application in recommendation systems and personalization.

The first group of studies from the above used probabilistic approaches to analyze social media data. For example, Bayesian networks and hidden Markov models (HMM) were used to simulate the evolution of user sentiments and online discussions, in particular in the papers [1; 2]. Similarly, Monte Carlo Markov chain techniques (MCMCs) have been used to estimate the probability of specific user behaviors, such as content distribution and the likelihood of interaction at work [3]. As a result, the methods described allow us to investigate the stochastic nature of online interactions and improve behavioral predictions, which is important for general understanding.

The second group is the Monte Carlo methods in social network forecasting, used to estimate uncertainty in social media analytics. Studies [4] have demonstrated their effectiveness in predicting content virality by modeling different distribution scenarios based on historical interaction data. The papers [5; 6] integrated Monte Carlo methods with deep learning to refine trend predictions in social networks, particularly in dynamic environments where deterministic models face difficulties due to rapid changes in content.

The next group of related works can be attributed to studies of user behavior and network dynamics, namely how users interact with content on social networks using stochastic and networked approaches. Agent-based modeling and Monte Carlo simulations were combined to study cascades of information and the role of influencers in content distribution as implemented in the work [7].

Regarding the use in recommendation systems and personalization, the paper [8; 9] shows that Monte Carlo-based reinforcement methods can improve adaptive recommendation systems by predicting how users will react to different types of content.

Presentation of the main material and discussion of the results

To collect potential data for the study, publicly available social media datasets were selected, including user interaction (likes, shares, comments) and content functions (text, images, and metadata), and the time of publication was performed based on [10]. The main purpose of the data collected is to display

user engagement metrics across different types of publications, including news articles, user-generated entertainment content.

The collected data has been pre-processed to remove irrelevant information, such as spam or duplicate posts. Also, emotions or empty comments appear on social networks most often. Data preprocessing provides a clean and representative set of real-world data that is well standardized for display and modeling. For example, the collected dataset on the social network Instagram mostly contains collections of photos and comments, likes and shares, that is, engagement indicators that are associated with popular hashtags. The data obtained from the social network Facebook contains publications from public pages and groups that are not combined with each other by type of content, but are active in user interaction with publications.

The Monte Carlo method was used to model user behavior under different conditions and aspects of user behavior. The Monte Carlo method involves the use of not only based simulations of user activities in social networks, but also improved approaches of this method using Markov chains. Basic Monte Carlo simulations estimate engagement levels by generating random samples using probability distributions. When simulating more advanced methods. For example, using the Monte Carlo method with Markov chains (MCMC). This method is used to simulate complex user behavior and check the virality of content.

In Fig. 1 it is shown that the data collected by likes, shares on comments from the social network Instagram and Facebook, as the main parameters of engagement when modeling user behavior. This parameter allows you to simulate user interactions for posts with different content characteristics and time of publication dependence, which shows the amount of potentially engaged for this type of content based on a random sample of engagement distributions of previous historical data.

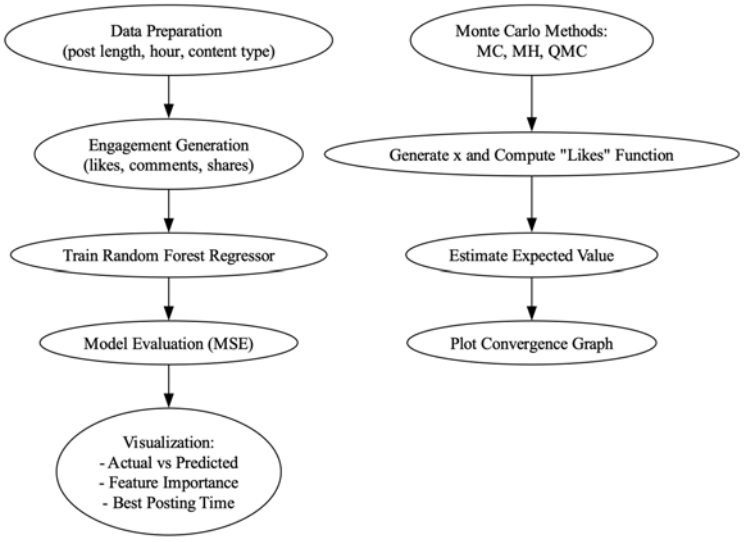


Fig. 1. Scheme of the program

The paper uses popular Python libraries, including NumPy and Pandas for generating and processing synthetic data, as well as Matplotlib for visualizing the results. Scikit-learn was used to build and evaluate the machine learning model, including the Random Forest Regressor algorithm and the standard error (MSE) metric.

In the first step, a dataset [10] is created that simulates social posts with three characteristics: length, time of publication, and type of content. Based on these parameters, interaction values are formed – likes, comments, and shares – with the addition of random noise. The built Random Forest Regressor model learns to predict the number of likes using these parameters. The accuracy of the prediction is estimated using the standard error. Additionally, graphs are created that visualize the correspondence between actual and predicted likes, the importance of signs and the optimal time to publish a post. The second stage analyzes three methods for estimating the expected number of likes: classic Monte Carlo, Metropolis-Hastings and Quasi-Monte Carlo. The classic Monte Carlo is simple and fast, but less accurate. Metropolis-Hastings is more accurate, but slower due to the use of Markov chains. Quasi-Monte Carlo provides the best balance between accuracy and efficiency through the use of Sobol sequences.

The number of values for research was chosen to be 10000. Organic user engagement and focused importance sampling, the basic Monte Carlo method, the improved Monte Carlo method using Markov chains, the Metropolis-Hastings and Quasi-Monte Carlo method are also displayed.

At the same time, the value of the error in each of the methods was significantly correlated. The estimation results show the values obtained by three different methods: the conventional Monte Carlo method (MC), the

Metropolis-Hastings method (MH), and the quasi-Monte Carlo (QMC) method. In this case, the convergence of methods is shown in Fig. 2. The MC method gave a score of 0.4064 and worked very quickly (0.0017 seconds). This shows that it is effective in simple tasks, although its accuracy depends on the number of simulations. The Metropolis-Hastings method, although it belongs to the improved stochastic methods, showed a significantly higher score (0.7124) and worked the longest (0.0888 seconds). This may indicate an insufficient number of iterations or an imperfect setting of the probabilistic transition, resulting in a less accurate result. This method requires careful adjustment to ensure stability and convergence to the correct solution. The quasi-Monte Carlo (QMC) method showed a score of 0.3929, which is close to the result of a conventional MC, but at the same time shows slightly higher stability. Although it took longer to complete (0.017 seconds), it often provides a lower variance in estimates, especially with fewer simulations.

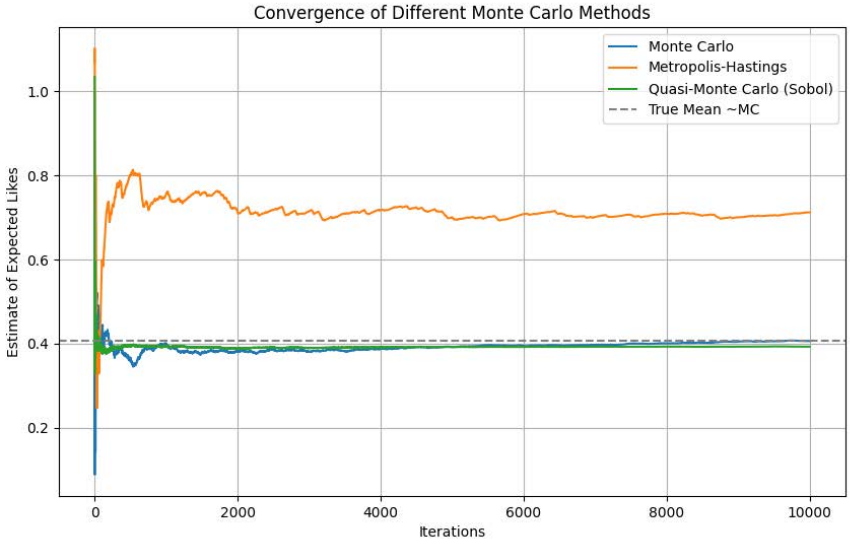


Fig. 2. Display of convergence by Monte Carlo methods (10000 values)

The next parameter that helps assess the likelihood of viral content due to active distribution and user engagement at the initial levels is the virality of the content.

Another parameter that allows you to check the dependence of sentiments on a certain topic or post are popularized over time, taking into account the activities of users, is the evolution of sentiments.

The simulation was carried out using the Monte Carlo method in the Python programming language. In addition, several basic NumPys were used to randomly sample the data and SciPy to statistically analyze the data obtained. And directly for data visualization, the most popular library matplotlib was used, which provides and provides an integrated approach to modeling uncertainty and predicting social media trends.

Data for modeling user behavior was obtained over 10,000 iterations to account for a wide range of possible outcomes and provide reliable statistical reliability. The user behavior model was built on the basis of a probabilistic structure that took into account user interaction and the engagement model from the distribution with the parameters evaluated from the dataset. The process of modeling content diffusion is used to investigate branching in the distribution of a publication and is determined by the level of engagement and its impact on the social network. Sentiment tracking simulation is also carried out. When tracking changes in user sentiment over time, sentiment analysis is applied on pre-trained models such as VADER (Valence Aware Dictionary and sEntiment Reasoner) to analyze the textual content of posts and comments. At the same time, the sentiment score is updated at each step of the simulation based on user activities (likes and comments, shares) to reflect the emotional tone of user interactions in social networks.

Modeling virality and content distribution (Fig. 3) is carried out by changing the initial number of interactions and the number of users and determining the factors that are responsible for the spread of viral content on social networks. Also, from the collected data, you can determine the dynamics of user sentiment among the provided post texts. Then the resulting parameter shows the evolution of sentiment regarding a certain trend (negative area of the Sentiment Score is negative, positive area is positive) or topics in social networks, taking into account external influences. For example, how holidays in Ukraine or marketing campaigns or news affect sentiment and how they fluctuate depending on user behavior and engagement.

Fig. 3. Displaying the distribution of content: likes, comments and shares

Modeling virality and content distribution is done by changing the initial number of interactions and the number of users and determining the factors that are responsible for the spread of viral content on social networks. Also, from the collected data, you can determine the dynamics of user sentiment. Then the resulting parameter will show the evolution of sentiment regarding a particular trend or topic in social networks, taking into account external influences. For example, how holidays in Ukraine or marketing campaigns or news affect sentiment and how they fluctuate depending on user behavior and engagement.

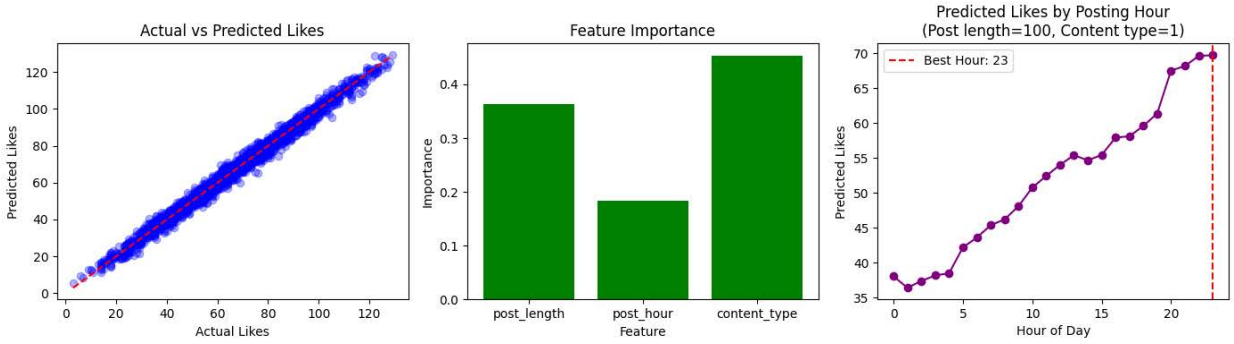


Fig. 4. Display of the Monte Carlo Efficiency Assessment

To evaluate the effectiveness of Monte Carlo modeling in predicting user behavior, as can be seen in Fig. 4, several performance indicators were used, such as: prediction accuracy, virality prediction, and exposure threshold. Prediction accuracy reflects the model's ability to predict actual engagement levels (e.g., the number of likes, shares, or comments) compared to real data, and virality prediction shows the accuracy of propagation modeling viral content, measured by the percentage of content that exceeded a predetermined engagement level. First, a dataset of 10.000 entries from prepared publications [10] and their interaction metrics such as likes, comments, and shares are presented. Next, this data is broken down into training and test samples to train the model. The Random Forest Regressor algorithm is used as a model, which learns to predict the number of likes. After training, the quality of the model is evaluated using the standard error (MSE), which shows the accuracy of the predictions. To better understand the results, three graphs are constructed: a comparison of actual and predicted likes, the importance of traits in the model, and recommendations for the best time to publish a post. The recommendation of the time of publication is based on the model's forecasts for different hours of the day, taking into account the fixed length of the post.

Conclusion

As part of the study, several variants of the Monte Carlo method were applied to model and predict user behavior in social networks. In particular, quasi-Monte Carlo methods turned out to be quasi-optimal for the selected dataset – they provided higher accuracy and stability of the results due to the reduction of sample

variance and more uniform coverage of the space of possible scenarios. Basic predictive analytics was developed, which, based on the results of simulations, can be integrated into recommendation systems. Such a system is potentially able to adapt the content delivery according to the intended level of user engagement, which opens up opportunities for personalizing marketing, optimizing news delivery and improving interaction with the audience. To analyze the emotional tone and identify the potential virality of the content, the lexical tool VADER (Valence Aware Dictionary and sEntiment Reasoner) was used, which made it possible to effectively classify the sentiments of messages in a large sample. This made it possible to better model the evolution of community sentiment and its impact on the dissemination of information. In general, the results of the study confirm the feasibility of using Monte Carlo methods, especially their quasi-deterministic variants, for a comprehensive analysis of user behavior in social networks.

References

1. Zheng, C. (2025). A comprehensive review of probabilistic and statistical methods in social network sentiment analysis. Advances in Engineering Innovation, 16(3), P. 38–43. DOI: https://doi.org/10.54254/2977-3903/2025.21918.

2. Li, H., Wang, Y., Zhang, J., & Liu, F. (2022). A Modified Attention Mechanism Powered by Bayesian Network for User Activity Analysis and Prediction. *Information Sciences, 603*, P. 115–130. DOI: https://doi.org/10.1016/j.ins.2022.02.003.

3. Wang, X., & Li, Q. (2022). Personality Trait Identification Based on Hidden Semi-Markov Model in Online Social Networks. *Proceedings of the 2022 7th International Conference on Intelligent Information Technology*, P. 30–35. DOI: https://doi.org/10.1145/3524889.3524898.

4. Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Composite Monte Carlo Decision Making under High Uncertainty of Novel Coronavirus Epidemic Using Hybridized Deep Learning and Fuzzy Rule Induction. arXiv preprint arXiv:2003.09868. DOI: https://doi.org/10.48550/arXiv.2003.09868.

5. Zhang, J. (2020). Modern Monte Carlo Methods for Efficient Uncertainty Quantification and Propagation: A Survey. arXiv preprint arXiv:2011.00680. DOI: https://doi.org/10.48550/arXiv.2011.00680.

6. Sosa, J., & Martínez, C. (2025). Bayesian Sociality Models: A Scalable and Flexible Alternative for Network Analysis. arXiv preprint arXiv:2503.14697. DOI: https://doi.org/10.48550/arXiv.2503.14697.

7. Garibay, I. et al. (2021). Deep Agent: Studying the Dynamics of Information Spread and Evolution in Social Networks. In: Yang, Z., von Briesen, E. (eds) *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*. CSSSA 2020. Springer Proceedings in Complexity. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-77517-9_11.

8. Meshram, R., & Kaza, K. (2021). Monte Carlo Rollout Policy for Recommendation Systems with Dynamic User Behavior. *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, P. 86–89. DOI: https://doi.org/10.1109/COMSNETS51098.2021.9352741.

9. Tommasel A., Diaz-Pace J. A. (2024). Leveraging Monte Carlo Tree Search for Group Recommendation. In Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24). *Association for Computing Machinery*, New York, NY, USA, P. 1136–1141. DOI: https://doi.org/10.1145/3640457.3691713.

10. Mysiuk I., Mysiuk R., Shuvar R. (2023). Collecting and analyzing news from newspaper posts in Facebook using machine learning. *Artificial Intelligence*. 28(1), P. 147–154 DOI: https://doi.org/10.15407/jai2023.01.147.

Дата першого надходження рукопису до видання: 26.05.2025
Дата прийнятого до друку рукопису після рецензування: 24.06.2025
Дата публікації: 25.07.2025

УДК 004.89
DOI https://doi.org/10.36994/2788-5518-2025-01-09-17

ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ ПІДТРИМКИ ВИКЛАДАЧА ПІД ЧАС ФОРМУВАННЯ ТЕСТІВ З МАТЕМАТИЧНИМ СКЛАДНИКОМ

Одрібець Н. В., *к.ф.-м.н., Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0008-0176-1211, sosn@ukr.net*

Одрібець С. П., *к.ф.-м.н., Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0002-2522-5794, s.odribets@gmail.com*

Богун Р. І., *к.ф.-м.н., Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0002-8519-2234, bohun.roma@gmail.com*

Анотація*. У статті розглянуто підхід до проєктування інтелектуальної системи підтримки викладача в процесі формування тестових завдань для навчальних дисциплін з математичним складником. Основну увагу приділено розробленню архітектури програмного рішення, що забезпечує поєднання сучасних мовних моделей з шаблонами, сформованими відповідно до рівнів таксономії Блума. Такий підхід дає змогу автоматизовано створювати тести з урахуванням рівня складності й форми відповіді.*

Запропонована система включає п'ять основних компонентів: модуль генерації запитань з підтримкою таксономії Блума, інтегрований LLM-модуль для обробки природної мови, математичний модуль для роботи з формулами у форматі LaTeX і перевірки еквівалентності математичних виразів, вебінтерфейс на базі Streamlit і сервіс експорту у формати LMS.

Окремо описано реалізацію модуля генерації математичних виразів і формул з використанням мови розмітки LaTeX, що забезпечує зручну візуалізацію контенту для викладача й студента. Обробка відкритих відповідей, які передбачають уведення формул або виразів, реалізована за допомогою бібліотеки символьних обчислень SymPy, що дає змогу здійснювати автоматизовану перевірку математичної коректності відповідей. Для побудови інтерфейсу системи генерації тестів використано фреймворк Streamlit, що забезпечує інтерактивну взаємодію та простоту розгортання. Запропонована система передбачає інтеграцію з популярними освітніми платформами, зокрема Moodle, і підтримує експорт згенерованих завдань у відповідні формати імпорту.

Проведено експериментальну перевірку функціональних можливостей системи в навчальному процесі, яка засвідчила підвищення ефективності оцінювання, скорочення часу на підготовку завдань і зменшення навантаження на викладачів. Особливу цінність система становить в умовах дистанційного та змішаного навчання, де автоматизація процесів оцінювання є критично важливою для забезпечення якості освіти.

Ключові слова: *інтелектуальні системи, генерація тестів, вища математика, мовні моделі, Streamlit, Moodle, таксономія Блума.*

INTELLIGENT TECHNOLOGIES FOR SUPPORTING TEACHERS IN THE FORMATION OF TESTS WITH MATHEMATICAL COMPONENT

Nataliia Odribets*, Ph.D., Ass. Prof., Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0008-0176-1211, sosn@ukr.net*

Serhii Odribets*, Ph.D., in Physical and Mathematical Sciences, Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0002-2522-5794, s.odribets@gmail.com*

Roman Bohun*, Ph.D., Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0002-8519-2234, bohun.roma@gmail.com*

Abstract. *The article considers an approach to designing an intelligent support system for a teacher in the process of creating test tasks for academic disciplines with a mathematical component. The main focus is on the development of a software solution architecture that combines modern language models with templates formed according to the levels of Bloom's taxonomy. This approach allows for the automated creation of tests, taking into account the level of difficulty and the form of the answer.*

The proposed system includes five main components: a question generation module with support for Bloom's taxonomy, an integrated LLM module for natural language processing, a mathematical module for working with formulas in LaTeX format and checking the equivalence of mathematical expressions, a web interface based on Streamlit, and a service for exporting to LMS formats.

Separately, the implementation of the module for generating mathematical expressions and formulas using the LaTeX markup language is described, which ensures convenient content visualization for the teacher and student. The processing of open-ended answers, which involve the input of formulas or expressions, is implemented using the SymPy symbolic computation library, which allows for the automated verification of the mathematical correctness of the answers. To build the interface of the test generation system, the Streamlit framework was used, which provides interactive interaction and ease of deployment. The proposed system provides for integration with popular educational platforms, in particular Moodle, and supports the export of generated tasks to the appropriate import formats.

An experimental verification of the system's functional capabilities in the educational process was carried out, which showed an increase in the efficiency of assessment, a reduction in the time for preparing tasks, and a decrease in the workload of teachers. The system is of particular value in the conditions of distance and blended learning, where the automation of assessment processes is critically important for ensuring the quality of education.

Key words: *intelligent systems, test generation, higher mathematics, language models, Streamlit, Moodle, Bloom's taxonomy.*

Вступ

Сучасні підходи до оцінювання знань студентів потребують не лише формального контролю, а й гнучких інструментів, здатних адаптуватися до змісту дисципліни, рівня підготовки здобувачів і цілей навчання. Особливо це стосується дисциплін математичного спрямування, де тестові завдання мають специфічну структуру: містять формули, графіки, аналітичні викладки та відкриті відповіді. Ручне створення таких завдань – складний і трудомісткий процес, що вимагає високої кваліфікації викладача, часу на підготовку й перевірку, а також знання технічних засобів подання математичних формул.

Застосування інтелектуальних технологій дає змогу автоматизувати рутинні етапи цього процесу, надаючи викладачеві інструменти для генерації, редагування й перевірки тестів у напівавтоматичному режимі. Особливої актуальності така автоматизація набуває в умовах дистанційного або змішаного навчання, де освітні платформи типу Moodle, Google Classroom або Canvas використовуються як основне середовище взаємодії з аудиторією.

У статті описано розроблення й дослідження системи, що поєднує можливості великих мовних моделей, інтерфейсних бібліотек на зразок Streamlit, обчислювальних інструментів для роботи з математичними виразами, а також підтримку різних форматів інтеграції з навчальними платформами. Система дає змогу не лише генерувати тестові завдання відповідно до рівнів таксономії Блума, а й формувати завдання з математичним змістом, здійснювати валідацію відповідей та експортувати завдання в придатному форматі для LMS.

Виконання досліджень

Інтелектуальна система підтримки викладача під час формування тестів з математичним складником побудована за модульною архітектурою, що забезпечує гнучкість, масштабованість і можливість інтеграції з різними освітніми платформами.

Основні компоненти такі:

1. Модуль генерації запитань.

– Формує тестові завдання на основі заданої теми, дисципліни та рівня складності.

– Підтримує шаблони відповідно до таксономії Блума.

– Генерує як текстові, так і формульні завдання для курсів з вищої математики, теорії ймовірності, інформаційних технологій, інтелектуального аналізу даних тощо.

2. Мовна модель (LLM) + LangChain.

– Забезпечує генерацію природномовних інструкцій, питань і відповідей.

– Використовує механізми семантичної фільтрації та перевірки відповідей.

3. Математичний модуль.

– Виконує генерацію формул, їх відображення у форматі LaTeX.

– Застосовує бібліотеку SymPy для спрощення виразів і перевірки еквівалентності (наприклад, $x^2 + x$ і $x(x+1)$).

4. Інтерфейс користувача.

– Реалізований на основі Streamlit.

– Дає викладачеві змогу задавати параметри генерації, переглядати, редагувати й експортувати завдання.

– Підтримує візуалізацію формул, табличну організацію та фільтрацію запитань.

5. Сервіс експорту в LMS.

– Генерує файли у форматах XML (Moodle), PDF, Word.

– Дає змогу групувати завдання за рівнем складності, темою чи типом запитання.

Таксономія освітніх цілей Бенджаміна Блума широко використовується для класифікації рівнів пізнавальної активності студентів. У контексті генерації тестів вона дає можливість розподілити

завдання за складністю й визначити мету кожного запитання: від простого відтворення факту до оцінювання або синтезу нової інформації.

У системі реалізовано розмежування завдань на шість рівнів:

1. Запам'ятовування: відтворення визначень, формул, властивостей (наприклад, означення похідної або правила добутку).
2. Розуміння: інтерпретація графіків, словесний опис математичних явищ.
3. Застосування: обчислення значень за формулою, підстановка чисел, вирішення типових задач.
4. Аналіз: порівняння розв'язків, виявлення помилок, розклад виразів на множники.
5. Оцінювання: перевірка правильності логіки міркувань, аргументація вибору методу.
6. Створення: побудова математичних моделей, самостійне формулювання умов задач.

Взаємодіючи із сучасними генеративними мовними моделями, система може генерувати запитання із заданим рівнем складності. Наприклад:

– Для курсу з теорії ймовірності: «Розрахуйте ймовірність того, що при киданні трьох гральних кубиків сума буде кратною трьом».

– Для модуля з нейромереж: «Поясніть, чому функція активації ReLU не використовується в останньому шарі мережі класифікації».

– Для розділу з вищої математики, у якому студенти вивчають аналітичну геометрію: «Побудуйте множину точок на площині, що задовольняють задану систему нерівностей з двома змінними».

У модулі генерації реалізовано механізм обробки інструкцій, що поєднує шаблон, ключові параметри, предметну ділянку й рівень Блума. На основі цього формується текст запитання, додаються математичні вирази, здійснюється перевірка синтаксичної коректності. Результати валідації проходять через фільтр, і лише узгоджені з критеріями запитання потрапляють у фінальний набір.

Генерація тестових запитань у системі відбувається у два основні етапи: семантичне конструювання запиту й синтаксична обробка результату. На першому етапі користувач (викладач) обирає предмет (наприклад, теорія ймовірностей, аналітична геометрія, основи нейромереж), указує бажаний рівень Блума й задає кількість завдань. Система за допомогою мовної моделі формує контекстно релевантні запити до шаблонного генератора. Далі згенеровані запитання проходять етап математичної обробки. Наприклад, числа у виразі

«Обчисліть значення виразу $a^2 + b^2$
при $a = 3$, $b = 4$»

автоматично заповнюються з використанням генератора цілих значень у допустимому діапазоні. Така модель є гнучкою для всіх рівнів складності: від простих обчислень до складніших задач на аналіз або моделювання.

Система підтримує різні типи запитань:

– з однією або декількома правильними відповідями,
– з відкритою відповіддю,
– на пошук відповідності,
– заповненням пропусків,
– комбіновані.

У завданнях з відкритими відповідями здійснюється перевірка на математичну еквівалентність: два вирази $x^2 + 2x + 1$ і $(x+1)^2$ розпізнаються як тотожні. Це особливо актуально під час створення тестів з вищої математики або теорії ймовірності, де студент може ввести відповідь у різному вигляді. Якщо студент уводить аналітичний вираз, система спрощує обидва вирази (правильний і наданий студентом) за допомогою бібліотеки SymPy й порівнює їх у канонічній формі.

Для генерації запитань зі сфери інтелектуального аналізу даних або нейромереж передбачено вектор термінів, пов'язаних із відповідною дисципліною, що підвантажується з попередньо складених списків понять і тем. Наприклад, завдання з теми «Перцептрон» може автоматично генеруватися з варіаціями запитання про функцію активації або геометричну інтерпретацію межі класифікації.

Формування тестових завдань з математичним змістом супроводжується низкою технічних і методичних труднощів, пов'язаних із точністю запису, візуалізацією формул, підтримкою синтаксису й обробкою введених студентом відповідей. У системі реалізовано комплексний підхід до формулювання, збереження та перевірки таких завдань.

Однією з ключових проблем є представлення формул у форматі, придатному для онлайн-перегляду й редагування. Для цього використовується формат LaTeX, що є де-факто стандартом у системах Moodle, Open edX та інших LMS. Для їх візуалізації застосовуються бібліотеки MathJax або KaTeX, які сумісні з більшістю браузерів і мобільних пристроїв.

Наприклад, завдання типу:

«Обчисліть похідну функції $f(x) = \sqrt{x^2 + 1}$»

записується як рядок LaTeX і передається у візуальний редактор. У системі забезпечено автоматичну генерацію таких формул із параметризованими змінними й додатковими стилістичними обмеженнями (наприклад, уникати дробів або логарифмів для початкового рівня складності).

Бібліотека Streamlit є сучасним інструментом для створення інтерактивних вебінтерфейсів до Python-додатків. Вона розроблена спеціально для задач аналітики, візуалізації та машинного навчання й дає змогу будувати повноцінні фронтенд-рішення з мінімальними зусиллями з боку розробника.

У контексті реалізації інтелектуальної системи формування тестів для математичних дисциплін Streamlit забезпечує низку ключових переваг:

1. Миттєвий запуск без вебпрограмування. Усі елементи інтерфейсу (кнопки, списки, слайдери, текстові поля) створюються безпосередньо у Python-коді.

2. Підтримка математичного формату. Streamlit нативно підтримує Markdown та LaTeX, що дає змогу легко рендерити формули, наприклад:

st.latex(r"f(x) = \int_0^1 x^2 dx")

3. Інтерактивність. Можна легко реалізувати логіку залежності між елементами форми: наприклад, при виборі рівня Блума змінюється перелік доступних шаблонів.

4. Збереження стану**.** Через st.session_state підтримується передача змін між запусками функцій, що дає змогу керувати кількома кроками генерації (вибір параметрів → генерація → редагування → експорт).

5. Візуалізація. Інтерфейс може містити графіки, таблиці з даними, індикатори складності завдань тощо – усе це використовується для зворотного зв'язку з викладачем.

6. Підключення до зовнішніх джерел. Streamlit дає змогу підключатися до баз даних, API сервісів Moodle, Google Sheets або LMS через бібліотеки requests, sqlalchemy тощо.

7. Інтеграція з машинним навчанням**.** Якщо генерація тестів включає LLM або моделі класифікації, результат їхньої роботи може миттєво відображатися у вебінтерфейсі.

Streamlit забезпечує достатній рівень безпеки й може розгортатися як на локальному сервері, так і в хмарі (наприклад, Streamlit Cloud, Heroku, Google Cloud). Серед недоліків варто відзначити обмежену кастомізацію інтерфейсу порівняно з React/Angular, проте для завдань академічного тестування це не є критичним.

Вибір Streamlit як платформи для побудови інтерфейсу зумовлений його швидкістю розробки, простотою інтеграції з Python-бібліотеками та фокусом на потреби користувачів без досвіду програмування.

Висновок

У статті розглянуто концепцію та реалізацію інтелектуальної системи підтримки викладача під час формування тестових завдань з математичним змістом. Показано, що поєднання сучасних мовних моделей, бібліотек генерації математичних виразів і систем керування навчанням (LMS) відкриває нові можливості для автоматизації освітнього процесу. Такий підхід не лише зменшує навантаження на викладача, а й підвищує якість оцінювання завдяки адаптації до рівня знань студента, підтримці різних типів завдань і формулювань у таких галузях, як вища математика, теорія ймовірності, інтелектуальний аналіз даних і нейромережі.

Запропонована система має модульну архітектуру, що дає змогу гнучко налаштовувати процес генерації завдань для різних дисциплін, автоматизовано перевіряти математичну еквівалентність відповідей, інтегрувати систему з платформами Moodle, Google Classroom і зберігати індивідуальні траєкторії навчання.

Подальші дослідження доцільно спрямувати на реалізацію зворотного зв'язку з LMS, інтеграцію із системами оцінювання на основі IRT, а також на забезпечення підтримки математичних задач у середовищах доповненої або віртуальної реальності.

Таким чином, запропоноване рішення є перспективним кроком у напрямі створення адаптивних, інклюзивних і масштабованих освітніх технологій для математичних дисциплін.

Література

1. Використання відкритого освітнього середовища з елементами штучного інтелекту для професійного розвитку вчителів: метод. рекомендації / О. О. Гриб'юк та ін. ; за ред. М. П. Шишкіної. Київ : ІЦО НАПН України, 2024. 118 с.

2. Гриценчук О. Використання штучного інтелекту в освіті: тенденції та перспективи в Україні та за кордоном. *Вісник кафедри ЮНЕСКО «Неперервна професійна освіта XXI століття»*. 2024. Вип. 10. С. 152–161.

3. Створення інформаційно-освітнього середовища сучасного закладу освіти України : матеріали Всеукраїнської науково-практичної конференції (м. Київ, 15 березня 2019 р.) / за заг. ред. Г. А. Коломоєць, О. М. Мельник, С. М. Грицай, А. В. Вознюк. Суми : НВВ КЗ СОІППО, 2019. 124 с.

4. Lord F. M. Applications of Item Response Theory to Practical Testing Problems. Hillsdale (NJ): Lawrence Erlbaum Associate s, 1980. 288 с.

5. Мар'єнко М. В. Моделювання хмаро орієнтованої методичної системи підготовки вчителів природничо-математичних предметів до роботи в науковому ліцеї. *Фізико-математична освіта.* 2020. № 2(24). С. 87–93.

6. Хмаро орієнтовані системи відкритої науки у навчанні і професійному розвитку вчителів : монографія / М. П. Шишкіна, С. Г. Литвинова, М. В. Мар'єнко та ін. ; за наук. ред. М. П. Шишкіної. Київ : ІЦО НАПН України, 2023. 176 с.

7. MathJax documentation. URL: https://docs.mathjax.org.

8. System architecture: MoodleDocs. URL: https://docs.moodle.org/dev.

References

1. Hrybiuk, O. O. (2024). Vykorystannia vidkrytoho osvitnoho seredovyshcha z elementamy shtuchnoho intelektu dlia profesiinoho rozvytku vchyteliv: Metodychni rekomendatsii [Using an open educational environment with elements of artificial intelligence for teachers' professional development: Methodical recommendations] (M. P. Shyshkina, Ed.). Kyiv: ITsO NAPN Ukrainy [in Ukrainian].

2. Hrytsenchuk, O. (2024). Vykorystannia shtuchnoho intelektu v osviti: Tendentsii ta perspektyvy v Ukraini ta za kordonom [The use of artificial intelligence in education: Trends and prospects in Ukraine and abroad]. Visnyk kafedry UNESCO «Neperervna profesiina osvita XXI stolittya» [Bulletin of the UNESCO Department "Lifelong Professional Education of the 21st Century"], 10, 152–161 [in Ukrainian].

3. Kolomoiets, H. A., Melnyk, O. M., Hrytsai, S. M., & Vozniuk, A. V. (Eds.). (2019). Stvorennia informatsiino osvitnoho seredovyshcha suchasnoho zakladu osvity Ukrainy: Materialy Vseukrainskoi naukovo praktychnoi konferentsii (m. Kyiv, 15 bereznia 2019 r.) [Creating an information educational environment of a modern educational institution of Ukraine: Proceedings of the All Ukrainian scientific and practical conference (Kyiv, 15 March 2019)]. Sumy: NVV KZ SOIPPO [in Ukrainian].

4. Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates [in English].

5. Marienko, M. V. (2020). Modeliuvannia khmaro oriientovanoi metodychnoi systemy pidhotovky vchyteliv pryrodnycho matematych nykh predmetiv do roboty v naukovomu litsei [Modeling a cloud oriented methodological system for training teachers of natural and mathematical subjects to work in a scientific lyceum]. Fiziko matematychna osvita [Physics & Mathematics Education], 2(24), 87–93 [in Ukrainian].

6. Shyshkina, M. P., Lytvynova, S. H., Marienko, M. V., Nosenko, Yu. H., Kovalenko, V. V., Luparenko, L. A., & Sukhikh, A. S. (2023). Khmaro oriientovani systemy vidkrytoi nauky u navchanni i profesiinomu rozvytku vchyteliv: Monohrafiia [Cloud oriented open science systems in teacher education and professional development: Monograph]. Kyiv: ITsO NAPN Ukrainy [in Ukrainian].

7. The MathJax Consortium. (n.d.). MathJax documentation. From: https://docs.mathjax.org [in English].

8. MoodleDocs. (n.d.). System architecture. From : https://docs.moodle.org/dev [in English].

Дата першого надходження рукопису до видання: 16.05.2025
Дата прийнятого до друку рукопису після рецензування: 12.06.2025
Дата публікації: 25.07.2025

УДК 004.42:004.8
DOI https://doi.org/10.36994/2788-5518-2025-01-09-18

МУЛЬТИАГЕНТНА СИСТЕМА ДЛЯ ВИЯВЛЕННЯ ІНФОРМАЦІЙНИХ МАНІПУЛЯЦІЙ В ОНЛАЙН-СЕРЕДОВИЩІ

Постовий О. В., *Національний університет «Львівська політехніка», Львів, Україна. https://orcid.org/0009-0002-3935-6912, o.postovyi@gmail.com*

Шкраб Р. Р., *ст. викл., Національний університет «Львівська політехніка», Львів, Україна. https://orcid.org/0000-0001-9813-2877, roman.r.shkrab@lpnu.ua*

Анотація. *Поширення інформаційних маніпуляцій становить значну загрозу для сучасного суспільства, особливо в контексті війни та політичної нестабільності, з якою стикається Україна. У статті досліджено сучасні методи автоматизованого виявлення, аналізу й оцінювання достовірності новинного контенту. Запропоновано новий, мультиагентний підхід, де спеціалізовані програмні агенти, чий потік виконання керується великими мовними моделями, беруть на себе чітко визначені різні ролі для всебічного дослідження актуального новинного контенту. Використання декількох агентів, що конкурують між собою та намагаються проаналізувати різні погляди на новинний контент, продемонструвало ефективність мультиагентного підходу в галузі дослідження інформаційного простору й виявлення інформаційних маніпуляцій. Система інтегрує великі мовні моделі для ретельного виконання широкого спектру завдань, пов'язаних із глибоким аналізом текстового контенту, розумінням контексту, факт-чекінгом, формуванням обґрунтованих висновків і генерацією звітів, а також використовує передові бібліотеки для оркестрації та відстеження дій агентів, що керуються мовними моделями. Для забезпечення всебічного аналізу новинного контенту система звертається до різноманітних зовнішніх ресурсів: пошукових систем для знаходження релевантної інформації в інтернеті, новинних агрегаторів для доступу до статей передових ЗМІ, платформ для збору текстових даних із соціальних мереж. Отримані результати підтверджують, що впровадження великих мовних моделей та агентно-орієнтованого підходу в рамках N-рівневої архітектури дає змогу підвищити рівень аналізу інформаційного середовища й удосконалити процес виявлення інформаційних маніпуляцій.*

Метою статті є визначення необхідності використання мультиагентних систем для масштабування й поглиблення методів виявлення інформаційних маніпуляцій.

Ключові слова: *дезінформація, інформаційні маніпуляції, штучний інтелект, великі мовні моделі, мультиагентні системи, аналіз новин, Langchain, Langgraph, виявлення фейків.*

MULTI-AGENT SYSTEM FOR DETECTING INFORMATION MANIPULATION IN THE ONLINE ENVIRONMENT

Oleksii Postovyi, *Lviv Polytechnic National University, Lviv, Ukraine. https://orcid.org/0009-0002-3935-6912, o.postovyi@gmail.com*

Shkrab Roman, *Sr. Lect., Lviv Polytechnic National University, Lviv, Ukraine. https://orcid.org/0000-0001-9813-2877, roman.r.shkrab@lpnu.ua*

Abstract. *The spread of information manipulation poses a significant threat to modern society, especially in the context of war and political instability faced by Ukraine. This paper investigates modern methods of automated detection, analysis, and evaluation of news content reliability. The paper proposes a new, multi-agent approach, where specialized software agents, whose execution flow is guided by large language models, take on well-defined different roles to comprehensively investigate relevant news content. The use of several agents competing with each other and trying to analyze different views on news content has demonstrated the effectiveness of the multi-agent approach in the field of information space research and detection of information manipulation. The system integrates large language models to thoroughly perform a wide range of tasks related to in-depth analysis of textual content, contextual understanding, fact-checking, drawing reasonable conclusions and generating reports, and uses advanced libraries to orchestrate and track the actions of agents guided by language models. To provide a comprehensive analysis of news content, the system uses a variety of external resources: search engines to find relevant information on the Internet, news aggregators to access articles from leading media outlets, and platforms to collect text data from social networks. The results obtained confirm that the introduction of large language models and an agent-based approach within the framework of the N-level architecture allows to increase the level of analysis of the information environment and improve the process of detecting information manipulation.*

The purpose of the article is to determine the need to use multi-agent systems to scale and deepen the methods of detecting information manipulation.

Key words: *disinformation, information manipulation, artificial intelligence, large language models, multi-agent systems, news analysis, Langchain, Langgraph, fake detection.*

Вступ

У сучасному світі Україна стикається з проблемою інформаційної війни, яка є частиною російської агресії. Інформаційні маніпуляції використовуються для впливу на громадську думку, дестабілізації суспільства та підриву демократичних процесів. Маніпулятивний контент активно поширюються через соціальні мережі, месенджери, новинні сайти й навіть традиційні медіа, створюючи загрозу інформаційній безпеці українського суспільства.

Одним із основних джерел маніпуляцій є пропагандистські ресурси, які поширюють вигідні наративи, спрямовані на формування викривленого уявлення про події в Україні та світі. Такі наративи часто містять емоційно забарвлені меседжі, спекуляції на соціально чутливих темах, викривлення історичних фактів і нав'язування недостовірних тверджень. Дослідження показують, що джерела маніпуляцій мають інший стиль написання, спрямований на легкість прочитання й поширення [1]. Дезінформація легша для обробки з погляду когнітивного зусилля (на 3% легша для читання, на 15% менш лексично різноманітна) і більш емоційна (у 10 разів більше покладається на негативний сентимент, на 37% більше апелює до моралі) (рис. 1), ніж надійні джерела [1].

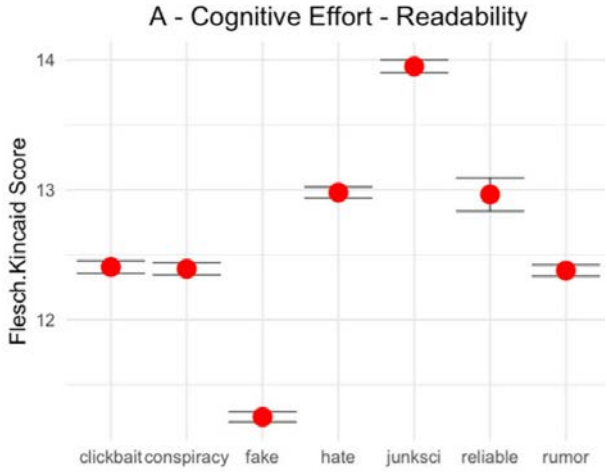


Рис. 1. Показники читабельності для різних типів інформаційних маніпуляцій

Чим легше читати новинний контент, тим імовірніше він належить до категорій клікбейту, теорій змови, фейкових новин або чуток [1]. Наприклад, фейкові новини в середньому на 8% простіші для читання й на 18% простіші в плані лексики, а також на 18% більше покладаються на негативні емоції та на 45% більше апелюють до моралі, ніж фактологічні новини [1]. Використання моральної та емоційної мови є одним із основних факторів комунікаційних стратегій маніпуляції [1]. Фактологічні новини, на відміну від них, є, по суті, нейтральними за емоційним забарвленням (-0.19). Для порівняння, розпалювання ненависті має найвище негативне значення емоційного забарвлення (-6.01), за ним ідуть фейкові новини (-3.51) і теорії змови (-3.41) (рис. 2).

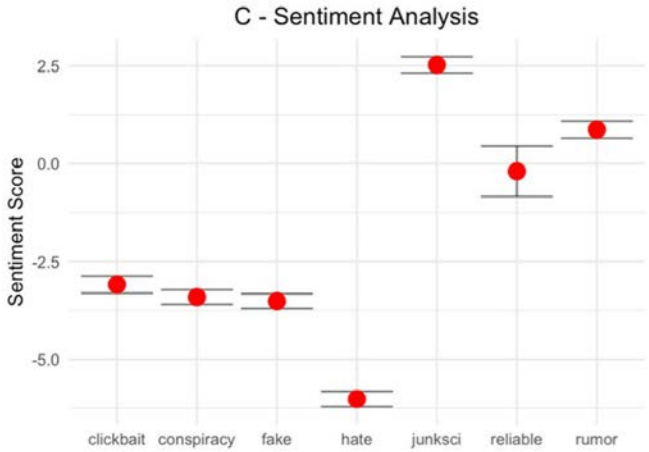


Рис. 2. Показники емоційного забарвлення для різних типів інформаційних маніпуляцій

Інформаційні маніпуляції в українському інформаційному просторі можуть стосуватися різних тем – від війни та міжнародної політики до економічних криз і внутрішньополітичних процесів. Окремо варто виділити маніпуляції, спрямовані на підрив довіри до державних інституцій, зокрема до армії, правоохоронних органів та уряду. Часто такі матеріали подаються під виглядом «аналітики» або «експертних думок» у соціальних мережах, що ускладнює їх викриття без спеціальних методів аналізу й додаткового пошуку інформації про джерело новини. Платформи на кшталт Facebook, Instagram, X, Telegram, TikTok і YouTube містять значний обсяг контенту, який важко модерується та використовується для інформаційних атак.

Виявлення інформаційних маніпуляцій в онлайн-середовищі є критично важливим завданням, що вимагає автоматизованих підходів через безпрецедентний обсяг контенту, який щодня публікується у вищезгаданих соціальних мережах, новинних ресурсах і блогах [1; 2].

Автоматизовані методи аналізу контенту дають змогу ефективно знаходити потенційно маніпулятивний контент шляхом аналізу стилістичних, лексичних і семантичних особливостей [1].

Одним із ключових етапів є аналіз тексту й класифікація інформаційних маніпуляцій. Дослідження вказують на можливість використання моделей машинного навчання для виявлення фейкових новин [12]. У публікації автори порівнюють Naïve Bayes, SVM і KNN для класифікації новин українською мовою [12]. Цей підхід передбачає наявність датасету з текстовими даними, препроцесинг тексту у вигляді токенізації та векторизації за допомогою TF-IDF, що дає змогу натренувати вищезгадані моделі машинного навчання та використати їх для визначення фейкової новини на раніше не баченому моделлю контенті [12].

Інші ж підходи передбачають використання агента з певною послідовністю етапів аналізу [13]. В основі такого агента використовується велика мовна модель, яка керує потоком управління програми. Агент обладнаний інструментами пошуку та вебскрапінгу для пошуку інформації про новину в зовнішніх джерелах. LLM аналізує зібрану інформацію, оцінює її достовірність, синтезує докази й, зрештою, виносить вердикт щодо правдивості новини, надаючи пояснення. Цей процес може бути ітеративним, де LLM уточнює запити та збирає додаткові дані за потреби. Стаття демонструє ефективність FactAgent на трьох реальних датасетах (Snopes, PolitiFact та GossipCop) і порівнює її з іншими методами виявлення фейкових новин. Результати показують, що цей агент перевершує інші підходи, зокрема ті, що ґрунтуються на навчених моделях, стандартних запитах і ланцюгах думок [13].

Архітектура проєктного рішення

У розробленій програмній системі вирішено використовувати агентно-орієнтований підхід з декількома кроками для формування фінального висновку.

Головною особливістю розробленої системи є використання мультиагентної взаємодії та створення конкуренції агентів для всебічного аналізу поданого контенту, на відміну від запропонованого підходу, що передбачає використання лише одного агента [13]. Також у цій системі є інструменти для аналізу соціальних мереж, таких як X (Twitter) і Telegram. Крім того, система може виявляти на помічати інформаційні маніпуляції наступними позначками: «Fake News», «Hate Speech», «Clickbait», «Conspiracy», «Rumours». За основу для класифікації взято наведені в дослідженні The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions типи маніпуляцій [1].

Незважаючи на використаний агентно-орієнтований підхід, за архітектуру взято N-рівневу систему. Сама ж взаємодія агентів відбувається на рівні бізнес-логіки.

Рівень бізнес-логіки включає такі сервіси:

• Сервіс аналізу новин. Центральний сервіс, що покроково реалізує процес аналізу новини. Він ініціює та координує роботу ШІ-агентів, а також відповідає за збереження результатів дослідження в БД.

• Агенти. За допомогою мовних моделей контролюють потік виконання програми й вирішують, які інструменти та сервіси використовувати. Типи створених агентів:

– Агент-«суддя» – головний агент, який керує процесом аналізу. Він отримує новину, ініціює роботу інших агентів і на основі їхніх звітів формує фінальний висновок, базуючись на звітах інших агентів. Агент використовує LLM o4-mini від OpenAI. Для реалізації цього типу агента, що виносить фінальний вердикт щодо інформаційних маніпуляцій у новині, у системі використано патерн створення агентів Orchestrator [9].

– Агент-аналітик. Є спільним класом для «прокурора» (шукає ознаки маніпуляції) й «адвоката» (шукає підтвердження правдивості новини). Кожен із цих агентів використовує LLM від Google (Gemini Flash 2.5) і набір інструментів для збору інформації. Принциповою різницею між «адвокатом» і «прокурором» є системний промпт, який задає тон для взаємодії аналітика із середовищем. Для реалізації агентів-аналітиків використано ReAct патерн [11]. У процесі розробки граф цього типу агента модифіковано для гарантованого написання фінального звіту й витягування використаних джерел для подальшого зберігання в базі даних.

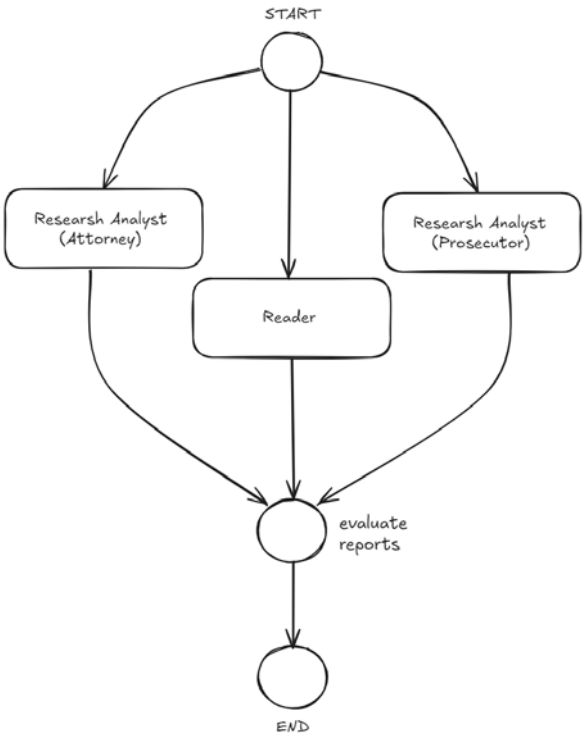


Рис. 3. Архітектура агента-«судді»

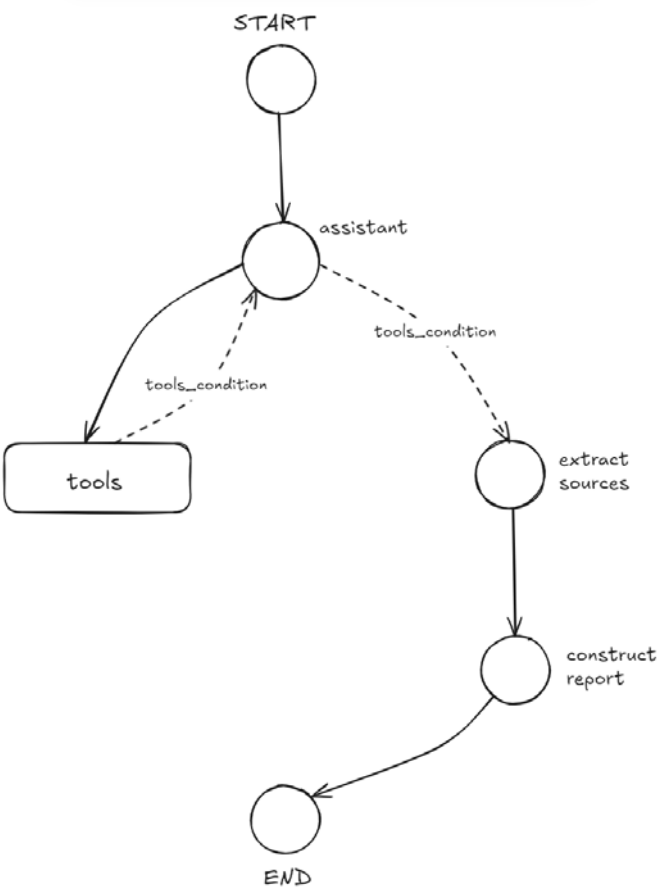


Рис. 4. Архітектура агента-«аналітика»

– Агент-«читач». Аналізує наданий текст новини безпосередньо для виявлення можливих типів маніпуляцій на основі визначених промптів для LLM. Нині реалізує лінійний потік виконання, проте в майбутньому може бути наділений спеціалізованими інструментами й керувати потоком виконання самостійно.

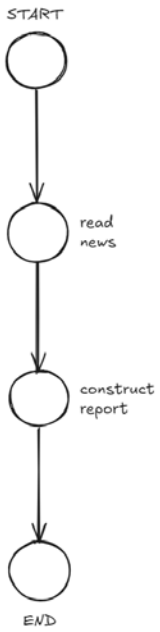


Рис. 5. Діаграма архітектури агента-«читача»

• Інструментарії для агентів. Сервіси, що містять різні типи інструментів для агентів. Для реалізації за основу взято концепцію Toolkit (інструментарія) із фреймворку Langchain [10]:

– Інструментарій для веб-пошуку використовує Tavily Search API й OpenAI Search API для пошуку релевантної інформації в Інтернеті.

– Інструментарій для скрапінгу новин взаємодіє з NewsAPI для пошуку новинних статей, пов’язаних з аналізованим контентом.

– Інструментарій для соціальних мереж призначений для збору даних із соціальних мереж, таких як Telegram та X, за допомогою платформи Apify.

Виконання досліджень

Для демонстрації роботи системи й оцінювання її ефективності у виявленні інформаційних маніпуляцій розроблено декілька сценаріїв дослідження. Кожен сценарій моделює типову ситуацію, з якою може зіткнутися користувач при аналізі новинного контенту.

Сценарій 1. Аналіз новини з ознаками «мови ворожнечі» (hate speech). Користувач надає системі текстовий фрагмент новини, яка містить ознаки мови ворожнечі й розпалювання ненависті (рис. 6).

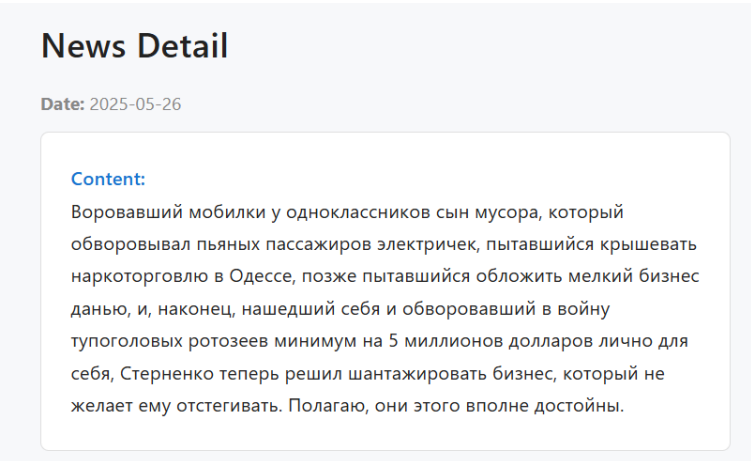


Рис. 6. Новина з мовою ворожнечі

Очікуваний результат: система класифікує новину як мову ворожнечі, надає докази її неправдивості й посилання на авторитетні джерела, що її спростовують.

Сценарій 2. Аналіз новини, що ймовірно поширює теорію змови (Conspiracy) або чутки (Rumours). Користувач надає текст новини, яка просуває необґрунтовану теорію змови або поширює неперевірені чутки. Така новина часто посилається на анонімні джерела, «експертів» без чіткої ідентифікації або будує складні, але логічно не підкріплені зв'язки між подіями (рис. 7).

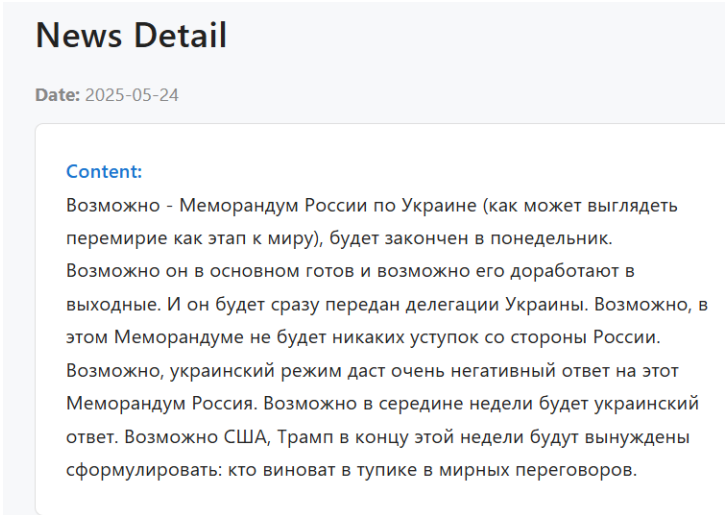


Рис. 7. Новина, що поширює слухи та/або теорії змови

Очікуваний результат: система описує новину, яка поширює необґрунтовану теорію змови або чутки, пояснюючи відсутність доказів і можливі мотиви поширення такої інформації, або в разі недостатніх доказів на користь присутності теорії змови пояснює, чому новина є правдивою.

Сценарій 3. Аналіз фактологічної новини з надійного джерела. Користувач надає системі текст новини, опублікованої відомим та авторитетним ЗМІ. Новина висвітлює реальну подію, подає факти збалансовано, посилається на конкретні джерела або офіційних осіб, уникає надмірної емоційності й сенсаційності (рис. 8).

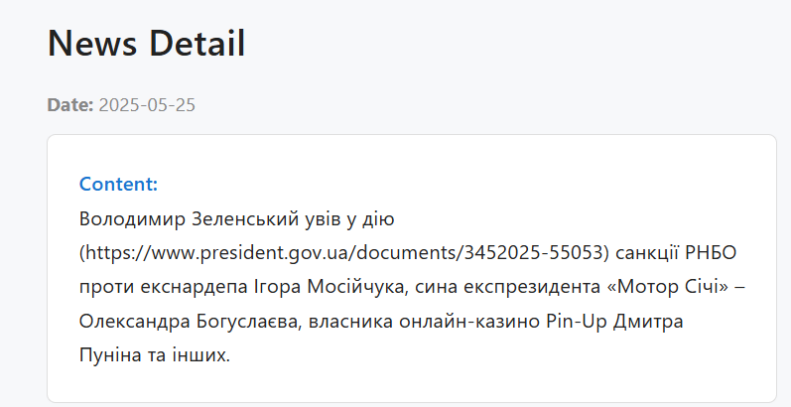


Рис. 8. Приклад новини без інформаційних маніпуляцій

Очікуваний результат: система описує новину як достовірну й таку, що не містить інформаційних маніпуляцій. У звіті наводяться посилання на інші авторитетні джерела, що підтверджують викладені факти, і зазначається відсутність виявлених маніпулятивних технік.

Метою дослідження було оцінити потенційну ефективність системи у виявленні різних типів інформаційних маніпуляцій, а також її здатність коректно ідентифікувати фактологічно достовірні новини. У ході дослідження також перевірено роботу системи у різних сценаріях.

На основі детального розбору кожного сценарію (фейкові новини, клікбейт, теорії змови/чутки, фактологічна новина) система продемонструвала високу ефективність у досягненні поставлених цілей.

У випадку сценарію 1 (розпалювання ненависті) система успішно ідентифікувала новину як таку, що містить мову ворожнечі. Ключову роль відіграв агент «Читач», який виявив ознаки Hate Speech

у самому тексті. Фінальний висновок агента «Судді» чітко аргументував розпалювання ненависті в тексті новини (рис. 9).

Рис. 9. Висновок про новину, що поширює мову ворожнечі

У процесу аналізу також успішно ідентифіковано маніпуляції Hate Speech і Fake News (рис. 10).

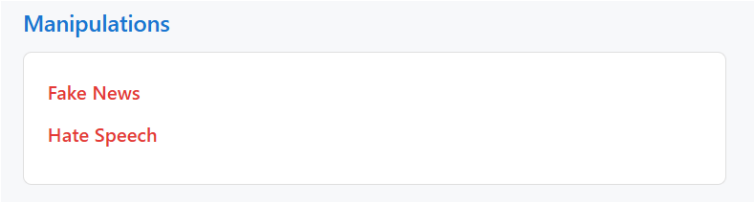


Рис. 10. Виявлені маніпуляції

Також помітно, що агент-«адвокат» для захисту тверджень у новині більше схилявся до проросійських джерел (рис. 11).

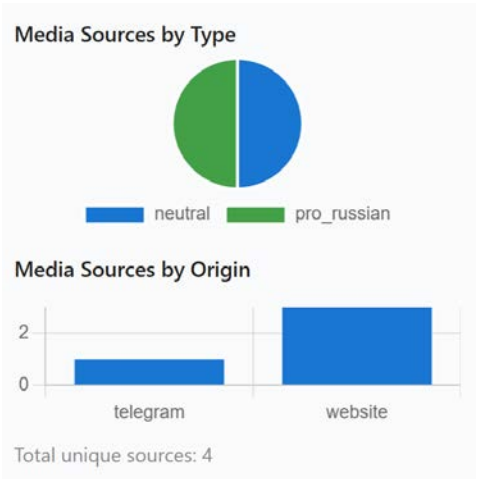


Рис. 11. Джерела, до яких звертався агент-«адвокат» для аналізу новини зі сценарію 1

Агент-«прокурор», у свою чергу, проаналізував більше джерел, серед яких були проукраїнські.

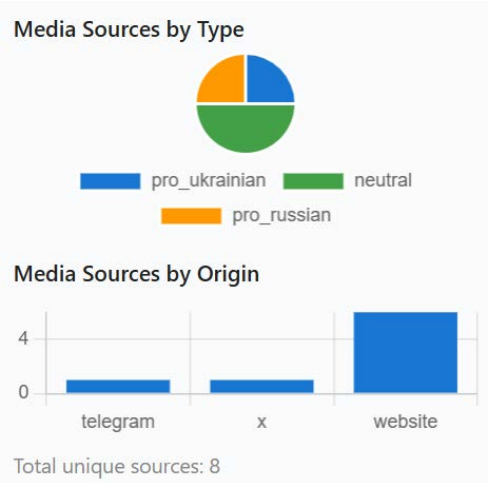


Рис. 12. Джерела, до яких звертався агент-«прокурор» для аналізу новини зі сценарію 1

У випадку сценарію 2 (теорія змови/чутки) система ефективно виявила ознаки чуток. Агент-«читач» указав на характерні риси Rumours. Його висновок став основою для висновку агента-«судді», оскільки агент-«читач» зміг виявити доволі необґрунтовані твердження, які не підтверджені ні агентом-«прокурором», ні агентом-«адвокатом» (рис. 13).

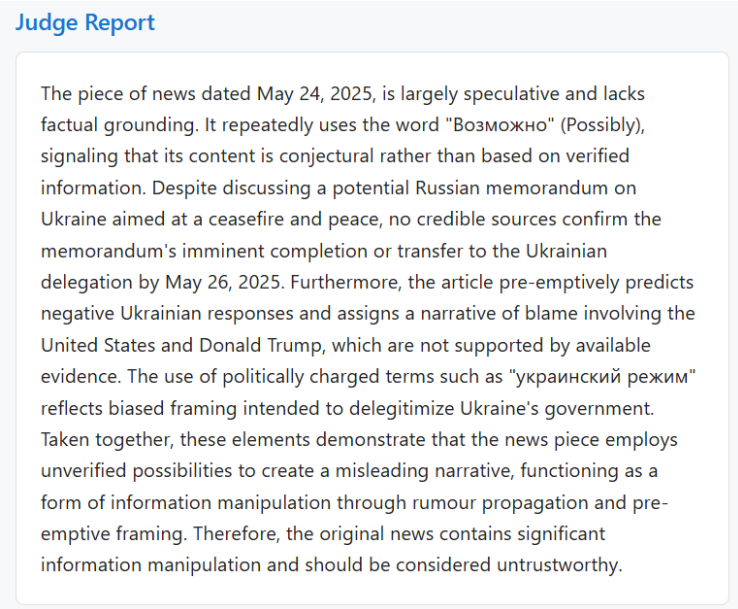


Рис. 13. Висновок системи про новину, що поширює теорії змови й/або чутки

У процесу аналізу також успішно ідентифіковано маніпуляцію Rumours (рис. 14).

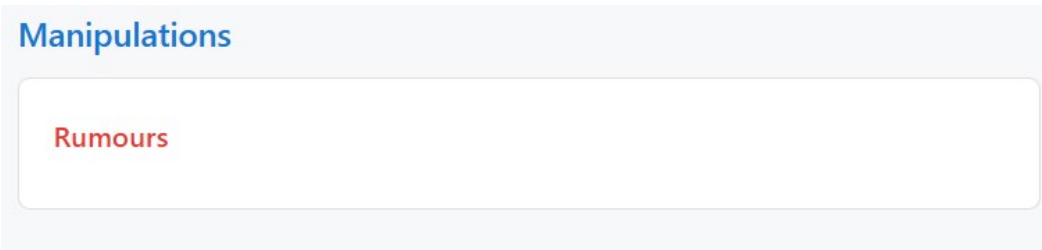


Рис. 14. Виявлені маніпуляції

Агент-«адвокат» намагався знайти підтвердження новини в нейтральних джерелах (рис. 15).

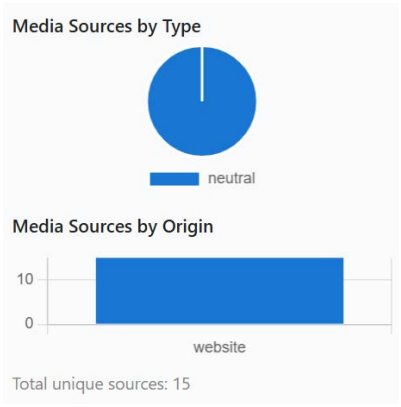


Рис. 15. Джерела, до яких звертався агент-«адвокат» для аналізу новини зі сценарію 2

Агент-«прокурор» теж використав лише нейтральні джерела (рис. 16).

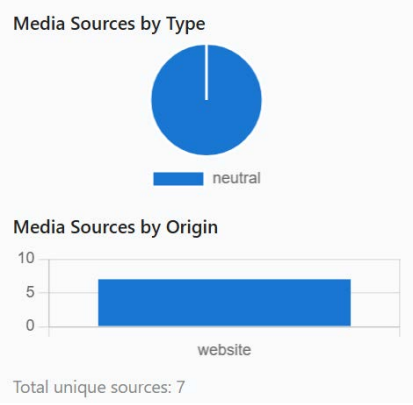


Рис. 16. Джерела, до яких звертався агент-«прокурор» для аналізу новини зі сценарію 2

У випадку сценарію 3 (фактологічна новина) система продемонструвала здатність правильно ідентифікувати достовірну інформацію. «Читач» не виявив маніпуляцій, «прокурор» зміг довести, що новина правдива, надавши численні підтвердження з авторитетних джерел, а «адвокат» не зміг знайти переконливих доказів неправдивості новини. Висновок «судді» підтвердив фактологічність і відсутність маніпуляцій (рис. 17).

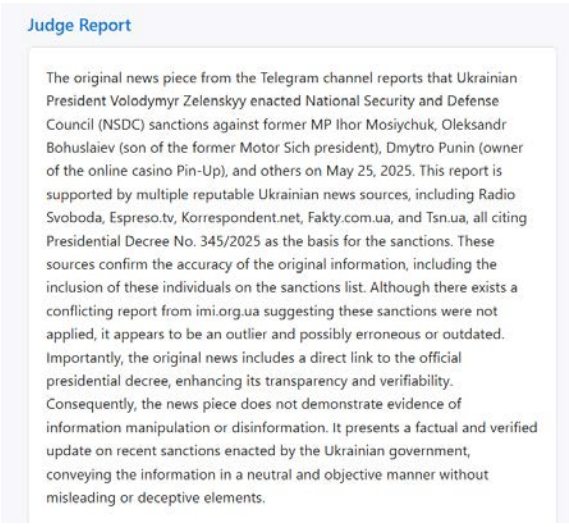


Рис. 17. Фінальний висновок системи для новини без інформаційних маніпуляцій

Агент-«адвокат» використав більше нейтральних джерел, хоча шукав інформацію також і в проукраїнських джерелах (рис. 18).

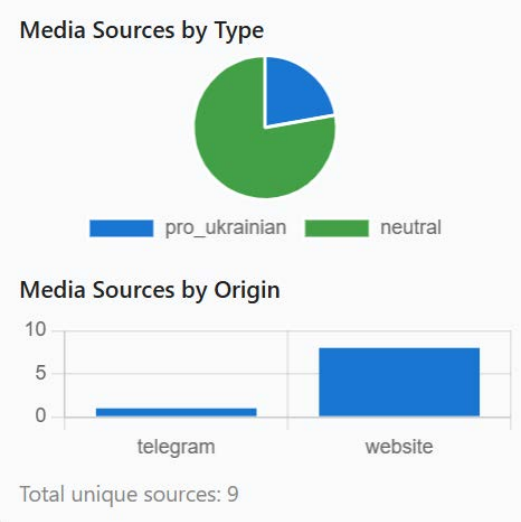


Рис. 18. Джерела, до яких звертався агент-«адвокат» для аналізу новини зі сценарію 3

Агент-«прокурор» використав більше проукраїнських джерел, ніж агент-«адвокат», проте теж переважно переглядав нейтральні джерела (рис. 19).

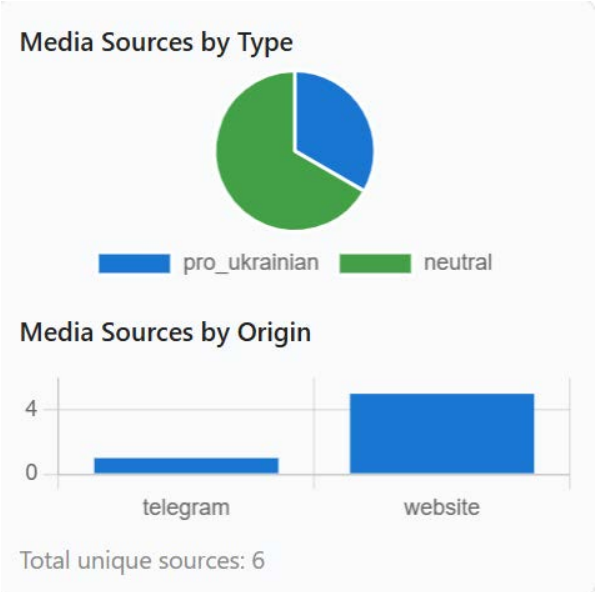


Рис. 19. Джерела, до яких звертався агент-«прокурор» для аналізу новини зі сценарію 3

Висновок

Дослідження дало змогу встановити, що мультиагентний підхід для виявлення інформаційних маніпуляцій є доволі ефективним. У всіх трьох сценаріях, які змодельовано в ході дослідження, система видала правильний результат, надавши детальний розбір новини з підкріпленням із різноманітних джерел.

Розроблена мультиагентна архітектура також дає змогу гнучко інтегрувати різноманітні нові джерела даних та інструменти, даючи змогу забезпечити ще глибше дослідження кожної новини.

Література

1. Carrasco-Farré C. The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities & Social Sciences Communications*. 2022. Vol. 9. Art. 162. DOI: 10.1057/s41599-022-01174-9.

2. Golovchenko Y. Measuring the scope of pro-Kremlin disinformation on Twitter. *Humanities & Social Sciences Communications*. 2020. Vol. 7. Art. 176. DOI: 10.1057/s41599-020-00659-9.

3. Vicari R., Komendatova N. Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters. *Humanities & Social Sciences Communications*. 2023. Vol. 10. Art. 332. DOI: 10.1057/s41599-023-01838-0.

4. Mantis Analytics. AI-Powered Early Warning System for Information Environment Threats : веб-сайт. URL: https://mantisanalytics.com/ (дата звернення: 07.05.2025).

5. Reality Defender. The Only Comprehensive Deepfake & AI-Generated Content Detection Platform : веб-сайт. URL: https://www.realitydefender.com/ (дата звернення: 07.05.2025).

6. LangGraph. Introduction. *LangGraph Documentation* : веб-сайт. URL: https://langchain-ai.github.io/langgraph/ (дата звернення: 08.05.2025).

7. Apify. Welcome to Apify Docs. *Apify Documentation* : веб-сайт. URL: https://docs.apify.com/ (дата звернення: 08.05.2025).

8. SQLAlchemy. SQLAlchemy 2.0 Documentation : веб-сайт. URL: https://docs.sqlalchemy.org/en/20/ (дата звернення: 08.05.2025).

9. LangGraph. Workflows and Agents. *LangGraph.js Documentation* : веб-сайт. URL: https://langchain-ai.github.io/langgraphjs/tutorials/workflows/#orchestrator-worker (дата звернення: 08.05.2025).

10. Langchain. Toolkits. *Langchain Python Documentation* : веб-сайт. URL: https://python.langchain.com/docs/concepts/tools/#toolkits (дата звернення: 12.05.2025).

11. Langchain. create_react_agent. *Langchain Python API Reference* : веб-сайт. URL: https://python.langchain.com/api_reference/langchain/agents/langchain.agents.react.agent.create_react_agent.html (дата звернення: 08.05.2025).

12. Джоші А., Павленко В. І. Порівняння методів машинного навчання для визначення фейкових новин. *Інфокомунікаційні та комп'ютерні технології*. 2024. Vol. 1, № 07. С. 46–55. DOI: 10.36994/2788-5518-2024-01-07-06.

13. Li X., Zhang Y., Malthouse E.C. Large Language Model Agent for Fake News Detection. arXiv. 2024. URL: https://arxiv.org/html/2405.01593v1 (дата звернення: 01.05.2025).

References

1. Carrasco-Farré C. The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities & Social Sciences Communications*. 2022. Vol. 9. Art. 162. DOI: 10.1057/s41599-022-01174-9 [in English].

2. Golovchenko Y. Measuring the scope of pro-Kremlin disinformation on Twitter. *Humanities & Social Sciences Communications*. 2020. Vol. 7. Art. 176. DOI: 10.1057/s41599-020-00659-9 [in English].

3. Vicari R., Komendatova N. Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters. *Humanities & Social Sciences Communications*. 2023. Vol. 10. Art. 332. DOI: 10.1057/s41599-023-01838-0 [in English].

4. Mantis Analytics. AI-Powered Early Warning System for Information Environment Threats : website. URL: https://mantisanalytics.com/ (accessed: 07.05.2025) [in English].

5. Reality Defender. The Only Comprehensive Deepfake & AI-Generated Content Detection Platform : website. URL: https://www.realitydefender.com/ (accessed: 07.05.2025) [in English].

6. LangGraph. Introduction. *LangGraph Documentation* : website. URL: https://langchain-ai.github.io/langgraph/ (accessed: 08.05.2025) [in English].

7. Apify. Welcome to Apify Docs. *Apify Documentation* : website. URL: https://docs.apify.com/ (accessed: 08.05.2025) [in English].

8. SQLAlchemy. SQLAlchemy 2.0 Documentation : website. URL: https://docs.sqlalchemy.org/en/20/ (accessed: 08.05.2025) [in English].

9. LangGraph. Workflows and Agents. *LangGraph.js Documentation* : website. URL: https://langchain-ai.github.io/langgraphjs/tutorials/workflows/#orchestrator-worker (accessed: 08.05.2025) [in English].

10. Langchain. Toolkits. *Langchain Python Documentation* : website. URL: https://python.langchain.com/docs/concepts/tools/#toolkits (accessed: 12.05.2025) [in English].

11. Langchain. create_react_agent. *Langchain Python API Reference* : website. URL: https://python.langchain.com/api_reference/langchain/agents/langchain.agents.react.agent.create_react_agent.html (accessed: 08.05.2025) [in English].

12. Dzhoshi A., Pavlenko V.I. Porivniannia metodiv mashynnoho navchannia dlia vyznachennia feikovykh novyn [Comparison of machine learning methods for detecting fake news]. *Infokomunikatsiini ta kompiuterni tekhnolohii* [Infocommunication and computer technologies]. 2024. Vol. 1. № 07. P. 46–55. DOI: 10.36994/2788-5518-2024-01-07-06 [in Ukrainian].

13. Li X., Zhang Y., Malthouse E.C. Large Language Model Agent for Fake News Detection. arXiv. 2024. URL: https://arxiv.org/html/2405.01593v1 (accessed: 01.05.2025) [in English].

Дата першого надходження рукопису до видання: 27.05.2025
Дата прийнятого до друку рукопису після рецензування: 23.06.2025
Дата публікації: 25.07.2025

УДК 004.021, 004.588, 004.78
DOI https://doi.org/10.36994/2788-5518-2025-01-09-19

СИМУЛЯТОРИ ДРОНІВ: ТРЕНУВАННЯ В БЕЗПЕЧНИХ УМОВАХ

Середа А. В., *Відкритий міжнародний університет розвитку людини «Україна», Інститут комп'ютерних технологій, Київ, Україна. https://orcid.org/0000-0002-4779-5475, andrew.sereda@gmail.com*

Даценко І. П.*, к.т.н., доц., Відкритий міжнародний університет розвитку людини «Україна», Інститут комп'ютерних технологій, Київ, Україна. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Анотація. *У статті розглядається застосування віртуальних симуляторів дронів як безпечного, ефективного й економічно доцільного інструмента підготовки операторів безпілотних літальних апаратів (БПЛА). З огляду на стрімкий розвиток безпілотних технологій і зростання потреби у висококваліфікованих операторах у різних галузях – від оборонної до аграрної – питання організації якісного й безпечного навчання є надзвичайно актуальним. Традиційна підготовка потребує значних ресурсів, пов'язана з ризиками ушкодження обладнання, а також залежить від погодних умов. Тому все частіше використовуються симулятори, які дають змогу тренувати навички керування у віртуальному середовищі. Метою дослідження було провести порівняльну оцінку шести популярних на ринку програмних платформ: DRL Simulator, Liftoff, Uncrashed, AI Drone Simulator, FPV SkyDive та FPV Freerider. Оцінювали за п'ятьма критеріями: реалістичністю моделювання, адаптивністю до різних сценаріїв, зручністю інтерфейсу користувача, можливістю інтеграції з реальним обладнанням (пультами, окулярами, контролерами) і загальною доступністю (вартість, сумісність, системні вимоги). Для об'єктивного вибору використано метод аналізу ієрархій (AHP), який передбачає побудову ієрархічної моделі з урахуванням вагових коефіцієнтів критеріїв. Проведено опитування 10 експертів: викладачів навчальних центрів, інженерів БПЛА-платформ і сертифікованих операторів. Респонденти оцінили важливість кожного критерію за 10-бальною шкалою, після чого дані нормалізовані для усунення індивідуальних упереджень. За отриманими оцінками розраховано інтегральний бал P для кожного симулятора. Лідером став DRL Simulator, який показав найвищі результати завдяки реалістичності польоту й широким можливостям налаштування сценаріїв. За ним – Liftoff та Uncrashed. Найнижчі бали отримали FPV Freerider та FPV SkyDive. Результати дослідження демонструють ефективність AHP як методу підтримки прийняття рішень при виборі навчального ПЗ. Стаття містить рекомендації для освітніх центрів, дослідницьких лабораторій, операторів дронів і розробників тренажерів, зацікавлених у впровадженні ефективних навчальних рішень у сфері БПЛА.*

Ключові слова: *БПЛА, віртуальні навчальні полігони, дрон, тренажер, моделювання середовища, сценарії місій, адаптивне навчання, цифрові тренування, алгоритми польоту, стимуляційні платформи, автоматизація управління.*

DRONE SIMULATORS: TRAINING IN SAFE CONDITIONS

Andrii Sereda*, Open University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0000-0002-4779-5475, andrew.sereda@gmail.com*

Ivan Datsenko*, Ph.D., Open University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Abstract. *The article explores the implementation of virtual drone simulators as a safe, efficient, and cost-effective alternative to traditional UAV (Unmanned Aerial Vehicle) operator training. With the increasing demand for skilled drone pilots in both civilian and defense sectors, simulation-based training offers an attractive solution by reducing risk, lowering operational costs, and providing consistent learning conditions. The study evaluates six widely used drone simulation platforms – DRL Simulator, Liftoff, Uncrashed, AI Drone Simulator, FPV SkyDive, and FPV Freerider – according to five comprehensive criteria: (1) realism of flight dynamics modeling, (2) adaptability to various flight environments and training scenarios, (3) usability and intuitiveness of the user interface, (4) capability to integrate with real UAV hardware and controllers, and (5) overall accessibility, including cost, hardware requirements, and availability across platforms. The Analytic Hierarchy Process (AHP) was employed as a structured methodology for multi-criteria decision-making. Ten subject-matter experts – including certified drone pilots, UAV instructors, and aerospace engineers – participated in a survey to assess the importance of each criterion on a 10-point scale. The collected data were normalized to reduce personal bias, and final scores (denoted as P) were computed for each simulator. The results showed that DRL Simulator ranked highest due to its superior realism and training adaptability. Liftoff and Uncrashed followed, showing strong overall performance. Simulators like AI*

Drone Simulator, FPV SkyDive, and FPV Freerider scored lower, often limited by hardware integration and interface quality. The findings, visualized through comparative charts, confirm the effectiveness of AHP in simulator evaluation and offer practical guidance for training centers, research institutions, and engineers choosing optimal software for UAV operator development.

Key words: *UAV, virtual training grounds, drone, simulator, environment modeling, mission scenarios, adaptive learning, digital training, flight algorithms, simulation platforms, control automation.*

Вступ

У сучасних умовах стрімкого розвитку безпілотних технологій зростає потреба у висококваліфікованих операторах дронів. Одним із ключових елементів ефективної підготовки є використання віртуальних навчальних полігонів – спеціалізованих програмних платформ, відомих як симулятори дронів. Вони моделюють широкий спектр реальних умов керування безпілотними літальними апаратами (далі – БПЛА), що дає змогу навчатися без ризику пошкодження техніки або загрози для безпеки людей. Симулятори дронів стають важливим складником сучасної освітньої інфраструктури, аналогічно до мобільної освіти, яка забезпечує доступність і гнучкість навчального процесу. Завдяки адаптивним сценаріям тренування оператори можуть набувати навичок реагування на нестандартні ситуації, відпрацьовувати техніку польоту в різних погодних умовах і вдосконалювати стратегії місій. Таке безпечне середовище дає змогу значно прискорити процес засвоєння необхідних знань і вмінь. Розвиток симуляторів відкриває нові можливості для стандартизації підготовки операторів, зменшуючи витрати на реальні тренування й підвищуючи загальний рівень безпеки у сфері експлуатації БПЛА.

Виконання досліджень

Оцінювання симуляторів дронів здійснювалося методом аналізу ієрархій (AHP) за низкою ключових критеріїв, що визначають ефективність тренувального середовища. До переліку критеріїв увійшли реалістичність моделювання польотної динаміки, адаптивність до різноманітних сценаріїв використання, зручність і доступність інтерфейсу користувача, можливість інтеграції симулятора з реальними апаратними платформами, а також загальна доступність програмного забезпечення для кінцевого користувача. Для оцінювання обрано шість найбільш популярних на ринку симуляторів: FPV SkyDive, Uncrashed, Liftoff, FPV Freerider, AI Drone Simulator і The Drone Racing League Simulator. Кожен симулятор оцінювався за визначеними критеріями шляхом експертного опитування. У дослідженні взяли участь 10 експертів, яких поділено на три основні категорії: викладачі навчальних центрів із підготовки операторів БПЛА, технічні інженери безпілотних платформ і сертифіковані оператори дронів. Учасникам запропоновано оцінити важливість кожного критерію за 10-бальною шкалою (де 1 – мінімальна важливість, 10 – максимальна). Отримані результати нормалізовані для усунення суб'єктивних відхилень. На основі нормалізації сформовано підсумкову таблицю середніх оцінок критеріїв і розраховано нормалізовані вагові коефіцієнти для подальшого аналізу.

Таблиця 1

Середнє значення й нормалізована вага по критеріях

Критерій	Середнє значення	Нормалізована вага
Реалістичність моделювання (R)	8,8	0,29
Адаптивність сценаріїв (A)	7,6	0,25
Інтерфейс користувача (U)	5,2	0,17
Інтеграція з обладнанням (I)	5,4	0,18
Вартість/доступність (C)	3,0	0,10

Отримали таблицю середніх оцінок симуляторів згідно з критеріями, наданими експертами.

Таблиця 2

Оцінка симуляторів за критеріями

Симулятор	R	A	U	I	C
FPV SkyDive	7	6	9	6	10
Uncrashed	9	8	8	9	7
Liftoff	8	10	8	9	6
FPV Freerider	6	5	7	5	10
AI Drone Simulator	8	9	7	8	7
DRL Simulator	9	10	9	10	6

Застосуємо інтегральний розрахунок методом аналізу ієрархій для обчислення балу кожного із симуляторів:

$$P = R \cdot 0.29 + A \cdot 0.25 + U \cdot 0.17 + I \cdot 0.18 + C \cdot 0.10 \quad (1),$$

де
- P – інтегральний бал симулятора,
- R – оцінка реалістичності моделювання,
- A – оцінка адаптивності до різних сценаріїв,
- U – оцінка зручності інтерфейсу користувача,
- I – оцінка інтеграції з реальним обладнанням,
- C – оцінка доступності програмного забезпечення,
- коефіцієнти 0,29, 0,25, 0,17, 0,18 і 0,10 – нормалізовані ваги відповідних критеріїв, отримані шляхом опитування експертів.

Отримані значення P для шести симуляторів, що оцінюються, відображено на графіку.

Висновок

FPV SkyDive – безкоштовний симулятор від компанії ORQA, орієнтований на початківців. Експерти високо оцінили інтерфейс користувача (9/10): він інтуїтивний, з мінімальним навантаженням для новачків. Модель польоту досить спрощена (7/10), тому симулятор краще підходить для базового навчання, а не професійного тренування. Сценарії обмежені (6/10), проте симулятор відмінно справляється з імітацією стартових умов. FPV SkyDive підтримує стандартні геймпади та джойстики (6/10) і має головну перевагу – повну безкоштовність (10/10).

Uncrashed FPV Drone Simulator високо оцінений за реалістичністю моделі польоту (9/10), що робить його придатним для просунутих тренувань. Адаптивність сценаріїв (8/10) дає змогу створювати треки, що відповідають різним умовам польотів, зокрема в закритих приміщеннях і на вулиці. Інтерфейс простий, але функціональний (8/10). Важливою перевагою є гарна інтеграція з більшістю FPV-контролерів (9/10). Ціна помірна (7/10), що робить цей симулятор збалансованим варіантом для індивідуального й навчального використання.

Liftoff – один із найвідоміших FPV-симуляторів, який активно використовують як любителі, так і професіонали. Він отримав найвищу оцінку за адаптивність сценаріїв (10/10): у грі доступна велика кількість трас, редактор треків, режим кар'єри та мультиплеєр. Реалістичність польоту (8/10) також високо оцінена, хоча деякі пілоти відзначають незначні спрощення в аеродинаміці. Інтерфейс зручний (8/10), а підтримка обладнання майже повна (9/10). Недоліком є вартість (6/10), проте вона виправдана широким функціоналом.

FPV Freerider – один із найперших симуляторів FPV-дронів. Він вирізняється своєю простотою та доступністю, має низькі системні вимоги й безкоштовну демо-версію. Але реалістичність польоту – на базовому рівні (6/10), що обмежує можливості професійної підготовки. Адаптивність (5/10) мінімальна, усього кілька треків без гнучкого редагування. Інтерфейс простий, але застарілий (7/10). Підтримка контролерів не завжди стабільна (5/10). Основна перевага – повна доступність (10/10), що робить його хорошим стартом для новачків.

AI Drone Simulator – симулятор з відкритим кодом, орієнтований не лише на FPV, а й на дослідження в галузі штучного інтелекту. Його фізика польоту оцінена на рівні 8/10, оскільки базується на реалістичному русі. Гнучкість сценаріїв висока (9/10), включаючи підтримку кастомних сцен і симуляцію сенсорів. Інтерфейс дещо технічний (7/10), тому новачкам може бути складно. Інтеграція з реальним обладнанням добра (8/10), особливо через API. Ціна – безкоштовний (7/10), хоча потрібно більше часу для налаштування.

The Drone Racing League (DRL) Simulator – офіційний симулятор Ліги перегонів дронів, розроблений спеціально для професійного FPV-тренування. Він лідирує в усіх категоріях: реалістичність – 9/10 (польоти майже ідентичні реальним трасам DRL), адаптивність – 10/10 (великий вибір треків, щотижневі змагання, тренувальні місії), інтерфейс – 9/10 (професійний, але інтуїтивний). Підтримка контролерів повна (10/10), особливо фірмового пульта DRL. Вартість не найнижча (6/10), але виправдана якістю. Ідеально підходить для серйозної підготовки гонщиків-дронів.

За результатами AHP-аналізу, The DRL Simulator є найкращим FPV-симулятором для підготовки операторів дронів. Його переваги: виняткова реалістичність, кастомні треки, інтеграція з реальними пультами й актуальна підтримка індустрії. Liftoff – відмінний вибір для тренувань у середовищі симуляції з відкритими треками й мультиплеєром. Uncrashed займає третє місце завдяки балансу між реалістичністю й ціною. Також хочеться відмітити FPV Freerider за низькі системні вимоги та легкий поріг входу для новачків.

Література

1. Холл Дж., Браун М. Симуляція польоту FPV-дронів : технічний посібник. DroneTech Publishing, 2020.
2. Команда розробників Liftoff. Liftoff: Симулятор перегонів FPV-дронів : офіційний посібник користувача. LuGus Studios, 2023.
3. Чень К., Чжао Х. Оцінка симуляторів БПЛА для підготовки пілотів на основі фізичних рушіїв реального часу. *Журнал безпілотних систем.* 2022. № 10 (3). С. 115–127.
4. Команда ORQA. FPV.SkyDive: Нотатки з розробки та принципи симуляції. ORQA Ltd, 2021.

References

1. Hall, J., & Brown, M. (2020). FPV Drone Flight Simulation: A Technical Guide. DroneTech Publishing.
2. Liftoff Dev Team. (2023). Liftoff: FPV Drone Racing Simulator – Official User Manual. LuGus Studios.
3. Chen, K., & Zhao, H. (2022). Evaluation of UAV Simulators for Pilot Training Based on Real-Time Physics Engines. Journal of Unmanned Systems, 10(3), 115–127.
4. ORQA Team. (2021). FPV.SkyDive: Development Notes and Simulation Principles. ORQA Ltd.

Дата першого надходження рукопису до видання: 20.05.2025
Дата прийнятого до друку рукопису після рецензування: 12.06.2025
Дата публікації: 25.07.2025

УДК 004.9
DOI https://doi.org/10.36994/2788-5518-2025-01-09-20

СИНТЕЗ ПЕРЦЕПТРОННОЇ СТРУКТУРИ ДЛЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ЕЛЕМЕНТІВ ТЕКСТУ

Топалов А. М., *к.т.н., доц., Національний університет кораблебудування ім. адмірала Макарова, Миколаїв, Україна. https://orcid.org/0000-0002-0959-4357, andrii.topalov@nuos.edu.ua*

Герасін О. С., *к.т.н., доц., Національний університет кораблебудування ім. адмірала Макарова, Миколаїв, Україна. https://orcid.org/0000-0001-5107-9677, oleksandr.gerasin@nuos.edu.ua*

Мануляк І. З., *к.т.н., доц. Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. https://orcid.org/0000-0002-0072-1532, manulyak-iryna@ukr.net*

Стрілецький Ю. Й., *д.т.н. проф., Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. https://orcid.org/0000-0002-0105-8306, momental@ukr.net*

Анотація. *У статті розглядається проблема синтезу нейронної мережі на основі перцептронів кореляційного типу для задачі автоматизованого розпізнавання символів латинського алфавіту. Розроблена структура мережі орієнтована на використання вбудованих засобів програмного середовища Matlab, зокрема його спеціалізованих функцій і бібліотек машинного навчання. Як навчальні дані для формування вхідних і вихідних параметрів нейронної мережі застосовано стандартний генератор prprob, який створює бінарні растрові представлення латинських літер розміром 5×7 пікселів. Результатом є матриця вхідних даних X і цільових значень T, що формують основу для навчання мережі.*

Основна увага приділена синтезу двошарової нейронної мережі прямого поширення сигналу. Перший шар побудовано з 25 прихованих модифікованих перцептронів, що використовують сигмоїдні функції активації, другий шар є лінійним. Це забезпечує підвищену гнучкість порівняно з класичними перцептронами з пороговими функціями активації. Окремо розглядається питання стабілізації випадкової ініціалізації вагових коефіцієнтів для забезпечення відтворюваності результатів навчання. Навчання першої нейронної мережі виконано на еталонних зображеннях літер.

Також детально описано експериментальні дослідження стійкості нейронної мережі до зашумлених вхідних даних. Навчання другої мережі проводилося з урахуванням додаткового шуму у вхідних образах. Аналіз результатів показав, що модель, навчена на зашумлених даних, демонструє менший відсоток помилок розпізнавання при різних рівнях шуму, що підтверджено графічними й числовими результатами. Подано алгоритм обчислення середньоквадратичної помилки (MSE) для оцінювання якості навчання й результати числового моделювання роботи нейромереж, навчених за різної кількості навчальних прикладів.

Результати дослідження можуть бути використані для розробки адаптивних систем розпізнавання тексту, здатних працювати в умовах зовнішніх перешкод, а також як приклад побудови практичних нейронних мереж у середовищі Matlab для навчальних цілей.

Ключові слова: *перцептрон, імовірнісні оцінки, цифрові компоненти, нейронне навчання, логічні функції.*

SYNTHESIS OF A PERCEPTRONIC STRUCTURE FOR THE PROBLEM OF RECOGNITION OF TEXT ELEMENTS

Peter Topalov, *Ph.D., Ass. Prof., Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. https://orcid.org/0000-0002-0959-4357, andrii.topalov@nuos.edu.ua*

Oleksandr Gerasin, *Ph.D., Ass. Prof., Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. https://orcid.org/0000-0001-5107-9677, oleksandr.gerasin@nuos.edu.ua*

Iryna Manulyak, *Ph.D., Ass. Prof., Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. https://orcid.org/0000-0002-0072-1532, manulyak-iryna@ukr.net*

Yurii Striletskyi*, Doctor of Engineering Science Prof. Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. https://orcid.org/0000-0002-0105-8306, momental@ukr.net*

Abstract. *The article considers the problem of synthesizing a neural network based on correlation-type perceptrons for the task of automated recognition of Latin alphabet characters. The developed network structure is oriented towards the use of built-in tools of the Matlab software environment, in particular its specialized functions and machine learning libraries. As training data for forming the input and output*

parameters of the neural network, the standard prprob generator was used, which creates binary raster representations of Latin letters with a size of 5×7 pixels. The result is a matrix of input data X and target values T, which form the basis for training the network.

The main attention is paid to the synthesis of a two-layer neural network of direct signal propagation. The first layer is built from 25 hidden modified perceptrons using sigmoid activation functions, the second layer is linear. This provides increased flexibility compared to classical perceptrons with threshold activation functions. The method of stabilizing random initialization of weight coefficients to ensure reproducibility of training results is considered separately. The first neural network was trained on reference letter images.

The article also describes in detail experimental studies of the stability of the neural network to noisy input data. The training of the second network was carried out taking into account additional noise in the input images. Analysis of the results showed that the model trained on noisy data demonstrates a lower percentage of recognition errors at different noise levels, which is confirmed by graphical and numerical results. An algorithm for calculating the mean square error (MSE) for assessing the quality of training is presented and the results of numerical simulation of neural networks trained on different numbers of training examples are presented.

The results of the study can be used to develop adaptive text recognition systems capable of operating in conditions of external interference, as well as as an example of building practical neural networks in the Matlab environment for training purposes.

Key words: *perceptron; probabilistic estimates; digital components; neural learning; logical functions.*

Вступ

Однією з основних проблем, що виникає при розв'язанні задач, пов'язаних із розпізнаванням рукописного тексту, є складність формалізації алгоритмічного підходу. Традиційні методи обробки інформації вимагають чіткого визначення правил, за якими система має ідентифікувати кожен символ. Проте в умовах варіативності рукописного письма – зміни нахилу, товщини ліній, індивідуальних особливостей написання – створення універсального алгоритму стикається з численними винятками й унікальними випадками, які складно передбачити на етапі проектування.

З огляду на ці труднощі, значного поширення набув альтернативний підхід на основі штучних перцептронних мереж, які дають змогу вирішувати проблему розпізнавання рукописних символів із використанням навчальних прикладів. Цей підхід ґрунтується на збиранні великого обсягу рукописних символів, які формують навчальну вибірку. На основі цієї вибірки створюється модель, здатна навчатися, виявляючи приховані закономірності, притаманні цим символам.

Нейронна мережа, сприймаючи навчальні приклади, формує власну систему правил для розпізнавання, що не задається явно, а виникає внаслідок узагальнення досвіду. Таким чином, що більше прикладів буде представлено під час навчання, то вища ймовірність того, що модель досягне високої точності в розпізнаванні нових, раніше не бачених символів. Перцептрон, як базова модель штучного нейрона, відіграє фундаментальну роль у структурі та функціонуванні нейронних мереж, що використовуються для задач розпізнавання рукописного тексту. Цей елемент, запропонований Френком Розенблаттом, є першою спробою математичного моделювання біологічного нейрона й поклав початок розвитку сучасних нейромережевих технологій. У контексті розпізнавання рукописних символів, перцептрон виступає як елементарна одиниця обчислення, що дає змогу класифікувати вхідні дані шляхом обчислення лінійної комбінації вхідних сигналів із наступним застосуванням функції активації. Саме завдяки здатності до навчання – зміни вагових коефіцієнтів на основі помилок – перцептрони дають змогу моделі адаптуватися до різноманітних зразків написання символів. Більше того, у багатошарових нейронних мережах, де прості перцептрони об'єднуються в складні архітектури, забезпечується можливість апроксимації складних функцій і розпізнавання високорівневих патернів. Завдяки цій властивості перцептрони стали невід'ємною частиною алгоритмів глибокого навчання, що демонструють високу ефективність у задачах класифікації, зокрема у сфері оптичного розпізнавання символів (далі – OCR).

Упродовж останніх років число дослідників у галузі розпізнавання рукописного тексту нейронними мережами значно зросло. Це зумовлено широким застосуванням OCR у цифровізації документів, автоматизації вводу даних та освітніх проєктах. Особливо активні дослідження ведуться в Україні й за кордоном: учені орієнтуються на оптимізацію моделей на базі перцептронів, CNN, RNN, Transformer-архітектур. Велику увагу приділяють адаптації підстроювання під особливості мови й рукописні зразки різних культур і мов. Автори [1] запропонували архітектуру на базі лише 1D-CNN шарів, яка дає змогу досягти продуктивності, порівнянної з RNN/Transformer-моделями. Уведено нову стратегію аугментації «Tiling and Corruption (TACO)» та модуль Squeeze and Excitation, що значно підвищують точність. Модель показала найкращі результати на базі IAM при роботі з обмеженим набором тренувальних даних. У праці [2] запропоновано підхід «Text Perceptron», що є скінчено-регулярною моделлю для розпізнавання тексту нестандартної форми. Завдяки Shape Transform Module забезпечується єдина модель для детекції й розпізнавання зі склеєним оптимізаційним фронтом. Досягнуто високої точності на ICDAR і SCUT CTW1500. У роботі [3] модель Self ONN із самоорганізо-

ваними операційними нейронами замінює CNN, використовуючи нелінійну параметризацію нейронів. Використання deformable convolutions забезпечило кращу адаптацію до різноманітності рукописних зразків. Результати на IAM і HADARA80P показали зменшення CER і WER порівняно з класичними CNN-архітектурами. У статті [4] модель DAN об'єднує сегментацію та розпізнавання в єдиний трансформер-базований конвеєр без явних сегментаційних міток. Для покрокового виведення символів зі структурою XML використовується FCN-енкодер і стек Transformer-декодерів. Досягнуто CER = 3,43 % на READ 2016, що є конкурентоспроможним результатом. У праці [5] представлено рекурсивні нейромережні моделі для зменшення розміру моделі без утрати точності. Використано повторне застосування ваг, масштабування ядерних функцій і повторювані шари, що дає змогу зменшити обсяг пам'яті. Показано, що такі моделі зберігають або підвищують точність на HTR-бенчмарках порівняно зі стандартними CNN/RNN. Український дослідник у роботі [6] порівняв щільні нейронні мережі (DNN) і CNN у розпізнаванні рукописних цифр. Виявлено, що DNN втрачають продуктивність зі збільшенням розміру вхідного зображення, тоді як CNN забезпечують сталість і точність. Робота підтверджує важливість використання перцептронів у шарі класифікації та архітектурних рішеннях для інтеграції локальних ознак.

Таким чином, значення перцептронів у нейронних мережах полягає не лише в їхній історичній ролі, а й у практичній ефективності як базового елемента, що забезпечує здатність системи до узагальнення, адаптації та точного розпізнавання складних візуальних структур.

Мета дослідження полягає в розробленні системи розпізнавання символів за допомогою штучної нейронної мережі, реалізованої у вигляді програмного засобу. Передбачається створення програмного засобу, здатного приймати на вхід літери, здійснювати їх обробку за допомогою навченої моделі та повертати показники точності розпізнавання.

Особливу увагу в рамках дослідження планується приділити динаміці навчального процесу, а саме забезпеченню стабільного зростання точності мережі на кожному етапі навчання.

Особливості функціонування перцептрона та вимоги до системи розпізнавання елементів тексту

Перцептрон є однією з фундаментальних моделей штучних нейронних мереж, яка має суттєве аналітичне значення в теорії машинного навчання й обчислювального інтелекту [7]. Його робота ґрунтується на обробці кількох вхідних сигналів, які, як правило, подаються у вигляді двійкових значень. Кожен вхідний сигнал позначається як x_i, де й змінюється від 1 до n, що відповідає кількості входів нейронного елемента.

Основним функціональним елементом класичного перцептрона є так званий пороговий елемент, який здійснює лінійне зважування вхідних сигналів і приймає рішення про активацію виходу на основі порівняння результату цієї операції з фіксованим або адаптивним порогом. Формально вихідний сигнал у цього елемента може набувати лише двох дискретних значень: 0 або 1, що відображає принцип бінарної класифікації [8].

Математичний опис роботи порогового елемента має вигляд:

$$y = \begin{cases} 0, & якщо \sum_{i=1}^{n} w_i x_i \geq w_0, \\ 1, & якщо \sum_{i=1}^{n} w_i x_i < w_0, \end{cases}$$

де w_i – вагові коефіцієнти, що відповідають значущості кожного входу, а w_0 − величина порога спрацьовування елемента.

Першу технічну модель зорового аналізатора, яку розробив Ф. Розенблатт, автор назвав перцептроном (від англійського слова *perception* − сприйняття). Структура первинної моделі Розенблатта включала рецепторний шар S, що складався з 400 фотоелементів, які формували прямокутне рецепторне поле розміром 20×20 елементів. Кожен фотоелемент функціонував як елементарний сенсорний детектор, що перетворював оптичне зображення на електричні сигнали. Інформація з рецепторного шару передавалася на вхід елементів наступного рівня − порогових нейронів перетворюючого шару (елементів А). Усього в моделі реалізовано 512 таких елементів. Кожен нейрон шару А мав по 10 входів, що випадковим чином підключалися до різних фотоелементів рецепторного поля. Половина із цих входів мала збуджуючий характер із коефіцієнтом посилення +1, інша половина – гальмівний характер із коефіцієнтом –1. Порогове значення для активації нейрона встановлювалося рівним нулю, що відповідало бінарній моделі роботи елемента (активація або неактивація). Вихідні сигнали нейронів перетворюючого шару А подавалися на вхід реагуючого нейрона – елемента R, який виконував роль інтегративного вузла, здійснюючи остаточне рішення про розпізнавання вхідної інформації [8–11].

Розглянемо математичну модель функціонування класичного перцептрона, яка стала фундаментом для подальшого розвитку штучних нейронних мереж:

1. **Формування вхідного сигналу.** На першому етапі в рецепторному полі формується сигнал, який відповідає зовнішньому подразнику. Цей сигнал подається у вигляді векторної величини x. Кож-

не рецепторне закінчення (рецептор) передає досить простий сигнал – або генерує імпульс, або не генерує його. З математичного погляду це означає, що вектор x є бінарним, тобто його компоненти приймають значення лише з множини 0 або 1. Таким чином, рецепторне поле здійснює кодування вхідної інформації у вигляді бінарного вектора.

2. **Передача й перетворення сигналів другим шаром нейронів.** Сформований бінарний вектор x передається до другого шару нейронів, де відбувається його трансформація у вихідний вектор y, який також є бінарним. Цей процес описується функціональним співвідношенням $y = f(x)$. Характер перетворення визначається двома основними принципами:

а) **пороговий характер обчислень:** трансформація вхідних сигналів здійснюється за допомогою порогових елементів, кожен із яких активується, коли сума вхідних сигналів, помножених на відповідні вагові коефіцієнти, перевищує заданий поріг;

б) **випадкове з'єднання входів:** зв'язки між рецепторами першого шару й пороговими елементами другого шару формуються випадковим чином, що імітує випадкову організацію нейронних зв'язків у біологічних мережах.

3. **Класифікація на основі активації вихідних нейронів.** Модель передбачає, що перцептрон відносить вхідний вектор до певного поняття j, якщо активується лише j-й реагуючий нейрон, а всі інші нейрони залишаються неактивними. Інакше кажучи, вихідний бінарний вектор $y = (y_1, y_2, \ldots, y_m)$ задовольняє таку систему нерівностей:

$$\sum_{i=1}^{m} w_{i,j} y_i \geq 0,$$

$$\sum_{i=1}^{m} w_{i,k} y_i < 0, \text{для всіх} \; k \neq j.$$

У цих нерівностях $w_{i,k}$ – коефіцієнти посилення *k*-го реагуючого нейрона.

4. Процес формування понять у структурі перцептрона, запропонованого Ф. Розенблаттом, ґрунтується на коректному налаштуванні вагових коефіцієнтів, що відповідають кожному елементові шару реагуючих нейронів (позначимо цей шар як R). Основною метою є встановлення таких значень ваг, які забезпечують правильне реагування моделі на вхідні сигнали.

Нехай до цього моменту існують деякі ваги елементів R і $w_{1,j}$, $w_{2,j}$, …, $w_{n,j}$ − ваги j-того елемента. У момент часу t для класифікації на вхід перцептрона надходить сигнал, що описується вектором x(t). Вектор x(t) може або відповідати поняттю j або не відповідати йому. Розглянемо обидва ці випадки.

Випадок перший. У цьому випадку очікуваною реакцією j-го нейрона є активація, тобто правильна відповідь системи полягає в тому, що нейрон збуджується у відповідь на поданий сигнал x(t). Математично це виражається у вигляді такої нерівності:

$$\sum_{i=1}^{m} w_{i,j} y_i \geq 0.$$

Якщо ваги елемента j забезпечують правильну реакцію на вектор x(t), то вони не змінюються. Якщо ж ваги не забезпечують правильної реакції елемента j, тобто вони такі, що

$$\sum_{i=1}^{m} w_{i,j} y_i < 0,$$

то ваги елемента j змінюються за правилом:

$$w_{i,j}(t) = w_{i,j}(t-1) + y_i.$$

Випадок другий. У ситуації, коли вхідний вектор x(t) не належить до поняття j, правильна реакція перцептрона полягає у відсутності активації відповідного нейронного елемента. Це означає, що елемент j має залишатися неактивним у відповідь на цей стимул. Математично така умова виражається у вигляді такої нерівності:

$$\sum_{i=1}^{m} w_{i,j} y_i \geq 0,$$

Якщо поточні значення ваг забезпечують правильну реакцію елемента j, тобто коли наведена нерівність виконується, ваги залишаються незмінними. Однак якщо вага елемента j не забезпечує правильної реакції, тобто виконується протилежна умова:

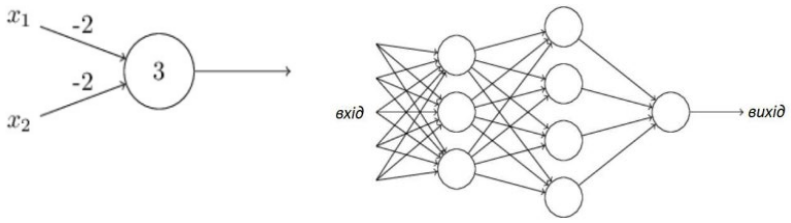
$$\sum_{i=1}^{m} w_{i,j} y_i \geq 0,$$

то ваги змінюються відповідно до правила:

$$w_{i,j}(t) = w_{i,j}(t-1) - y_i.$$

При здійсненні навчання аналогічно змінюються ваги всіх R-елементів перцептрона. Таким чином, коригування вагових коефіцієнтів усіх R-елементів забезпечує узгоджену адаптацію всієї мережі, що сприяє підвищенню точності класифікації.

Вигляд перцептрона та зібраної з перцептронів структури показаний на рис. 1. Дані зображення перцептрона й перцептронної структури вважаються загальними для першочергового етапу проєктування систем розпізнавання [12].



а б

Рис. 1. Перцептрон і його структура: а) окремий перцептрон з ваговими коефіцієнтами та загальним зміщенням; б) математична модель мережі перцептронів

Аналізуючи схему, представлену на рисунку 1, б, можемо спостерігати наявність декількох вертикальних груп елементів, які утворюють структуровані ряди, або шари перцептронів [13]. Така багаторівнева структура є характерною рисою багатошарових нейронних мереж, де кожен наступний шар здійснює обробку інформації, яка є результатом роботи попереднього шару.

Перший шар перцептронів відповідає за первинний рівень прийняття рішень, здійснюючи обчислення на основі зваженого підсумовування вхідних сигналів. Кожен нейрон цього шару аналізує сирі дані, що надходять безпосередньо із зовнішнього середовища або рецепторних елементів. Другий шар перцептронів приймає на вхід результати обробки, здійсненої першим шаром, і виконує більш складну обробку, комбінуючи й переосмислюючи ці проміжні результати. Таким чином, другий шар здатен виділяти складніші закономірності та зв'язки, ніж перший. Аналогічно на третьому рівні (третьому шарі) відбувається ще глибше узагальнення інформації, що дає змогу нейронній мережі формувати висновки ще вищого рівня абстракції. Цей підхід демонструє ієрархічну природу обробки даних у багатошарових нейронних мережах, де кожен наступний шар виконує дедалі складніші функції аналізу. Окремо варто зазначити, що в класичному вигляді перцептрон, як модель окремого нейрона, має лише один вихідний сигнал, що відповідає його бінарному рішенню − активації або відсутності активації. Однак на схемі, представленій на рисунку 1, для наочності показано, що цей вихід одночасно може бути підключений до кількох нейронів наступного шару. Таким чином, формується уявлення про множинне використання одного вихідного сигналу: один вихідний сигнал перцептрона може виступати як кілька вхідних сигналів для різних нейронів наступного шару. Це не означає наявність декількох фізичних виходів у перцептрона, а лише відображає можливість розгалуження інформаційного потоку у структурі мережі.

Отже, наведена модель демонструє принцип пошарової обробки інформації, де кожен шар виконує функцію узагальнення й ускладнення прийнятих рішень, а вихідні сигнали кожного нейрона можуть одночасно впливати на декілька наступних елементів мережі, формуючи таким чином складну багаторівневу систему взаємозв'язків.

Виконання досліджень

Розглянемо задачу розпізнавання будь-якого із символів латинського алфавіту за допомогою перцептронної нейронної мережі, яка створюється та навчається за допомогою вбудованих функцій Matlab [14]. Для вирішення цієї задачі доцільно розробити відповідну методику, яка містить такі кроки:

1. Генерування масиву вхідних даних (букв латинського алфавіту). Для цього необхідно запустити вбудований скрипт Matlab:

[X, T] = prprob;

Скрипт prprob визначає матрицю X з 26 колонками, по одній для кожної букви алфавіту. Кожна колонка має 35 значень, які можуть дорівнювати або 1 або 0. Кожен стовпець із 35 значень являє собою растрове зображення букви 5x7. Матриця T − це одинична матриця розміром 26x26, яка відображає 26 вхідних векторів у 26 класів.

Необхідно зазначити, що для роботи команди prprob необхідно встановлювати tollbox deep learn або повністю задавати її як функцію в коді програми перед викликом.

2. Вибір і побудова зображення еталонної літери для розпізнавання (N − номер літери):

plotchar (X (:, N));

У результаті виконання цієї команди отримується зображення еталонної літери у вигляді бітовій карти.

3. Створення першої двошарової перцептронної нейронної мережі прямого зв'язку з 25 прихованими нейронами для класифікації букв латинського алфавіту:

```
setdemorandstream(pi);
net1 = feedforwardnet(25);
view(net1)
```

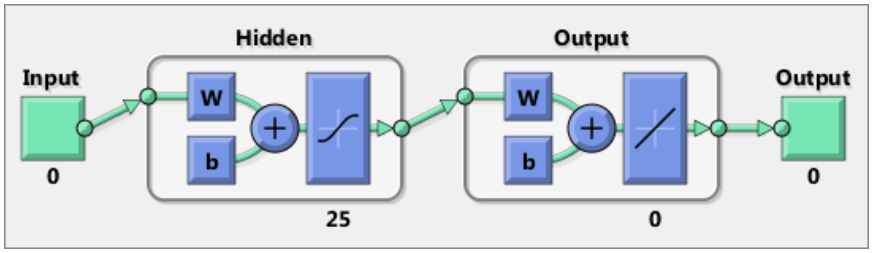


Рис. 2. Ініційована нейронна мережа

Оскільки нейронна мережа ініціюється випадковими початковими вагами, результати після навчання будуть незначно відрізнятися кожного разу після запуску скрипту. Щоб позбутися цієї випадковості, випадкова стала за допомогою функції *setdemorandstream* встановлюється константою. У результаті виконання скипту отримується зображення нейронної мережі прямого поширення, що наведено на рис. 2. Ця мережа містить 2 шари, причому модифіковані перцептрони першого шару містять сигмоїдні функції активації, а другого – лінійні. Такий підхід дає змогу розширити межі застосування нейронних мереж на базі класичних перцептронів з пороговими функціями активації.

4. Навчання створеної нейронної мережі за допомогою функції train:

```
net1.divideFcn = '';
net1 = train(net1,X,T,nnMATLAB);
```

Функція *train* ділить вхідні дані на три масиви – для навчання, валідації (перевірки) і тестування (випробування) нейронної мережі. Навчальний набір використовується для оновлення вагових коефіцієнтів мережі. Валідаційний набір даних потрібен для зупинки мережі, перш ніж відбудеться перенасичення навчальними даними. Набір для тестування виступає як повністю незалежна міра того, наскільки добре можна очікувати виконання поставленої задачі мережею на нових вхідних сигналах. Навчання припиняється, коли мережа вже не може покращити свої результати за рахунок навчальних або валідаційних наборів даних. Властивість *divideFcn* об'єкта нейронної мережі визначає, яким чином дані автоматично поділятимуться на ці три набори (підмножини) перед навчанням. Для першої нейронної мережі вся вибірка використовується для навчання. Структура навченої нейронної мережі наведена на рисунку 3.

Навчання першої нейронної мережі тривало 11 ітерацій (4 с) методом Левенберг-Марквардта за критерієм мінімуму середньоквадратичної помилки ($mse_{min} = 1{,}2 \cdot 10^{-16}$).

Далі доцільно дослідити, як мережа буде розпізнавати не лише ідеально сформовані літери, а і їх зашумлені версії. Тому необхідно створити й навчити другу мережу на зашумлених даних і порівняти її здатність до узагальнення з першою мережею.

5. Створення масиву вхідних даних з додатковим забрудненням сигналів:

```
numNoise = 30;
Xn = min(max(repmat(X,1,numNoise)+randn(35,26*numNoise)*0.2,0),1);
Tn = repmat(T,1,numNoise);
```

Цей скрипт створює набори даних із 30 копій кожної букви з масиву X, що додатково забруднені шумом. Причому забрудненість шумом виражається в тому, що значення відповідних клітин масиву представлення частини літери (5x7) обмежені на проміжку між 0 і 1. Також визначено відповідні цілі Tn.

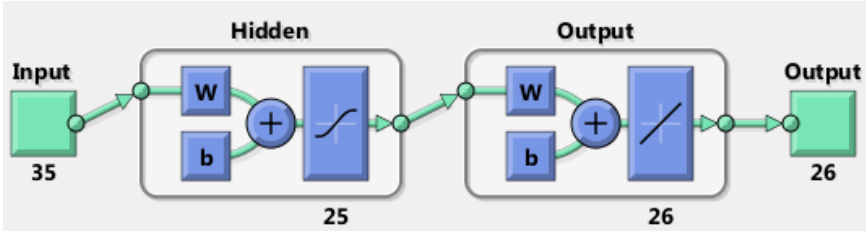
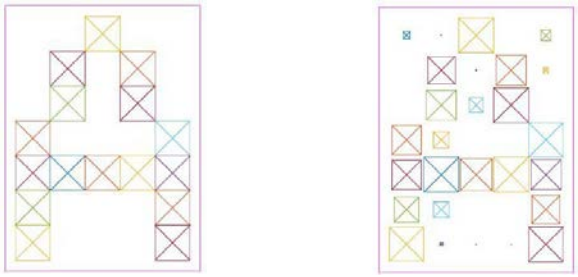


Рис. 3. Структура навченої перцептронної нейронної мережі
для класифікації букв латинського алфавіту

6. Побудова зображення еталонної букви, забрудненої шумом:

```
figure
plotchar(Xn(:,N));
```

Порівняння еталонного зображення й забрудненого шумом зображення для літери «А» наведено на рис. 4 (результати виконання кроків 2 та 6 методики).



а б

Рис. 4. Візуалізація літер у вигляді бітових карт:
а – еталонне зображення; б – забруднене шумом зображення

7. Створення й навчання нейронної мережі для роботи з забрудненими даними:

```
net2 = feedforwardnet(25);
net2 = train(net2,Xn,Tn,nnMATLAB);
```

Навчання другої нейронної мережі тривало 28 ітерацій (101 с) методом Левенберг-Марквардта за критерієм мінімуму середньоквадратичної помилки ($mse_{min} = 1{,}345 \cdot 10^{-3}$).

8. Аналіз роботи обох синтезованих нейронних мереж (першої та другої) для 30 наборів зашумлених даних за допомогою такого скрипту:

```
noiseLevels = 0:.05:1;
numLevels = length(noiseLevels);
percError1 = zeros(1,numLevels);
percError2 = zeros(1,numLevels);
for i = 1:numLevels
  Xtest = min(max(repmat(X,1,numNoise)+randn(35,26*numNoise)*noiseLevels(i),0),1);
  Y1 = net1(Xtest);
  percError1(i) = sum(sum(abs(Tn-compet(Y1))))/(26*numNoise*2);
  Y2 = net2(Xtest);
  percError2(i) = sum(sum(abs(Tn-compet(Y2))))/(26*numNoise*2);
end
figure
plot(noiseLevels,percError1*100,'--',noiseLevels,percError2*100);
title('Percentage of Recognition Errors');
xlabel('Noise Level');
ylabel('Errors');
legend('Network 1','Network 2','Location','NorthWest';
```

Результат порівняння роботи двох нейромереж для різних рівнів шуму наведений на рисунку 5. Виходячи із цього зображення (рис. 5), перша мережа, навчена без шуму, має більші значення помилок через шум, ніж друга мережа, яка була навчена з шумом. Зроблено висновок, що при вирішенні завдання розпізнавання елементів тексту в умовах присутності шуму доцільно проводити навчання нейромережевої структури одразу на основі зашумлених даних.

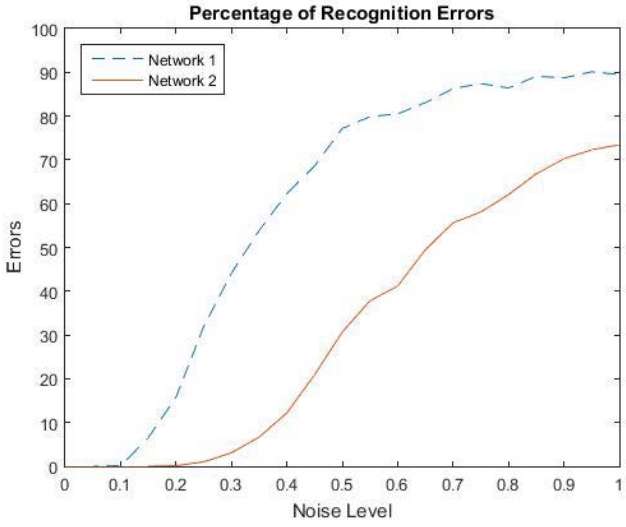


Рис. 5. Порівняння помилок розпізнавання символів першою
та другою перцептронними нейронними мережами

9. Визначення кількості наборів зашумлених даних, необхідних для навчання перцептронної нейромережі.

Для визначення необхідної кількості копій кожної літери необхідно виконати кроки 5–8 навчання другої мережі на різній кількості наборів вхідних зашумлених даних. Фрагмент результатів моделювання перцептронних нейромереж, навчених на різній кількості наборів вхідних даних, наведені в таблиці 1, де MSE_k − показник середньоквадратичної помилки, розрахований за виразом:

MSE_k=(1/length(percError2))(percError2-MSE_0)^2;

де MSE_0 – еталонне значення помилки, MSE_0 = 0.

Таблиця 1

Результати моделювання роботи нейромереж

Кількість наборів зашумлених вхідних даних, k	**Кількість ітерацій навчання**	**Час навчання, с**	**Помилка валідації, 10^{-3}**	**MSE_k**
10	15	22	9,024	2,57
20	25	61	2,749	1,97
30	28	101	1,345	2,08
40	**22**	**106**	**2,538**	**1,71**
50	32	180	0,962	2,36

Виходячи з отриманих результатів (таблиця 1), для вирішення задачі розпізнавання елементів тексту більш придатною є перцептронна нейронна мережа, для навчання якої використовувалося по 40 копій зашумлених літер. Перевагою цього набору також є невелика кількість ітерацій і прийнятний час навчання порівняно з іншими наборами.

Висновок

У роботі наведено методику синтезу перцептронної структури для задачі розпізнавання елементів тексту, написаного латиницею. Розглянуто механізми генерації вхідних навчальних даних для еталонних і зашумлених літер у векторному форматі, який придатний для подальшої обробки нейронною мережею, показано спосіб їх візуалізації. Отримано двошарову структуру нейронної мережі прямого поширення з модифікованими перцептронами, у яких, замість порогових функцій активації, використані сигмоїдальні (перший, прихований, шар) і лінійні (другий, вихідний шар). Далі виконано навчання двох таких структур перцептронних мереж: а) на еталонних даних і б) на даних, які містять шум. Реалізовано процес тестування навчених мереж на даних із різним рівнем шуму, обчислено відсоток помилок розпізнавання для кожної мережі й побудовано графіки залежності помилки від рівня шуму. Аналіз ефективності роботи двох мереж показав вищу пристосованість (менший відсоток помилок) до розпізнавання зашумлених даних перцептронною мережею, яка навчена на зашумлених даних. Крім того, визначено необхідну кількість зашумлених копій кожної літери для навчання перцептронної нейронної мережі, щоб мінімізувати середньоквадратичну помилку.

Література

1. Chaudhary K., Bali R. Easter 2.0: Improving convolutional models for handwritten text recognition. *Computer Vision and Pattern Recognition*. 2022. 12 p.

2. Qiao L., Tang S., Cheng Z., Xu Y., Niu Y., Pu S., Wu F. Text Perceptron: Towards End-to-End Arbitrary-Shaped Text Spotting. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020. P. 11899–11907.

3. Hassen Mohammed H., Malik J., Al-Madeed S., Kiranyaz S. 2D Self-Organized ONN Model For Handwritten Text Recognition. *Computer Vision and Pattern Recognition*. 2022. 12 p.

4. Coquenet D., Chatelain C., Paquet T. DAN: A Segmentation-free Document Attention Network for Handwritten Document Recognition. *Computer Vision and Pattern Recognition*. 2022. 13 p.

5. Mas-Candela E., Calvo-Zaragoza J. Exploring recursive neural networks for compact handwritten text recognition models. *International Journal on Document Analysis and Recognition*. 2024. № 27(3). P. 213–223.

6. Матвійчук А. М. Аналіз архітектур нейронних мереж для розпізнавання рукописних цифр. *Телекомунікаційні та інформаційні технології.* 2023. № 4. С. 118–127. DOI: 10.31673/2412-4338.2023.041827.

7. Гороховатський О. В., Тесленко О. В. Розпізнавання зображень літер із використанням проекцій та багатошарового персептрону. *Системи обробки інформації.* 2017. Вип. 2. С. 163–167.

8. Jeng J.-H., Hsieh J.-G. Pathways to Machine Learning and Soft Computing. EHGBooks, 2018. 397 p.

9. Khan G. M. Evolution of Artificial Neural Development: In Search of Learning Genes. Springer, New York, 2018. 142 p.

10. McKeon R. Neural Networks for Electronics Hobbyists. Apress, 2018. 139 p.

11. Руденко О. Г., Бодянський Є. В. Штучні нейронні мережі Н навчальний посібник для студентів вищих навчальних закладів. К иїв : CMIT, 2006. 404 с.

12. Chapter 1: Using neural nets to recognize handwritten digits. Electronic resource. http://neuralnetworksanddeeplearning.com/chap1.html.

13. Chollet F. Deep Learning with Python. 2018. 384 p.

14. Коробко О. В. Комп'ютеризовані системи штучного інтелекту : методичні вказівки до лабораторних робіт. Миколаїв : НУК, 2014. 76 с.

References

1. Chaudhary K., Bali R. Easter2.0: Improving convolutional models for handwritten text recognition. *Computer Vision and Pattern Recognition*, 2022. 12 p.

2. Qiao L., Tang S., Cheng Z., Xu Y., Niu Y., Pu S., Wu F. Text Perceptron: Towards End-to-End Arbitrary-Shaped Text Spotting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. 11899–11907

3. Hassen Mohammed H., Malik J., Al-Madeed S., Kiranyaz S. 2D Self-Organized ONN Model For Handwritten Text Recognition. *Computer Vision and Pattern Recognition*, 2022. 12 p.

4. Coquenet D., Chatelain C., Paquet T. DAN: A Segmentation-free Document Attention Network for Handwritten Document Recognition. *Computer Vision and Pattern Recognition*, 2022. 13 p.

5. Mas-Candela E., Calvo-Zaragoza J. Exploring recursive neural networks for compact handwritten text recognition models. *International Journal on Document Analysis and Recognition*, 2024. 27(3): 213–223.

6. Matviichuk A. M. Analiz arkhitektur neironnykh merezh dlia rozpiznavannia rukopysnykh tsyfr. *Telecommunications and Information Technologies*, 2023. № 4, 118–127. DOI: 10.31673/2412-4338.2023.041827 [in Ukrainian].

7. Gorokhovatskyi O. V., Teslenko O. V. Recognition of letter images using projections and a multilayer perceptron. Information Processing Systems, 2017. Issue 2. 163–167 [in Ukrainian].

8. Jeng J.-H., Hsieh J.-G. Pathways to Machine Learning and Soft Computing. EHGBooks, 2018. 397 p.

9. Khan G. M. Evolution of Artificial Neural Development: In Search of Learning Genes. Springer, New York, 2018. 142 p.

10. McKeon R. Neural Networks for Electronics Hobbyists. Apress, 2018. 139 p.

11. Rudenko O. G., Bodiansky E. V. Artificial neural networks Textbook for students of higher educational institutions K.: SMIT, 2006, 404 p. [in Ukrainian].

12. Chapter 1: Using neural nets to recognize handwritten digits. Electronic resource. http://neuralnetworksanddeeplearning.com/chap1.html.

13. Chollet F. Deep Learning with Python. 2018. 384 p.

14. Korobko O. V. Kompiuteryzovani systemy shtuchnoho intelektu: Metodychni vkazivky do laboratornykh robit. Mykolaiv: NUK. 2014. 76 p. [in Ukrainian].

Дата першого надходження рукопису до видання: 16.05.2025
Дата прийнятого до друку рукопису після рецензування: 18.06.2025
Дата публікації: 25.07.2025

НАУКОВЕ ВИДАННЯ

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 1 (09) 2025

Підписано до друку: 25.07.2025.
Формат 60x84/8. Arial.
Папір офсет. Цифровий друк. Ум. друк. арк. 17,91. Замов. № 0925/745. Наклад 300 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, Україна, м. Одеса, вул. Інглезі, 6/1
Телефон +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб'єкта видавничої справи
ДК № 7623 від 22.06.2022 р.