



ISSN 2788-5518

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

№ 2 (10), 2025

НАУКОВИЙ ЖУРНАЛ



Видавничий дім
«Гельветика»
2025



ISSN 2788-5518

INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 2 (10), 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 Copernicus, Vernadsky National Library of Ukraine.

Мова видання: українська, англійська

Рекомендовано до друку вченою радою Відкритого міжнародного університету розвитку людини «Україна» (протокол № 4 від 23.12.2025 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.

Редколегія

Головний редактор: Таланчук Петро Михайлович, д.т.н., проф., університет «Україна», ORCID 0000-0001-8748-6127, Київ, Україна. talanshuk_pm@ukr.net

Заступник головного редактора: Забара Станіслав Сергійович, д.т.н., проф., університет «Україна», ORCID 0000-0001-9554-7575, Київ, Україна. staszabara37@gmail.com

Відповідальний секретар: Павленко Володимир Іванович, к.ф.-м.н., доц., університет «Україна», ORCID 0000-0002-3958-0415, Київ, Україна. pavlenko.v@i.ua

Члени редколегії

Безрук Валерій Михайлович, д.т.н., проф., Харківський національний університет радіоелектроніки, Харків, Україна, ORCID-0000-0003-2349-7788, valeriy_bezruk@ukr.net

Блаунштейн Натан Шаєвич, д.ф.-м.н., проф., Університет Бен-Гуріона, Біг-Шева, Ізраїль, ORCID-0000-0003-2945-9379, nathan.blaunstein@hotmail.com

Бойко Юлій Миколайович, д.т.н., проф., Хмельницький національний університет, Хмельницький, Україна, ORCID-0000-0003-0603-7827, boiko_julius@ukr.net

Бескровний Олексій Іванович, к.т.н., доц., Університет «Україна», Київ, Україна, ORCID-0000-0002-7043-7801, obezkrovnyy@gmail.com

Гепко Ігор Олександрович, д.т.н., проф., Український державний центр радіочастот, Київ, Україна, ORCID-0000-0002-4138-021X, gepko@ucrf.gov.ua

Климаш Михайло Миколайович, д.т.н., проф., Національний університет «Львівська політехніка», Львів, Україна, ORCID-0000-0003-2867-1482, mklimash@polynet.lviv.ua

Козловський Валерій Валерійович, д.т.н., проф., Національний авіаційний університет, Київ, Україна, ORCID-0000-0002-8301-5501, vv_k@nau.edu.ua

Кушнір Микола Ярославович, к.ф.-м.н., доц., Чернівецький національний університет ім. Юрія Федьковича, Чернівці, Україна, ORCID-0000-0001-9480-3856, kushnirnick@gmail.com

Лунтовський Андрій Олегович, д.т.н., проф., Державна академія Саксонії «Беруфсакадемія», Дрезден, Німеччина, ORCID-0000-0001-7038-6955, andriy.luntovskyy@ba-sachsen.de

Мельник Ігор Віталійович, д.т.н., проф., НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, ORCID-0000-0003-0220-0615, imelnik@phbme.kpi.ua

Наконечний Олександр Григорович, д.ф.-м.н., проф., Київський національний університет ім. Тараса Шевченка, Київ, Україна, ORCID-0000-0002-8705-3070, a.nakonechniy@gmail.com

Наритник Теодор Миколайович, к.т.н., проф., НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, ORCID-0000-0001-8071-6356, t.narytnyk@ukr.net

Петренко Анатолій Іванович, д.т.н., проф., НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, ORCID-0000-0001-6712-7792, tolja.petrenko@gmail.com

Попов Валентин Іванович, д.ф.-м.н., проф., Ризький технічний університет, Рига, Латвія, ORCID-0000-0002-2040-9319, popovs@latnet.lv

Почерняєв Віталій Миколайович, д.т.н., проф., Державний університет інтелектуальних технологій і зв'язку, Одеса, Україна, ORCID-0000-0001-7130-8668, vpochernyaev@gmail.com

Рогоза Валерій Станіславович, д.т.н., проф., НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, ORCID-0000-0003-2327-156X, rosvetnik@gmail.com

Самарай Валерій Петрович, к.т.н., с.н.п., Університет «Україна», Київ, Україна, ORCID-0000-0003-4419-1366, samaraj@ukr.net

Семенко Анатолій Іларіонович, д.т.н., проф., Університет «Україна», Київ, Україна, ORCID-0000-0002-7043-7801, setel@ukr.net

Сидоренко Анатолій Сергійович, д.ф.-м.н., проф., академік Академії наук Молдови, Кишинів, Республіка Молдова, ORCID-0000-0001-7433-4140, anatoli.sidorenko@kit.edu

Стрелковська Ірина Вікторівна, д.т.н., проф., Міжнародний гуманітарний університет, Одеса, Україна, ORCID-0000-0002-1813-0554, irina7000370@gmail.com

Тимошенко Анатолій Григорович, к.т.н., доц., Університет «Україна», Київ, Україна, ORCID-0000-0003-0954-3186, timoshag@i.ua

Ящишін Євген Михайлович, д.т.н., проф., Варшавський технологічний університет, Варшава, Польща, ORCID-0000-0001-8378-1399, e.jaszczyszyn@ire.pw.edu.pl

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 23.12.2025)

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

Editorial board

Chief Editor: Peter Talanchuk, Dr. habil., Prof., University "Ukraine", ORCID 0000-0001-8748-6127, Kyiv, Ukraine. talanshuk_pm@ukr.net

Deputy Chief Editor: Stanislav Zabara, Dr. habil., Prof., University "Ukraine", ORCID 0000-0001-9554-7575, Kyiv, Ukraine. staszabara37@gmail.com

Executive secretary: Vladimir Pavlenko, Ph.D., Ass.Prof. University "Ukraine", ORCID 0000-0002-3958-0415, Kyiv, Ukraine. pavlenko.v@i.ua

Members of the Editorial Board

Valeriy Bezruk, Dr. habil., Prof., Kharkiv national University of Radioelectronics, ORCID: 0000-0003-2349-7788, Kharkiv, Ukraine. valeriy_bezruk@ukr.net

Natan Blaunstein, Dr. habil. Phys., Prof., Ben-Gurion University, ORCID: 0000-0003-2945-9379, Beer Sheva, Israel. nathan.blaunstein@hotmail.com

Julius Boiko, Dr. habil., Prof., Khmelnytsky national University, ORCID: 0000-0003-0603-7827, Khmelnytskyi, Ukraine. boiko_julius@ukr.net

Alexey Beskrovnyy, Ph.D., Ass.Prof., University "Ukraine", ORCID-0000-0002-7043-7801, Kyiv, Ukraine. obezkrovnyy@gmail.com

Igor Gepko, Dr. habil., Prof., Ukrainian State Center of Radio Frequencies, ORCID-0000-0002-4138-021X, Kyiv, Ukraine. gepko@ucrf.gov.ua

Mykhailo Klymash, Dr. habil., Prof., Lviv polytechnic national University, ORCID: 0000-0003-2867-1482, Lviv, Ukraine. mklmash@polynet.lviv.ua

Valeriy Kozlovsky, Dr. habil., Prof., National Aviation University, ORCID: 0000-0002-8301-5501, Kyiv, Ukraine. vv_k@nau.edu.ua

Mykola Kushnir, Ph.D., Phys., Ass.Prof., Yuriy Fedkovych Chernivtsi national University, ORCID: 0000-0001-9480-3856, Chernivtsi, Ukraine. kushnirnick@gmail.com

Andriy Luntovsky, Dr. habil., Prof., Saxon Study Academy "Berufsakademie", ORCID: 0000-0001-7038-6955, Dresden, Germany. andriy.luntovskyy@ba-sachsen.de

Igor Melnyk, Dr. habil., Prof., National technical University of Ukraine "Igor Sikorsky Kyiv polytechnic Institute", ORCID: 0000-0003-0220-0615, Kyiv, Ukraine. imelnik@phbme.kpi.ua

Alexander Nakonechny, Dr. habil. Phys., Prof., Taras Shevchenko national University of Kyiv, ORCID: 0000-0002-8705-3070, Kyiv, Ukraine. a.nakonechniy@gmail.com

Teodor Narytnyk, Ph.D. Prof. National technical University of Ukraine "Igor Sikorsky Kyiv polytechnic Institute", ORCID: 0000-0001-8071-6356, Kyiv, Ukraine. t.narytnyk@ukr.net

Anatoliy Petrenko, Dr. habil., Prof., National technical University of Ukraine "Igor Sikorsky Kyiv polytechnic Institute", ORCID: 0000-0001-6712-7792, Kyiv, Ukraine. tolja.petrenko@gmail.com

Valentin Popov, Dr. habil. Phys., Prof., Riga technical University, ORCID: 0000-0002-2040-9319, Riga, Latvia. popovs@latnet.lv

Vitaliy Pochernyaev, Dr. habil., Prof., University of Intellectual Technologies and Communications, ORCID: 0000-0001-7130-8668, Odesa, Ukraine. vpochernyaev@gmail.com

Valeriy Rogoza, Dr. habil., Prof., National technical University of Ukraine "Igor Sikorsky Kyiv polytechnic Institute", ORCID: 0000-0003-2327-156X, Kyiv, Ukraine. rosvetnik@gmail.com

Valeriy Samaraj, Ph.D., SSR., University "Ukraine", ORCID: 0000-0003-4419-1366, Kyiv, Ukraine. samaraj@ukr.net

Anatoliy Semenکو, Dr. habil., Prof., University "Ukraine", ORCID: 0000-0002-7043-7801, Kyiv, Ukraine. setel@ukr.net

Anatoliy Sidorenko, Dr. habil. Phys., Prof., Academician of AS of Republic of Moldova, ORCID: 0000-0001-7433-4140, Kishinev, Republic of Moldova. anatoli.sidorenko@kit.edu

Irina Strelkovskaya, Dr. habil., Prof., International Humanitarian University, ORCID: 0000-0002-1813-0554, Odessa, Ukraine. irina7000370@gmail.com

Anatoliy Tymoshenko, Ph.D., Ass. Prof., University "Ukraine", ORCID: 0000-0003-0954-3186, Kyiv, Ukraine. tymoshag@i.ua

Eugeniy Yashchyszyn, Dr. habil., Prof., Warsaw University of Technology, ORCID: 0000-0001-8378-1399, Warsaw, Poland. e.jaszczyszyn@ire.pw.edu.pl

ЗМІСТ

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

Гнатюк В. О., Литвинюк О. М. АНАЛІЗ СУЧАСНИХ ПІДХОДІВ ДО УПРАВЛІННЯ РЕСУРСАМИ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ ІЗ ВИКОРИСТАННЯМ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ	9
Почерняєв В. М., Сивкова Н. М. КОРЕКТОР ПОЛЯРИЗАЦІЇ АНТЕННО-ФІДЕРНОГО ТРАКТУ МОБІЛЬНОЇ ЦИФРОВОЇ ТРОПОСФЕРНОЇ СТАНЦІЇ	15
Fant M. O., Vakaliuk T. A. ARCHITECTURE AND IMPLEMENTATION OF A MODEL CONTEXT PROTOCOL-BASED DOCUMENTATION ASSISTANT: A CASE STUDY WITH AI-POWERED CONTEXT GENERATION	23

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

Бабич С. М., Ляшук Т. Г. ЕВОЛЮЦІЯ АРХІТЕКТУР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ОБЧИСЛЮВАЛЬНИХ СИСТЕМ: ВІД МОНОЛІТНИХ ДО МІКРОСЕРВІСНИХ	30
Безрук В. М., Голобородько Ю. М., Заболотний В. І. ЗАХИСТ ВІД АКТИВНИХ ЗАВАД БЕЗДРОВОТНИХ СИСТЕМ ЗВ'ЯЗКУ З АДАПТИВНИМИ ФАЗОВАНИМИ АНТЕННИМИ РЕШІТКАМИ	37
Ващіліна О. В., Голотюк В. М. ПРОЄКТУВАННЯ ПРОГРАМНОЇ СИСТЕМИ ПЛАНУВАННЯ ЧАСУ НА ОСНОВІ КРОСПЛАТФОРМНИХ ТЕХНОЛОГІЙ	43
Гнатюк В. О., Зандер К. Ю. МЕТОД ДИНАМІЧНОГО БАЛАНСУВАННЯ ПРОДУКТИВНОСТІ СЕРВЕРІВ У ІНТЕРАКТИВНИХ СИСТЕМАХ	51
Дьяконов О., Ушкаренко О., Обрубов А., Нянько С. В., Сірівчук А. ЗАСТОСУВАННЯ АЛГОРИТМІВ ДЕТЕКЦІЇ ОБ'ЄКТІВ У БОРТОВИХ СИСТЕМАХ ВІДЕОАНАЛІЗУ БПЛА З ВИКОРИСТАННЯМ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ	65
Забара С. С., Дуднік А. С., Ізварін І. В., Яремак В. І., Веркалець І. Д. ОБРОБКА ВЕЛИКИХ ДАНИХ ТРАНЗАКЦІЙ БАНКІВСЬКОЇ СИСТЕМИ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ	75
Завгородня Г. А., Завгородній В. В. АВТОМАТИЗАЦІЯ ТЕСТУВАННЯ ІГРОВИХ МЕХАНІК ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ	83
Кулик Ю. Р., Батюк А. Є. РОЗРОБЛЕННЯ ТА АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ І ОБЧИСЛЮВАЛЬНИХ МЕТОДІВ ДЛЯ СИМУЛЯЦІЇ ПРОЦЕСІВ У ВІРТУАЛЬНІЙ РЕАЛЬНОСТІ	89
Medvinskyi S. V. THE USE OF CROSS-CORRELATION AS AN INTERACTION TOOL FOR COMPUTER SYSTEMS BY INDIVIDUALS WITH MUSCULOSKELETAL DISORDERS	98
Петросян А. Р. ОГЛЯД СУЧАСНИХ МЕТОДІВ ОБРОБКИ СЕНСОРНОЇ ІНФОРМАЦІЇ ТА ПРОГРАМНИХ АРХІТЕКТУР БОРТОВИХ СИСТЕМ УПРАВЛІННЯ БЕЗПІЛОТНИХ ПОВІТРЯНИХ СУДЕН	105
Полуектова Н. Р., Литвиненко Т. М. МОДЕЛЮВАННЯ ГЛИБИННИХ ПІДКРІПЛЮВАЛЬНИХ МЕТОДІВ ДЛЯ АДАПТИВНОЇ МАРШРУТИЗАЦІЇ В КОМП'ЮТЕРНІЙ SDN МЕРЕЖІ З ВИКОРИСТАННЯМ PYTHON	113
Свінцицька О. М., Коротун О. В., Марчук Д. К., Желізо В. В. АЛГОРИТМИ АВТОМАТИЗОВАНОЇ ПЕРЕВІРКИ ПАРАМЕТРІВ АКАДЕМІЧНИХ ДОКУМЕНТІВ	119
Стисло Т. Р. АНАЛІТИКА ОСВІТНІХ ДАНИХ ЯК ІНСТРУМЕНТ ВИЗНАЧЕННЯ ГЕНДЕРНИХ ДИСБАЛАНСІВ У ЦИФРОВІЙ ОСВІТІ	128
Фуріхата Д. В., Вакалюк Т. А. КЛАСИЧНІ ТА ГІБРИДНІ МЕТОДИ ПРОГНОЗУВАННЯ ДАНИХ ЛІЧИЛЬНИКІВ ЗІ СПОЖИВАННЯ РЕСУРСІВ	135
Шевченко С. С. АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ВЕРИФІКАЦІЇ ТА ВИКОНАННЯ КООПЕРАТИВНИХ СЦЕНАРІЇВ У РОЗПОДІЛЕНИХ ТРЕНАЖЕРНИХ СИСТЕМАХ	142

CONTENTS

INFOCOMMUNICATION TECHNOLOGIES

Hnatiuk V. O., Lytvyniuk O. M. ANALYSIS OF MODERN APPROACHES TO RESOURCE MANAGEMENT IN TELECOMMUNICATION SYSTEMS USING ARTIFICIAL INTELLIGENCE METHODS.....	9
Pochernyaev V. M., Syvkova N. M. POLARIZATION CORRECTOR OF THE ANTENNA-FEEDER PATH OF A MOBILE DIGITAL TROPOSCATTER STATION	15
Fant M. O., Vakaliuk T. A. ARCHITECTURE AND IMPLEMENTATION OF A MODEL CONTEXT PROTOCOL-BASED DOCUMENTATION ASSISTANT: A CASE STUDY WITH AI-POWERED CONTEXT GENERATION	23

COMPUTER TECHNOLOGIES

Babych S. M., Liashuk T. G. THE EVOLUTION OF COMPUTING SYSTEM SOFTWARE ARCHITECTURES: FROM MONOLITHIC TO MICROSERVICES	30
Bezruk V. M., Holoborodko Yu. M., Zabolotnyi V. I. PROTECTION AGAINST ACTIVE INTERFERENCE OF WIRELESS COMMUNICATION SYSTEMS USING ADAPTIVE PHASED ARRAY ANTENNAS	37
Vashchilina O. V., Holotiuk V. M. DESIGN OF TIME MANAGEMENT SOFTWARE SYSTEM BASED ON CROSS-PLATFORM TECHNOLOGIES	43
Gnatyuk V. O., Zander K. Yu. DYNAMIC SERVER PERFORMANCE BALANCING METHOD FOR INTERACTIVE SYSTEMS	51
Dyakonov O. S., Ushkarenko O. O., Obrubov A. V., Nianko S. V., Sirivchuk A. A. APPLICATION OF OBJECT DETECTION ALGORITHMS IN UAV ONBOARD VIDEO ANALYSIS SYSTEMS USING DEEP LEARNING MODELS	65
Zabara S. S., Dudnic A. S., Izvarin I. V., Yaremak V. I., Verkalets I. D. PROCESSING BIG DATA OF BANKING SYSTEM TRANSACTIONS USING MACHINE LEARNING	75
Zavhorodnia G. A., Zavhorodnii V. V. AUTOMATION OF GAME MECHANICS TESTING USING ARTIFICIAL INTELLIGENCE	83
Kulyk Yu. R., Batiuk A. Ye. DEVELOPMENT AND ANALYSIS OF MATHEMATICAL MODELS AND COMPUTATIONAL METHODS FOR SIMULATING PROCESSES IN VIRTUAL REALITY	89
Medvinskyi S. V. THE USE OF CROSS-CORRELATION AS AN INTERACTION TOOL FOR COMPUTER SYSTEMS BY INDIVIDUALS WITH MUSCULOSKELETAL DISORDERS	98
Petrosian A. R. A REVIEW OF MODERN METHODS FOR SENSOR DATA PROCESSING AND SOFTWARE ARCHITECTURES FOR ONBOARD CONTROL SYSTEMS OF UNMANNED AERIAL VEHICLES	105
Poluektova N. R., Lytvynenko T. M. MODELING OF DEEP REINFORCEMENT LEARNING METHODS FOR ADAPTIVE ROUTING IN A COMPUTER SDN NETWORK USING PYTHON	113
Svintsytska O. M., Korotun O. V., Marchuk D. K., Zhelizko V. V. ALGORITHMS FOR AUTOMATED VERIFICATION OF ACADEMIC DOCUMENT PARAMETERS	119
Styslo T. R. LEARNING ANALYTICS AS A TOOL FOR IDENTIFYING GENDER IMBALANCES IN DIGITAL EDUCATION	128
Furikhata D. V., Vakaliuk T. A. CLASSICAL AND HYBRID METHODS FOR FORECASTING RESOURCE CONSUMPTION METER DATA	135
Shevchenko S. S. ALGORITHMIC SUPPORT FOR VERIFICATION AND EXECUTION OF COOPERATIVE SCENARIOS IN DISTRIBUTED TRAINING SYSTEMS	142

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.8:621.39

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-01>

АНАЛІЗ СУЧАСНИХ ПІДХОДІВ ДО УПРАВЛІННЯ РЕСУРСАМИ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ ІЗ ВИКОРИСТАННЯМ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ

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

Литвинюк О. М., PhD аспірант, Державний університет «Київський авіаційний інститут», Київ, Україна. <https://orcid.org/0009-0002-2221-8505>, 6163394@stud.kai.edu.ua

Анотація. Швидка еволюція телекомунікаційних технологій призвела до збільшення складності мережі, її неоднорідності та динаміки трафіку даних. Традиційні підходи до управління ресурсами, засновані на статичних або евристичних методах, більше не можуть забезпечити адаптивність, масштабованість та якість обслуговування (QoS), необхідні сучасним додаткам. Тому штучний інтелект (ШІ) та машинне навчання (МН) стали ключовими факторами інтелектуальної автоматизації в телекомунікаційних системах, дозволяючи приймати рішення в режимі реального часу та самооптимізувати ресурси. У цій статті аналізуються сучасні підходи до управління ресурсами в телекомунікаційних системах на основі ШІ та систематизуються дослідження щодо використання контрольованого, неконтрольованого, підкріплюваного та федеративного навчання для оптимізації радіо-, обчислювальних та енергетичних ресурсів у мережах наступного покоління. Аналіз зосереджений на таких фреймворках, як відкриті мережі радіодоступу (O-RAN), багатоканальні периферійні обчислення (MEC) та інтегровані мережі «супутник-повітря-земля» (SAGIN), де інтелектуальна оркестрація забезпечує продуктивність, енергоефективність та безперервність обслуговування. Навчання з підкріпленням (RL) та глибоке навчання з підкріпленням (DRL) визначені як найперспективніші парадигми для динамічного розподілу ресурсів та мережевого сегментування, тоді як федеративне навчання (FL) дозволяє проводити децентралізоване навчання моделей зі збереженням конфіденційності. Механізми прогностного керування на основі штучного інтелекту, інтегровані в самоорганізовані мережі (SON), дозволяють проактивно адаптуватися до змін навантаження та мобільності. Ключові проблеми масштабного розгортання ШІ включають обмежені набори даних, обчислювальні витрати, інтерпретованість моделі та обмеження в режимі реального часу. Майбутні дослідження повинні зосередитися на гібридних моделях ШІ, які поєднують аналітичну оптимізацію з адаптацією на основі навчання, та на розробці стандартизованих контрольних показників для систем управління на основі ШІ. Інтеграція ШІ в управління телекомунікаційними ресурсами підвищує гнучкість, операційну ефективність та оптимізацію енергоспоживання, формуючи основу для інтелектуальних мережевих архітектур 5G/6G.

Ключові слова: телекомунікаційні системи, управління ресурсами, штучний інтелект, машинне навчання, QoS, 5G.

ANALYSIS OF MODERN APPROACHES TO RESOURCE MANAGEMENT IN TELECOMMUNICATION SYSTEMS USING ARTIFICIAL INTELLIGENCE METHODS

Viktor Hnatiuk, PhD (Engineering), Associate Professor, State University "Kyiv Aviation Institute", Kyiv, Ukraine; State Scientific and Research Institute of Cybersecurity Technologies, Kyiv, Ukraine. <https://orcid.org/0000-0002-4916-7149>, viktor.hnatiuk@npp.kai.edu.ua

Oleksandr Lytvyniuk, PhD Student, State University "Kyiv Aviation Institute", Kyiv, Ukraine. <https://orcid.org/0009-0002-2221-8505>, 6163394@stud.kai.edu.ua

Abstract. The rapid evolution of telecommunication technologies has increased network complexity, heterogeneity, and data traffic dynamics. Traditional resource management approaches based on static or heuristic methods can no longer ensure the adaptability, scalability, and quality of service (QoS) required by modern applications. Artificial intelligence (AI) and machine learning (ML) have therefore

become key enablers of intelligent automation in telecommunication systems, enabling real-time decision-making and self-optimization of resources. This paper analyzes modern AI-based approaches to resource management in telecommunication systems and systematizes research on the use of supervised, unsupervised, reinforcement, and federated learning for optimizing radio, computing, and energy resources in next-generation networks. The analysis focuses on frameworks such as Open Radio Access Networks (O-RAN), Multi-access Edge Computing (MEC), and Satellite–Air–Ground Integrated Networks (SAGIN), where intelligent orchestration ensures performance, energy efficiency, and service continuity. Reinforcement learning (RL) and deep reinforcement learning (DRL) are identified as the most promising paradigms for dynamic resource allocation and network slicing, while federated learning (FL) enables decentralized model training with privacy preservation. AI-based predictive control mechanisms integrated into self-organizing networks (SONs) allow proactive adaptation to changing loads and mobility. Key challenges to large-scale AI deployment include limited datasets, computational overhead, model interpretability, and real-time constraints. Future research should focus on hybrid AI models that combine analytical optimization with learning-based adaptation and on developing standardized benchmarks for AI-driven management systems. The integration of AI into telecommunication resource management enhances flexibility, operational efficiency, and energy optimization, forming a foundation for intelligent 5G/6G network architectures.

Key words: telecommunications systems, resource management, artificial intelligence, machine learning, QoS, 5G.

Вступ

Сучасні телекомунікаційні системи переживають етап фундаментальної трансформації, зумовлений переходом до мереж п'ятого та майбутніх поколінь (5G/6G), експоненційним зростанням обсягів мобільного трафіку та розвитком парадигми Інтернету речей (IoT). Ці зміни призвели до безпрецедентного ускладнення мережевої архітектури та висунули нові, значно жорсткіші вимоги до управління ресурсами. Ключова проблема полягає в тому, що традиційні (класичні) підходи до управління ресурсами виявляються неспроможними ефективно функціонувати в умовах сучасних гетерогенних та високодинамічних мереж. Класичні методи, що здебільшого базуються на статичних моделях, детермінованих алгоритмах та фіксованих правилах, були розроблені для значно простіших і більш передбачуваних умов. Їхні головні недоліки в контексті нових викликів:

1. Недостатня адаптивність: традиційні алгоритми не здатні в реальному часі реагувати на раптові зміни умов у мережі, такі як миттєві сплески трафіку, зміна місця розташування тисяч абонентів, динамічна поява джерел інтерференції або зміна вимог до якості обслуговування (QoS) для різних додатків.

2. Обчислювальна складність: задачі оптимізації розподілу ресурсів (спектру, потужності, обчислювальних потужностей) у багатовимірному просторі параметрів сучасних мереж (наприклад, Massive MIMO, Network Slicing) стають NP-складними. Спроби їх розв'язання класичними методами призводять або до неприйнятно великих затримок, або до знаходження субоптимальних рішень.

3. Реактивний, а не проактивний характер: класичні системи здебільшого реагують на проблеми (наприклад, перевантаження каналу) вже після їх виникнення. Вони не мають вбудованих механізмів для прогнозування майбутнього стану мережі та запобігання деградації якості обслуговування.

Таким чином, виникає гостре науково-технічне протиріччя між зростаючою складністю та динамічністю телекомунікаційних систем і обмеженими можливостями існуючих методів управління ресурсами. Це протиріччя призводить до неефективного використання дорогоцінного радіочастотного спектра, надлишкового енергоспоживання мережевого обладнання та нездатності гарантувати стабільну якість обслуговування (QoS) та якості сприйняття (QoE) для кінцевих користувачів.

Вирішення цієї проблеми вимагає переходу від статичних, реактивних моделей управління до нової парадигми – інтелектуального, автономного та проактивного управління, яке здатне навчатися, адаптуватися та приймати оптимальні рішення в умовах невизначеності. Саме методи штучного інтелекту, зокрема машинне навчання та навчання з підкріпленням, пропонують інструментарій, здатний подолати вказані обмеження.

Отже, аналіз та систематизація сучасних підходів до управління ресурсами на основі ШІ є важливою науково-прикладною задачею, вирішення якої дозволить підвищити ефективність, автономність та надійність телекомунікаційних систем майбутнього.

Аналіз останніх джерел

У сучасних телекомунікаційних системах управління ресурсами (Resource Management) є ключовим завданням, що визначає ефективність функціонування мереж п'ятого та шостого поколінь. За останні роки дослідники активно впроваджують методи штучного інтелекту (ШІ), зокрема машинного навчання (ML) та глибокого підкріплювального навчання (Deep Reinforcement Learning, DRL), для підвищення рівня автоматизації, адаптивності та енергоефективності мереж [1–5]. Згідно з узагальнюючим оглядом Elias Dritsas і Maria Trigka [1], методи машинного навчання відіграють фундаментальну роль у вирішенні задач оптимізації телекомунікаційних систем. Автори класифікують сучасні

ML-підходи за типами задач – прогнозування трафіку, управління пропускнуою здатністю, динамічне планування ресурсів та адаптивне управління QoS. Огляд також підкреслює важливість інтеграції ML у системи з багаторівневою архітектурою (cloud–edge–device) для мінімізації затримок і підвищення ефективності. В огляді Hayder Faeq Alhashimi та співавт. [2] проведено систематизацію підходів ШІ до управління ресурсами у гетерогенних мережах 6G. Зокрема, розглядаються питання сумісності інтелектуальних рішень із технологіями Network Slicing, Massive MIMO, RIS та edge computing. Автори вказують на переваги використання моделей глибокого навчання для розподілу обчислювальних і спектральних ресурсів, а також звертають увагу на проблеми енергоефективності та потребу у стандартизованих інтерфейсах обміну даними. Дослідження Pivoto та ін. [3], опубліковане в журналі IEEE Communications Surveys & Tutorials, охоплює широкий спектр ML-застосувань для управління ресурсами у бездротових системах. Огляд виділяє тенденцію переходу від централізованого до децентралізованого управління ресурсами на основі ШІ, де ключову роль відіграють агенти, здатні до самонавчання в умовах змінного навантаження. Окрему увагу у літературі приділено застосуванню підкріплювального навчання для управління радіоресурсами в архітектурах O-RAN. Galdino та співавт. [4] демонструють, що алгоритми RL здатні ефективно адаптувати політики розподілу радіоресурсів у реальному часі з урахуванням QoS-параметрів користувачів. Раніше подібні підходи узагальнив Alwaraфу та ін. [5], які показали, що DRL перевершує традиційні евристичні методи у складних динамічних сценаріях мобільних мереж. Робота Qian Chen і співавт. [6] розширює підхід управління ресурсами на середовище SAGIN (Space–Air–Ground Integrated Networks), поєднуючи комунікаційні й обчислювальні ресурси. Автори наголошують на необхідності створення уніфікованих моделей оптимізації, які враховують взаємодію між сегментами мережі. Дослідження Nong і Liang [7] демонструє практичну реалізацію оптимізації розподілу ресурсів у 5G slicing-мережах, де використано вагові експоненційні та логарифмічні функції для покращення QoS. Це підтверджує ефективність математично-інтелектуальних моделей у задачах управління із багатьма критеріями. Для підвищення масштабованості AI-рішень у розподілених середовищах запропоновано підхід Federated Deep Reinforcement Learning (FDRL) [8], який дозволяє навчати моделі на різних вузлах без централізованого збору даних. Це знижує навантаження на канал і забезпечує конфіденційність користувачів. Питання енергоефективності та «зелених» комунікацій детально розглянуте в роботі Alhussien і Gulliver [9], де підкреслено, що інтеграція ШІ в управління ресурсами є ключовим напрямом розвитку екологічно стійких 6G-мереж. Запропоновано підхід, який мінімізує споживання енергії базових станцій при збереженні QoS. Дослідження Vega та ін. [10; 11] демонструє можливість інтеграції AI у систему управління Network Slicing через створення модульної архітектури, що динамічно виділяє ресурси для кожного сервісного вирізу мережі. Це забезпечує високий рівень адаптивності та автоматизації управління.

Підсумовуючи аналіз, можна зробити висновок, що застосування методів ШІ в управлінні ресурсами телекомунікаційних систем формує нову парадигму інтелектуального управління, у якій основну роль відіграють глибоке навчання, розподілене навчання та енергоефективна оптимізація в контексті 6G.

Мета статті – провести комплексний аналіз сучасних підходів і тенденцій у застосуванні методів штучного інтелекту, зокрема машинного та глибокого підкріплювального навчання, для автоматизації управління ресурсами в телекомунікаційних системах нового покоління (5G/6G, O-RAN, MEC). На основі узагальнення результатів останніх наукових досліджень визначити ключові проблеми, виклики та перспективні напрями подальшого розвитку інтелектуальних моделей управління, спрямованих на підвищення ефективності, енергоощадності та якості обслуговування в гетерогенних мережах майбутнього.

Виконання досліджень

Сучасні телекомунікаційні системи нового покоління, зокрема 5G та майбутні 6G мережі, характеризуються високим рівнем динамічності, гетерогенності та зростанням вимог до якості обслуговування (QoS) і досвіду користувача (QoE). Традиційні методи управління ресурсами, що базуються на статичних або евристичних алгоритмах, стають неефективними в умовах швидких змін мережевого середовища, складної взаємодії між користувацькими, обчислювальними та транспортними ресурсами.

Одним із найперспективніших напрямів є використання методів штучного інтелекту (AI) та машинного навчання (ML) для автоматизації процесів управління мережевими ресурсами. Згідно з Dritsas та Trigka [1], алгоритми ML дедалі частіше застосовуються для прогнозування навантажень, адаптивного розподілу пропускнуої здатності та мінімізації енергоспоживання.

Pivoto та ін. [3] зазначають, що методи підкріплювального навчання (RL) виявились найбільш ефективними у сценаріях з динамічними параметрами середовища. У свою чергу, Galdino та ін. [4] продемонстрували успішне застосування RL у відкритій радіомережевій архітектурі (O-RAN), де агент адаптивно управляє радіоресурсами в режимі реального часу.

У таблиці 1 узагальнено основні напрями використання методів AI/ML у сучасних телекомунікаційних системах.

Таблиця 1

Основні напрями застосування методів AI у телекомунікаційних системах

Напрямок застосування	Тип алгоритму AI/ML	Очікуваний ефект	Приклади робіт
Розподіл радіоресурсів	Reinforcement Learning (RL)	Оптимізація пропускної здатності, зниження затримок	[3], [4], [5]
Управління енергоспоживанням	Supervised Learning	Зменшення втрат енергії, баланс навантаження	[2], [9]
Розподіл обчислювальних ресурсів	Deep RL, Federated Learning	Підвищення ефективності MEC-сервісів	[6], [8]
Network Slicing	Cognitive AI, Heuristic ML	Адаптивне виділення зрізів мережі	[7], [10], [11]

На рис. 1 показано узагальнену архітектуру системи AI-керованого управління ресурсами в телекомунікаційній мережі, що складається з рівнів збору даних, інтелектуального аналізу та прийняття рішень.

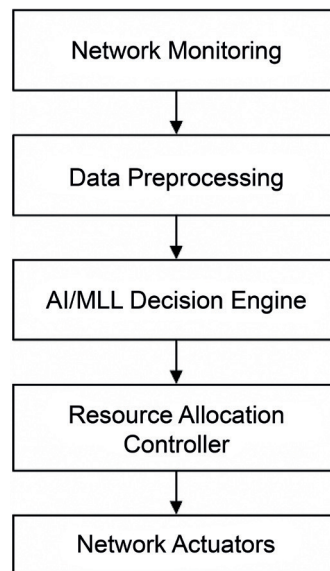


Рис. 1. Узагальнена архітектура AI-керованої системи управління ресурсами

Chen та ін. [6] розширили межі управління ресурсами до інтегрованих SAGIN-систем (Satellite–Air–Ground Integrated Networks), де AI допомагає координувати обчислювальні та комунікаційні ресурси. У таблиці 2 подано порівняння основних методів AI, які використовуються у сфері управління ресурсами.

Таблиця 2

Порівняння методів штучного інтелекту, застосовуваних у управлінні ресурсами

Метод	Тип навчання	Ключові переваги	Недоліки	Типові сфери застосування
Supervised Learning	Контрольоване	Висока точність при наявності навчальних даних	Не працює при нестачі даних	Енергоменеджмент, класифікація навантажень
Unsupervised Learning	Неконтрольоване	Виявлення прихованих закономірностей	Менш точний контроль результатів	Кластеризація користувачів, прогнозування QoE
Reinforcement Learning	Підкріплювальне	Адаптація до середовища, оптимізація дій	Високі обчислювальні витрати	Розподіл радіоресурсів, routing
Federated Learning	Розподілене	Конфіденційність, децентралізація	Складність синхронізації моделей	MEC, IoT
Deep RL	Глибинне підкріплення	Самонавчання, масштабованість	Необхідність потужних обчислень	6G, O-RAN, Network Slicing

Alhussien і Gulliver [9] підкреслюють значення енергоефективності у контексті «зелених» 6G-мереж, де AI використовується для динамічного регулювання потужності передавачів та прогнозування навантаження. Vega та ін. [10] демонструють, що інтеграція AI у керування зрізами мережі (Network Slicing) є ключем до реалізації Self-Organizing Networks (SON), що забезпечують автономну оптимізацію QoS та QoE.

На рис. 2 подано структурну схему роботи інтелектуальної системи управління ресурсами. Система функціонує у вигляді замкнутого циклу, що забезпечує постійне вдосконалення прийняття рішень. Monitoring Module здійснює збір даних про стан мережі, навантаження, QoS та енергоспоживання. Data Processing відповідає за фільтрацію, нормалізацію й попередню обробку інформації, формуючи вхідні параметри для моделі штучного інтелекту.

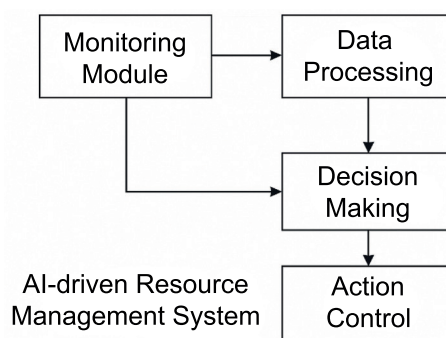


Рис. 2. Узагальнена схема процесу управління ресурсами в телекомунікаційній системі із використанням методів штучного інтелекту

Decision Making реалізує алгоритми машинного навчання або підкріплювального навчання для визначення оптимальних дій щодо розподілу ресурсів. Action Control виконує керувальні дії, спрямовані на адаптацію параметрів мережі (розподіл пропускної здатності, частотного ресурсу, потужності тощо). Зворотний зв'язок між контролером дій і модулем моніторингу забезпечує безперервне самонавчання системи на основі результатів її роботи.

Висновок

У результаті аналізу встановлено, що застосування методів штучного інтелекту в управлінні ресурсами телекомунікаційних систем значно підвищує ефективність і адаптивність мереж наступних поколінь. Проаналізовано сучасні централізовані, децентралізовані та гібридні підходи, показано перехід від класичних алгоритмів оптимізації до інтелектуальних моделей, здатних до самонавчання та прогнозування. Обґрунтовано доцільність використання AI для автоматизації моніторингу, аналізу й прийняття рішень у мережах 5G/6G завдяки можливості оброблення великих обсягів даних у реальному часі. Запропоновано узагальнену структурну модель AI-керованої системи управління ресурсами, що охоплює етапи збору, обробки даних, прийняття рішень і контролю. Модель може слугувати основою для створення прототипів інтелектуальних систем управління в операторських мережах. Перспективними напрямками подальших досліджень є інтеграція глибокого підкріпленого навчання, розвиток цифрових двійників мереж та побудова моделей багаторівневого управління у гетерогенних середовищах. Впровадження таких рішень є ключем до формування автономних телекомунікаційних мереж майбутнього.

Література

1. Dritsas Elias, Trigka Maria. Machine Learning in Information and Communications Technology: A Survey. *Information* 2025. 16 (1). P. 8. DOI: <https://doi.org/10.3390/info16010008>.
2. Alhashimi Hayder Faeq, Hindia Mhd Nour, Dimyati Kaharudin, Hanafi Effariza Binti, Alden Feras Zen, Qamar Faizan, Nguyen Quang Ngoc. Survey on AI-Enabled Resource Management for 6G Heterogeneous Networks: Recent Research, Challenges, and Future Trends. *Computers, Materials and Continua*. Volume 83. Issue 3. 19 May 2025. Pages 3585–3622. DOI: <https://doi.org/10.32604/cmc.2025.062867>.
3. Pivoto D. G. S., de Figueiredo F. A. P., Cavdar C., d. L. Tejerina G. R. and Mendes L. L. A Comprehensive Survey of Machine Learning Applied to Resource Allocation in Wireless Communications. *IEEE Communications Surveys & Tutorials*. DOI: 10.1109/COMST.2025.3552370.
4. Galdino Caio P., Souza Couto Rodrigo, de Medeiros Dianne Scherly Varel. *Reinforcement Learning for Radio Resource Management in O-RAN* : 2024 IEEE 13th International Conference on Cloud Networking (CloudNet). DOI: 10.1109/CloudNet62863.2024.10815702.
5. Alwarafy A., Abdallah M., Ciftler B. S., Al-Fuqaha A., Hamdi M. *Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey* // arXiv preprint arXiv:2106.00574. – 2021.
6. Chen Qian, Guo Zheng, Meng Weixiao, Han Shuai, Li Cheng, Quek Tony Q. S. A Survey on Resource Management in Joint Communication and Computing-Embedded SAGIN. DOI: <https://doi.org/10.48550/arXiv.2403.17400>.

7. Nong X., Liang X. Scheduling allocation in 5G slicing networks utilizing weighted exponential and logarithmic functions to improve QoS. *Sci Rep* 15, 32204 (2025). DOI: <https://doi.org/10.1038/s41598-025-17385-4>.
8. Zang L., Zhang X. and Guo B. Federated Deep Reinforcement Learning for Online Task Offloading and Resource Allocation in WPC-MEC Networks. *IEEE Access*. 2022 Vol. 10. Pp. 9856–9867. DOI: 10.1109/ACCESS.2022.3144415.
9. Alhussien N. and Aaron Gulliver T. Toward AI-Enabled Green 6G Networks: A Resource Management Perspective. *IEEE Access*. 2024. Vol. 12. Pp. 132972–132995. DOI: 10.1109/ACCESS.2024.3460656.
10. Bega D., Gramaglia M., Garcia-Saavedra A., Fiore M., Banchs A. and Costa-Perez X. Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management. *IEEE Communications Magazine*. June 2020. Vol. 58. No. 6. Pp. 32–38. DOI: 10.1109/MCOM.001.1900653.
11. ГНАТЮК В., ГОРБАЧОВ І. (2025). МЕТОД ДИНАМІЧНОГО УПРАВЛІННЯ SLICE-МЕРЕЖАМИ У 5G ДЛЯ ПОКРАЩЕННЯ QOS У VOIP. *Herald of Khmelnytskyi National University. Technical Sciences*. 353 (3.2). 104–109. DOI: <https://doi.org/10.31891/2307-5732-2025-353-12>.

References

1. Dritsas Elias, Trigka Maria. Machine Learning in Information and Communications Technology: A Survey, Information 2025. 16 (1). P. 8. DOI: <https://doi.org/10.3390/info16010008>.
2. Alhashimi Hayder Faeq, Hindia Mhd Nour, Dimyati Kaharudin, Hanafi Effariza Binti, Alden Feras Zen, Qamar Faizan, Nguyen Quang Ngoc. Survey on AI-Enabled Resource Management for 6G Heterogeneous Networks: Recent Research, Challenges, and Future Trends. *Computers, Materials and Continua*. Volume 83. Issue 3. 19 May 2025, Pages 3585–3622. DOI: <https://doi.org/10.32604/cmc.2025.062867>.
3. Pivoto D. G. S., de Figueiredo F. A. P., Cavdar C., d. L. Tejerina G. R. and Mendes L. L. A Comprehensive Survey of Machine Learning Applied to Resource Allocation in Wireless Communications. *IEEE Communications Surveys & Tutorials*. DOI: 10.1109/COMST.2025.3552370.
4. Galdino Caio P., Souza Couto Rodrigo, de Medeiros Dianne Scherly Varel. *Reinforcement Learning for Radio Resource Management in O-RAN* : 2024 IEEE 13th International Conference on Cloud Networking (CloudNet). DOI: 10.1109/CloudNet62863.2024.10815702.
5. Alwarafy A., Abdallah M., Ciftler B. S., Al-Fuqaha A., Hamdi M. Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey. 2021. arXiv preprint arXiv:2106.00574.
6. Chen Qian, Guo Zheng, Meng Weixiao, Han Shuai, Li Cheng, Quek Tony Q. S. A Survey on Resource Management in Joint Communication and Computing-Embedded SAGIN. DOI: <https://doi.org/10.48550/arXiv.2403.17400>.
7. Nong X., Liang X. Scheduling allocation in 5G slicing networks utilizing weighted exponential and logarithmic functions to improve QoS. *Sci Rep* 15, 32204 (2025). DOI: <https://doi.org/10.1038/s41598-025-17385-4>.
8. Zang L., Zhang X. and Guo B. Federated Deep Reinforcement Learning for Online Task Offloading and Resource Allocation in WPC-MEC Networks. *IEEE Access*. 2022. Vol. 10. Pp. 9856–9867. DOI: 10.1109/ACCESS.2022.3144415.
9. Alhussien N. and Aaron Gulliver T. Toward AI-Enabled Green 6G Networks: A Resource Management Perspective. *IEEE Access*. 2024. Vol. 12. Pp. 132972–132995. DOI: 10.1109/ACCESS.2024.3460656.
10. Bega D., Gramaglia M., Garcia-Saavedra A., Fiore M., Banchs A. and Costa-Perez X. Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management. *IEEE Communications Magazine*. June 2020. Vol. 58. No. 6. Pp. 32–38. DOI: 10.1109/MCOM.001.1900653.
11. Gnatyuk V., Gorbachov I. (2025). DYNAMIC SLICE NETWORK MANAGEMENT METHOD IN 5G FOR QOS ENHANCEMENT IN VOIP. *Herald of Khmelnytskyi National University. Technical Sciences*. 353 (3.2). 104–109. DOI: <https://doi.org/10.31891/2307-5732-2025-353-12>.

Дата першого надходження статті до видання: 23.10.2025
 Дата прийняття статті до друку після рецензування: 24.11.2025
 Дата публікації (оприлюднення) статті: 26.12.2025

УДК 621.396.67

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-02>

КОРЕКТОР ПОЛЯРИЗАЦІЇ АНТЕННО-ФІДЕРНОГО ТРАКТУ МОБІЛЬНОЇ ЦИФРОВОЇ ТРОПОСФЕРНОЇ СТАНЦІЇ

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

Анотація. Перспективні мобільні цифрові тропосферні станції мають антенно-фідерні тракту, побудовані на частково заповнених діелектриком хвиеводах. В склад антенно-фідерного тракту входять поляризаційні селектори, призначені для забезпечення роботи мобільних цифрових тропосферних станцій в режимах двох лінійних поляризацій – горизонтальної та вертикальної, що потребує забезпечення відповідної кросполяризаційної розв'язки. Коректор поляризації призначений для корекції паразитного поляризаційного зсуву, що вноситься деполаризаторами. Наведена конструкція поляризаційного селектора на частково заповненому діелектриком квадратному хвиеводі. В даній роботі пропонується пристрій, а саме коректор поляризації на частково заповнених діелектриком квадратному хвиеводі, який забезпечує більш високу кросполяризаційну розв'язку для лінійно-поляризованих хвиль. Цей пристрій коректує положення лінійно-поляризованої хвилі в горизонтальній та вертикальній площинах. Проведено аналіз тематичної літератури. Коректор поляризації представляє собою металізований плоско-поперечний стик з двома ортогональними щілинами з увімкненими в них відкритими нелінійними структурами. В роботі досліджено коректор поляризації у вигляді металізованого плоско-поперечного стику з двома ортогональними щілинами з увімкненими в них відкритими нелінійними структурами. В кожному з ортогональних щілин увімкнено відкриту нелінійну структуру у вигляді наноплівки. Наявність тонких ВНС в щілинах у вигляді наноплівок відповідає рівню нанотехнологій. Регулювання вольтамперних параметрів через відкриті нелінійні структури забезпечує необхідну реактивну провідність стику та модуля коефіцієнта трансформації, а ті в свою чергу необхідну кросполяризаційну розв'язку для лінійно-поляризованих хвиль. Записано коефіцієнт кросполяризаційної розв'язки в термінах *S*-матриці. Приведені власні векторні функції частково заповненого діелектриком квадратного хвиеводу, які виражені через функції Мат'є. Для розрахунків функції Мат'є представлені у вигляді тригонометричних рядів та рядів по функціям Бесселя першого роду. Отримані чисельні результати для коефіцієнта кросполяризаційної розв'язки в залежності від модуля коефіцієнта відображення при різних утриманнях членів рядів функції Мат'є.

Ключові слова: кросполяризаційна розв'язка, лінійно-поляризовані хвилі, відкрита нелінійна структура, наноплівка, частково заповнений діелектриком квадратний хвиевод, модуль коефіцієнта відображення, власні векторні функції, функції Мат'є.

POLARIZATION CORRECTOR OF THE ANTENNA-FEEDER PATH OF A MOBILE DIGITAL TROPOSCATTER STATION

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

Abstract. Promising mobile digital troposcatter stations have antenna-feeder paths built on waveguides partially filled by dielectric. The antenna-feeder path includes polarization selectors designed to ensure the operation of mobile digital troposcatter stations in two linear polarization modes – horizontal and vertical, which ensures the appropriate cross-polarization isolation. A polarization corrector is designed to correct the parasitic polarization shift introduced by depolarizers. The design of a polarization selector based on a square waveguide partially filled by dielectric is presented. In this article proposes a device – a polarization corrector based on a square waveguide partially filled by dielectric that provides higher cross-polarization isolation for linearly polarized waves. The device corrects the position of a linearly polarized wave in the horizontal and vertical planes. A literature review was conducted. The polarization corrector is a metallized flat transverse junction with two orthogonal slits containing open nonlinear structures. The article presents a polarization corrector in the form of a metallized plane-transverse junction with two orthogonal slits containing open nonlinear structures. Each orthogonal

slit contains an open nonlinear structure in the form of a nanofilm. The presence of thin open nonlinear structures in the gaps in the form of nanofilms corresponds to the level of nanotechnology. Adjusting the current-voltage parameters through open nonlinear structures ensures the required junction conductance and transformation ratio modulus, which in turn provides the necessary cross-polarization isolation for linearly polarized waves. The cross-polarization isolation coefficient is written in terms of the S-matrix. The eigenvector functions of a square waveguide partially filled by dielectric are presented, expressed in terms of Mathieu functions. For calculations, the Mathieu functions are presented as trigonometric series and series in terms of Bessel functions of the first kind. Numerical results are obtained for the cross-polarization decoupling coefficient depending on the modulus of the mapping coefficient for different retentions of the terms of the Mathieu function series.

Key words: cross-polarization decoupling, linearly polarized waves, open nonlinear structure, nanofilm, square waveguide partially filled by dielectric, modulus of the mapping coefficient, eigenvector functions, Mathieu functions.

Вступ

Перспективні мобільні цифрові тропосферні станції (МЦТрС) мають антенно-фідерні тракти, побудовані на частково заповнених діелектриком хвильоводах (ЧЗДХ). В склад антенно-фідерного тракту входять поляризаційні селектори, призначені для забезпечення роботи МЦТрС в режимах двох лінійних поляризацій – горизонтальної та вертикальної, що потребує забезпечення відповідної кросполяризаційної розв'язки. Коректор поляризації призначений для корекції паразитного поляризаційного зсуву, що вноситься деполаризаторами. Під деполаризаторами будемо розуміти елементи антенно-фідерного тракту або впливи на трасі розповсюдження сигналу, які призводять до нестійкості площин поляризації.

Запропонований спосіб покращення кросполяризаційної розв'язки для трактів, реалізованих на ЧЗДХ полягає у введенні в тракт після поляризаційного селектора [1] коректора поляризації. Мобільні комбіновані цифрові тропосферно-радіорелейні, тропосферно-космічні, тропосферно-іоносферні станції і мобільна вузлова цифрова тропосферна станція мають антенно-фідерні тракти, що реалізовані на ЧЗДХ [2; 3]. Деякі пристрої та елементи антенно-фідерних трактів МЦТрС та комбінованих цифрових станцій описані в роботах [4–15].

В роботі [1] досліджена конструкція поляризаційного селектора на ЧЗДХ (рис. 1). В даній роботі досліджується пристрій, а саме коректор поляризації на частково заповненому діелектриком квадратному хвильоводі (ЧЗДКХ), який забезпечує більш високу кросполяризаційну розв'язку для лінійно-поляризованих хвиль. Цей пристрій коректує положення лінійно-поляризованої хвилі в горизонтальній та вертикальній площинах.

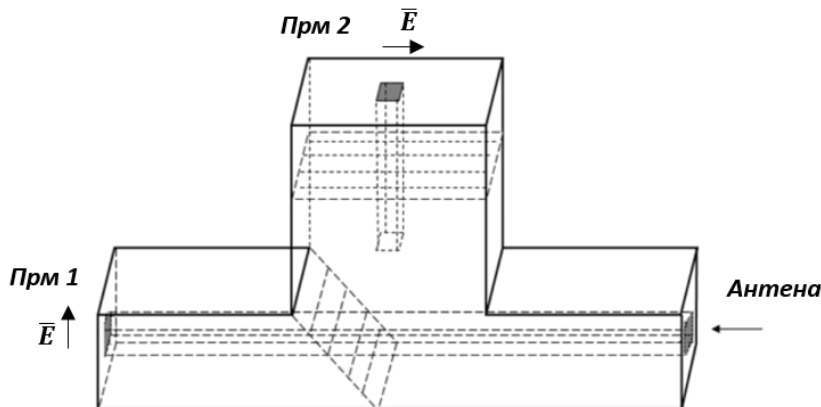


Рис.1. Конструкція поляризаційного селектора на ЧЗДКХ

Аналіз літератури

Аналіз літератури показує, що основні хвильоводні конструкції коректорів поляризації складаються з основного осьового круглого або квадратного хвильоводів та одномодового прямокутного хвильоводу, розташованого під кутом $\pi/2$ до основного осьового хвильоводу. В основному осьовому хвильоводі круглого поперечного перерізу встановлена поздовжня провідна пластина. Ця пластина дозволяє пропускати хвилю однієї лінійної поляризації і відбиває хвилю з іншою лінійною поляризацією в бічний прямокутний хвильовод. Конструкції поляризаційних селекторів, що мають найбільше застосування, показано у довіднику [16]. Також модифікації цих конструкцій досліджено в [17–19; 28; 32], а поляризаційна фільтрація в антенах різних типів і різних діапазонів проаналізовано в [20–27; 29–31; 33–35].

Метою роботи є розробка та дослідження коректора поляризації на ЧЗДКХ.

Основний зміст

Коректор поляризації представляє собою металізований плоско-поперечний стик з двома ортогональними щілинами з увімкненими в них відкритими нелінійними структурами (ВНС). На рис.2 наведено коректор поляризації у вигляді металізованого плоско-поперечного стика з двома ортогональними щілинами з увімкненими в них ВНС. В кожному з ортогональних щілин увімкнені ВНС у вигляді наноплівки. Регулювання вольтаперних параметрів через ВНС забезпечує необхідну jb та $|S_{11}|$, а ті в свою чергу необхідну кросполяризаційну розв'язку для лінійно-поляризованих хвиль.

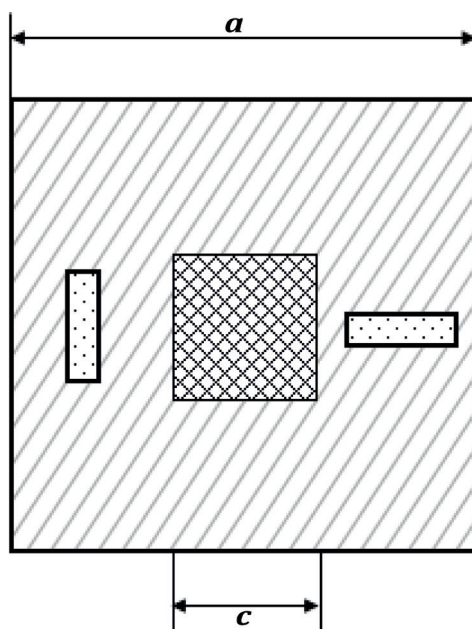


Рис. 2. Коректор поляризації у вигляді металізованого плоско-поперечного стика з двома ортогональними щілинами з увімкненими в них ВНС у вигляді наноплівки

На рис. 3 наведено фрагмент ЧЗДКХ з (увімкненим в нього) коректором поляризації.

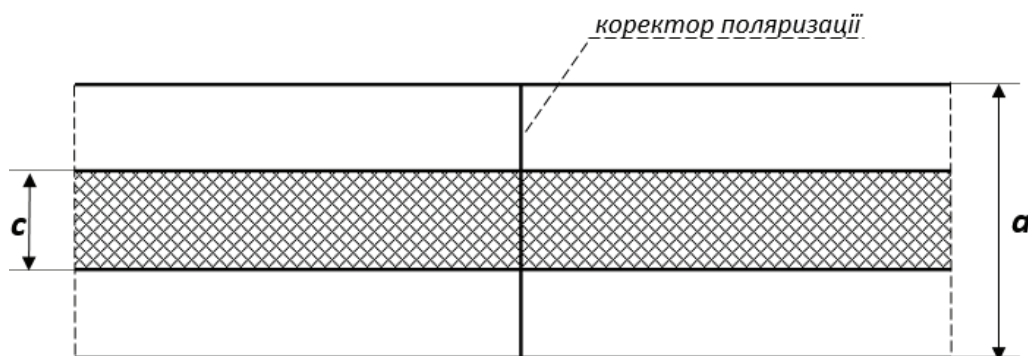


Рис. 3. Фрагмент ЧЗДКХ з (увімкненим в нього) коректором поляризації

Коефіцієнт кросполяризаційної розв'язки в термінах S-матриці визначається наступним чином:

$$X = 20 \lg \frac{|S_{11}| / \sqrt{1 - |S_{11}|^2}}{\sqrt{1 - \frac{|S_{11}|^2}{1 - |S_{11}|^2}}} = 20 \lg \frac{|S_{11}|}{\sqrt{1 - 2|S_{11}|^2}},$$

де $|S_{11}|$ – модуль коефіцієнта відображення від плоско-поперечного стика взаємно перпендикулярних отворів з ВНС.

Конструкція коректора поляризації з ортогональними щілинами еквівалентна ЕН-зчленуванням. Тому:

$$S_{11} = S_{11}^E + S_{11}^H, S_{11}^E = \frac{m_1^2}{2Q_1}, S_{11}^H = -\frac{m_2^2}{2Q_2},$$

$$Q_1 = n^2 + \frac{1}{2}m_1^2 + jb, Q_2 = n^2 + \frac{1}{2}m_2^2 + jb.$$

Такі ЕН-зчленування еквівалентні Е-зчленуванню та Н-зчленуванню подвійного хвильового трийника.

Для Е-зчленування, яке не володіє напрямленими властивостями, в якості функції, апроксимуючої розподіл дотичного електричного поля на отворі, візьмемо власну функцію отвору виду:

$$\bar{\Xi}_1 = \sqrt{\frac{128}{tl}} Ce_1(x, q) \bar{z}^0, \quad (1)$$

де t, l – ширина та довжина прямокутного отвору.

Вираз для коефіцієнту трансформації m_1 [7]:

$$m_1 = \int_S [\bar{n} \bar{\Xi}_1] \bar{\mathbf{S}}_{\perp} \cos \beta_{10} z dS,$$

$$\text{де } \bar{\mathbf{S}}_{\perp} = \sqrt{128 / dl (64 + q^2 + p^2 + q^2 p^2)} * [Ce_1(x, q) Ce_0(y, p) \bar{x}^0 + Ce_1(x, q) Ce_0(y, p) \bar{y}^0]$$

з врахуванням (1) та $\bar{n} = \bar{y}^0$ набуде вигляд:

$$m_1 = N_1 \int_S Ce_0(y, p) Ce_1(x, q) Se_1(x, q) \cos \beta_{10} z dS, \quad (2)$$

де $N_1 = 128 / dl \sqrt{(64 + q^2 + p^2 + q^2 p^2)(64 + q^2)}$, $Ce_0(y, p), Ce_1(x, q)$ – парні функції Мат'є; $Ce_0(y, p), Ce_1(x, q)$ – похідні парні функцій Мат'є; $Se_1(x, q)$ – непарна функція Мат'є; β_{10} – фазова постійна основної хвилі квазі- H_{10} ; q, p – параметри функцій Мат'є [7]; S – площа отвору.

Коефіцієнт трансформації m_1 знайдемо двома способами, в основу яких покладена різниця представлення функцій Мат'є та кількість врахованих членів ряду. Перший спосіб полягає в розкладанні функцій Мат'є в ряд за тригонометричними функціями, другий – в ряд по функціям Бесселя. Збіжність таких рядів досліджена в [36].

При розкладанні функцій Мат'є в ряд по тригонометричним функціям обмежимося членами другого порядку малості. Для цього достатньо враховувати члени ряду з коефіцієнтами q^k, p^k , де $k \leq 2$. Тоді

$$Ce_0(y, p) = 1 - \frac{p}{2} \cos \frac{2\pi}{d} y,$$

$$Ce_1(x, q) = \cos \frac{\pi}{l} x - \frac{q}{8} \cos \frac{3\pi}{l} x + \frac{q^2}{64} \left(\frac{1}{3} \cos \frac{5\pi}{l} x - \cos \frac{3\pi}{l} x \right),$$

$$Se_1(x, q) = \sin \frac{\pi}{l} x - \frac{q}{8} \sin \frac{3\pi}{l} x + \frac{q^2}{64} \left(\sin \frac{3\pi}{l} x + \frac{1}{3} \sin \frac{5\pi}{l} x \right). \quad (3)$$

Після підставлення (3) в (2) інтеграл зводиться до табличного [37].

Функції Мат'є, представлені рядами по функціям Бесселя, мають вигляд:

$$Ce_0(y, p) = A_0 \sum_{k=0}^{\infty} (-1)^k A_{2k} J_{2k} (4\sqrt{2p} \cos y),$$

$$Ce_1(x, q) = (A_1 / 2\sqrt{2q}) \sum_{k=0}^{\infty} (-1)^k A_{2k+1} J_{2k+1} (4\sqrt{2q} \cos x), \quad (4)$$

$$Se_1(x, q) = (B_1 / 2\sqrt{2q}) \operatorname{tg} x \sum_{k=0}^{\infty} (-1)^k (2k+1) B_{2k+1} J_{2k+1} (4\sqrt{2q} \cos x),$$

де A, B – коефіцієнти ряду [36]; J – функції Бесселя першого роду.

Ряди в виразах (4) мають рівномірний збіг [36], тому при розрахунку інтегралу (2) скористаємося результатами роботи [7]. Коефіцієнт трансформації m_1 в цьому випадку виражається через гіпергеометричну функцію Гауса.

Для H -зчленування, який так само не володіє напрямленими властивостями, апроксимуюча функція на отворі:

$$\bar{\Xi}_2 = \sqrt{\frac{128}{dl(64 + p^2)}} Ce_1(y, p) \bar{z}^0, \quad (5)$$

Вираз для коефіцієнту трансформації [7]:

$$m_2 = \left(\frac{k_{10}}{\beta_{10}} \right) \int_S [\bar{n} \bar{\Xi}_2] \bar{S}_z \cos \beta_{10} z dS,$$

$$\text{де } \bar{S}_z = \sqrt{128 / dl(64 + q^2 + p^2 + q^2 p^2)} Ce_1(y, p) Ce_0(x, q) \bar{z}^0,$$

З врахуванням (5) та $\bar{n} = \bar{x}^0$ набуде вигляд:

$$m_2 = N_2 \left(\frac{k_{10}}{\beta_{10}} \right) \int_S Ce_1(y, p) Ce_0(x, q) Se_1(y, p) \cos \beta_{10} z dS, \quad (6)$$

$$\text{де } N_2 = 128 / dl \sqrt{(64 + q^2 + p^2 + q^2 p^2)(64 + p^2)}.$$

На рис.4 показані залежності кросполяризаційної розв'язки від модуля коефіцієнта відображення. Ці залежності побудовані при величинах $c/a = 0,3$; $m_1 = 0,45$; $m_2 = 0,35$. Крива 1 побудована з використанням формули (3), крива 2 – з використанням формули (4) з утриманням двох членів ряду. Хрестиками біля кривої 1 показаний результат, щодо використання формули (4) з утриманням трьох членів ряду. З графіків рис.4 можна зробити висновок, що при $|S_{11}| \rightarrow 0,1$ величина $X \approx 10$ дБ, а при $|S_{11}| \rightarrow 0,01$ величина X вже досягає ≈ 30 дБ. На сучасних лініях радіозв'язку величина $X \geq 30$ дБ враховується достатньою для мобільних радіосистем.

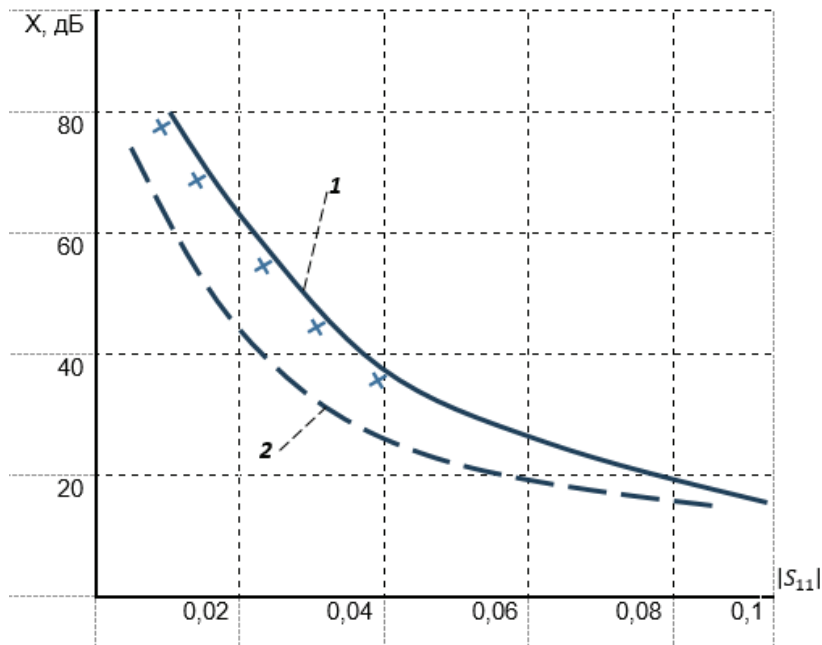


Рис. 4. Коефіцієнт кросполяризаційної розв'язки

Висновок

Новизна конструкції полягає в тому, що коректор поляризації на основі порожнистого круглого хвильоводу замінений на ЧЗДКХ як більш складну конструкцію, але яка дає можливість варіації більшої кількості параметрів. Така заміна основного осьового хвильоводу, яким раніше слугував порожнистий

круглий хвилевід на ЧЗДХ, шляхом підбору геометричних розмірів діелектричної пластини, її відносної діелектричної проникності та вольтамперних характеристик ВНС у вигляді наноплівок дозволяє корегувати поляризаційні характеристики в широкій смузі частот будь-якої радіосистеми НВЧ. При відсутності коректора поляризації потрапляємо в область значень $|S_{11}|$, при якій значення кросполяризаційної розв'язки становить ≈ 20 дБ.

Таким чином, наявність в щілинах ВНС у вигляді наноплівок відповідає рівню нанотехнологій і дозволяє отримувати високі поляризаційні характеристики.

Література

1. Почерняев В. М., Сивкова Н. М., Магомедова М. С. Поляризаційний селектор на частково заповнених діелектриком хвелеводах. *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Środowiska*. 2024. Т.14. № 4. С. 28–31.
2. Почерняев В. М., Сивкова Н. М., Магомедова М. С. Мобільна вузлова цифрова тропосферна станція. *Системи озброєння і військова техніка*. 2024. № 4 (76). С. 6–15.
3. Почерняев В. Н. Повхлеб В. С. Состояние и направления развития мобильных цифровых тропосферных систем связи. *Системи озброєння і військова техніка*. 2018. № 2. С. 51–60.
4. Почерняев В., Магомедова М. Вібаторна антенна решітка для мобільної цифрової іоносферної станції. *Інфокомунікаційні та комп'ютерні технології*. 2022. № 2 (04). С. 50–58. DOI: <https://doi.org/10.36994/2788-5518-2022-02-04-05>.
5. Pochernyaev V., Syvkova N. Broadband switch on partially filled by dielectric rectangular waveguide. *The scientific heritage*. 2021. № 60 (60). VOL 1. С. 49–52. ISSN 9215-0365.
6. Pochernyaev V., Syvkova N., Mahomedova M. Switch-filter on a rectangular waveguide partially filled by dielectric. *Informatyka, Automatyka, Pomiar W Gospodarce I Ochronie Środowiska*. 2022. № 12 (3). P. 8–11.
7. Почерняев В. Н. Устройства на частично заполненных диэлектриком волноводах. Киев : УКНИПСК, 2000. 223 с.
8. Почерняев В. М., Сивкова Н. М., Магомедова М. С. Обмежувач потужності НВЧ на частково заповнених діелектриком прямокутних хвелеводах. *Інфокомунікаційні та комп'ютерні технології*. 2022. № 1 (03). С. 90–101.
9. Почерняев В. М., Сивкова Н. М., Магомедова М. С. Пристрій регулювання потужністю НВЧ на частково заповнених діелектриком прямокутних хвелеводах. *Інфокомунікаційні та інформаційні технології*. 2021. № 2 (02). С. 161–171.
10. Почерняев В. М., Сивкова Н. М. Пристрій управління потужністю НВЧ на частково заповненому діелектриком прямокутному хвелеводі. *Інфокомунікаційні та інформаційні технології*. 2021. № 1 (01). С. 81–89.
11. Pochernyaev V., Mahomedova M., Syvkova N. Комутаційний фазообертач на частково заповненому діелектриком прямокутному хвелеводі. *Системи управління, навігації та зв'язку*. 2023. Т. 1 (71). С. 171–176.
12. Почерняев В. М., Магомедова М. С., Сивкова Н. М. Фазо-частотний пристрій на частково заповненому діелектриком прямокутному хвелеводі. *Системи управління, навігації та зв'язку*. 2022. № 4 (70). С. 158–161.
13. Почерняев В. М., Сивкова Н. М. Перемикач на хвелеводному триїнику, частково заповнений діелектриком. *The scientific heritage*. 2022. VOL 1. № 84 (84). С. 53–57. ISSN 9215-0365.
14. Почерняев В. М., Сивкова Н. М. Фільтр на хвелеводних триїниках, що частково заповнені діелектриком. *The scientific heritage*. 2021. № 75 (75). VOL 1. С. 42–47. ISSN 9215-0365.
15. Pochernyaev V., Syvkova N., Mahomedova M. Microwave mixer on rectangular waveguides partially filled by dielectric. *Informatyka, Automatyka, Pomiar W Gospodarce I Ochronie Środowiska*. 2023. 13 (4). 126–131.
16. Справочник по цифровым радиорелейным системам. Женева : MCE, 1996.
17. Matthias Nickel et al. A liquid crystal based tunable polarization selector in a microwave imaging radiometer. 11th German Microwave Conference – GeMiC. 2018. DOI: <http://doi.org/10.23919/GEMIC.2018.8335036>.
18. Qingsong Jia et al. A Multifunctional Broadband Linear-Circular Polarization Conversion Transmission Metasurface. International Conference on Microwave and Millimeter Wave Technology – ICMMT. 2023. DOI: <http://doi.org/10.1109/ICMMT58241.2023.10277184>.
19. Sergey V. Skovorodnikov et al. Application of Higher-order Wave Modes Filter for Measurement of Phased Antenna Array Elements. Antennas Design and Measurement International Conference – ADMInC. 2021. DOI: <http://doi.org/10.1109/ADMInC54110.2021.9671007>.
20. Sheng Jie Yang et al. Millimeter-Wave Dual-Polarized Filtering Antenna Design for 5G AiP Application. IEEE MTT-S International Microwave Filter Workshop – IMFW. 2021. DOI: <http://doi.org/10.1109/IMFW49589.2021.9642335>.
21. Wenwei Wang et al. Circularly Polarized Patch Antenna With Filtering Performance Using Polarization Isolation and Dispersive Delay Line. *IEEE Antennas and Wireless Propagation Letters*. 2020. 19 (8). Pp. 1457–1461 DOI: <http://doi.org/10.1109/LAWP.2020.3005709>.
22. Tao Xu et al. A Wideband Circularly Polarized Filtering Patch Antenna With Strip Network. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (12). Pp. 2826–2830. DOI: <http://doi.org/10.1109/LAWP.2023.3300168>.
23. Jixing Wang et al. A Wideband Circularly Polarized Filtering Antenna Based on Slot-Patch Structure. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (8). Pp. 1858–1862. DOI: <http://doi.org/10.1109/LAWP.2023.3267053>.
24. Kai Huang et al. Dual-Polarized Filtering Antenna Based on a Second-Order SIR Filter. *IEEE Antennas and Wireless Propagation Letters*. 2024. 23 (1). Pp. 154–158. DOI: <http://doi.org/10.1109/LAWP.2023.3320655>.
25. Pan-Pan Zhang et al. Design of Single-Layer Circularly Polarized Filtering Reflectarray Antenna. IEEE MTT-S International Wireless Symposium – IWS. 2023. DOI: <http://doi.org/10.1109/IWS58240.2023.10222869>.
26. Guangshang Cheng et al. A Stacked Circularly Polarized Filtering Antenna With Crossed Slot. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (12). Pp. 2935–2939. DOI: <http://doi.org/10.1109/LAWP.2023.3305600>.
27. Xinyu Song et al. Planar circularly polarization filter antenna design for the 2.4 GHz Bluetooth band. International Applied Computational Electromagnetics Society Symposium – ACES-China. 2022. DOI: <http://doi.org/10.1109/ACES-China56081.2022.10064915>.
28. Xiuxiang Zhong et al. Circularly Polarized Filtering Antenna Based on Integrated Substrate Gap Waveguide. IEEE MTT-S International Wireless Symposium – IWS. 2021 DOI: <http://doi.org/10.1109/IWS52775.2021.9499709>.

29. Zheng Y. Y. et al. A Compact Full-metal Circularly Polarized Filtering Antenna with an Embedded Metasurface. IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications – IMWS-AMP. 2022. DOI: <http://doi.org/10.1109/IMWS-AMP54652.2022.10106855>.
30. Zhou Jian et al. A Wideband Circularly Polarized Filtering Antenna Based on Stacked Patch Structure. IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications – IMWS-AMP. 2023. DOI: <http://doi.org/10.1109/IMWS-AMP57814.2023.10381421>.
31. Li Hui-Min et al. Compact Broadband Filtering Differential Dielectric Resonator Antenna for Dual-Polarized Applications. *IEEE Antennas and Wireless Propagation Letters*. 2024. 23 (1). DOI: <http://doi.org/10.1109/LAWP.2023.3325411>.
32. He Yejun et al. Dual-Polarized Filtering Antenna Array for 5G Base Station Applications. International Conference on Microwave and Millimeter Wave Technology – ICMMT. 2022. DOI: <http://doi.org/10.1109/ICMMT55580.2022.10022492>.
33. Zhao Weitao et al. A Millimeter-Wave Dual-Polarized Filtering Patch Antenna Array Supported by High-Order SIW Cavity. IEEE 5th International Conference on Electronic Information and Communication Technology – ICEICT. 2022. DOI: <http://doi.org/10.1109/ICEICT55736.2022.9909079>.
34. Yuan Hang et al. A Wideband and High Gain Dual-Polarized Filtering Antenna Based on Multiple Patches. *IEEE Transactions on Antennas and Propagation*. 2022. 70 (10). Pp. 9843–9848. DOI: <http://doi.org/10.1109/TAP.2022.3177494>.
35. Zhai Jian Hong et al. Millimeter-Wave Wideband Circularly Polarized Filtering Antenna for Satellite Communication. 16th UK-Europe-China Workshop on Millimetre Waves and Terahertz Technologies – UCMMT. 2023. DOI: <http://doi.org/10.1109/UCMMT58116.2023.10310621>.
36. McLachlan N. Theory and Application of Mathieu Functions. Oxford, 1947. 402 p.
37. Gradshteyn I. S., Ryzhik I. M. Table of Integrals, Series, and Products. Academic Press, 2007. 1121 p.

References

1. Pocherniaiev V. M., Syvkova N. M., Mahomedova M. S. Poliaryzatsiyni selektor na chastkovo zapovnenykh dielektrykom khvylevodakh. *Informatyka, Avtomatyka, Pomiary w Gospodarce i Ochronie Środowiska*, 2024. № 4. T. 14. S. 28–31.
2. 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.
3. Pocherniaiev V. N. Povkhele V. S. Sostoianye y napravleniya rozvytyia mobyl'nykh tsyfrovyykh troposfernykh system svyazy. *Systemy ozbroiennia i viiskova tekhnika*. 2018. № 2. S. 51–60.
4. Pocherniaiev V., Mahomedova M. Vibratorna antenna reshitka dla mobilnoi tsyfrovoy ionosfernoi stantsii. Infokomunikatsiini ta kompiuterni tekhnologii. 2023. № 2 (04). 50–58. DOI: <https://doi.org/10.36994/2788-5518-2022-02-04-05>.
5. Pochernyaev V., Syvkova N. Broadband switch on partially filled by dielectric rectangular waveguide. *The scientific heritage*. 2021. № 60 (60). VOL 1. P. 49–52. ISSN 9215-0365.
6. Pochernyaev V., Syvkova N., Mahomedova M. Switch-filter on a rectangular waveguide partially filled by dielectric. *Informatyka, Avtomatyka, Pomiary W Gospodarce I Ochronie Środowiska*. 2022. № 12. P. 8–11.
7. Pocherniaiev V. N. Ustroistva na chastychno zapolnennykh dyelektrykom volnovodakh. Kyev : UKNYPSPK, 2000. 223s.
8. Pocherniaiev V. M., Syvkova N. M., Mahomedova M. S. Obmezhuvach potuzhnosti NVCh na chastkovo zapovnenykh dielektrykom priamokutnykh khvylevodakh. *Infokomunikatsiini ta kompiuterni tekhnologii*. 2022. № 1 (03). S. 90–101.
9. Pocherniaiev V. M., Syvkova N. M., Mahomedova M. S. Prystii rehuliuвання potuzhnosti NVCh na chastkovo zapovnenykh dielektrykom priamokutnykh khvylevodakh. *Infokomunikatsiini ta informatsiini tekhnologii*. 2021. № 2 (02). S. 161–171.
10. Pocherniaiev V. M., Syvkova N. M. Prystii upravlinnia potuzhnosti NVCh na chastkovo zapovnenomu dielektrykom priamokutnomu khvylevodi. *Infokomunikatsiini ta informatsiini tekhnologii*. 2021. № 1 (01). S. 81–89.
11. Pochernyaev V., Mahomedova M., Syvkova N. Komutatsiyni fazoobertach na chastkovo zapovnenomu dielektrykom priamokutnomu khvylevodi. *Systemy upravlinnia, navihatsii ta zviazku*. 2023. T. 1 (71). S. 171–176.
12. Pocherniaiev V. M., Mahomedova M. S., Syvkova N. M. Fazo-chastotnyi prystii na chastkovo zapovnenomu dielektrykom priamokutnomu khvylevodi. *Systemy upravlinnia, navihatsii ta zviazku*. 2022. № 4 (70). S. 158–161.
13. Pocherniaiev V. M., Syvkova N. M. Peremykach na khvylevodnomu triinyku, chastkovo zapovnenyi dielektrykom. *The scientific heritage*. 2022. № 84 (84). VOL 1. S. 53–57. ISSN 9215-0365.
14. Pocherniaiev V. M., Syvkova N. M. Filtr na khvylevodnykh triinykakh, shcho chastkovo zapovneni dielektrykom. *The scientific heritage*. 2021. VOL 1. № 75 (75). S. 42–47. ISSN 9215-0365.
15. Pochernyaev V., Syvkova N., Mahomedova M. Microwave mixer on rectangular waveguides partially filled by dielectric. *Informatyka, Avtomatyka, Pomiary W Gospodarce I Ochronie Środowiska*. 2023. 13 (4). 126–131.
16. Spravochnik po tsyfrovym radyoreleinyim systemam. Zheneva : MSE, 1996.
17. Matthias Nickel et al. A liquid crystal based tunable polarization selector in a microwave imaging radiometer. 11th German Microwave Conference – GeMiC. 2018. DOI: <http://doi.org/10.23919/GEMIC.2018.8335036>.
18. Qingsong Jia et al. A Multifunctional Broadband Linear-Circular Polarization Conversion Transmission Metasurface. International Conference on Microwave and Millimeter Wave Technology – ICMMT. 2023. DOI: <http://doi.org/10.1109/ICMMT58241.2023.10277184>.
19. Sergey V. Skovorodnikov et al. Application of Higher-order Wave Modes Filter for Measurement of Phased Antenna Array Elements. Antennas Design and Measurement International Conference – ADMInC. 2021. DOI: <http://doi.org/10.1109/ADMInC54110.2021.9671007>.
20. Sheng Jie Yang et al. Millimeter-Wave Dual-Polarized Filtering Antenna Design for 5G AiP Application. IEEE MTT-S International Microwave Filter Workshop – IMFW. 2021 DOI: <http://doi.org/10.1109/IMFW49589.2021.9642335>.
21. Wenwei Wang et al. Circularly Polarized Patch Antenna With Filtering Performance Using Polarization Isolation and Dispersive Delay Line. *IEEE Antennas and Wireless Propagation Letters*. 2020. 19 (8). Pp. 1457–1461. DOI: <http://doi.org/10.1109/LAWP.2020.3005709>.
22. Tao Xu et al. A Wideband Circularly Polarized Filtering Patch Antenna With Strip Network. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (12). Pp. 2826–2830. DOI: <http://doi.org/10.1109/LAWP.2023.3300168>.

23. Jixing Wang et al. A Wideband Circularly Polarized Filtering Antenna Based on Slot-Patch Structure. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (8). Pp. 1858–1862. DOI: <http://doi.org/10.1109/LAWP.2023.3267053>.
24. Kai Huang et al. Dual-Polarized Filtering Antenna Based on a Second-Order SIR Filter. *IEEE Antennas and Wireless Propagation Letters*. 2024. 23 (1). Pp. 154–158. DOI: <http://doi.org/10.1109/LAWP.2023.3320655>.
25. Pan-Pan Zhang et al. Design of Single-Layer Circularly Polarized Filtering Reflectarray Antenna. *IEEE MTT-S International Wireless Symposium – IWS*. 2023. DOI: <http://doi.org/10.1109/IWS58240.2023.10222869>.
26. Guangshang Cheng et al. A Stacked Circularly Polarized Filtering Antenna With Crossed Slot. *IEEE Antennas and Wireless Propagation Letters*. 2023. 22 (12). Pp. 2935–2939. DOI: <http://doi.org/10.1109/LAWP.2023.3305600>.
27. Xinyu Song et al. Planar circularly polarization filter antenna design for the 2.4 GHz Bluetooth band. *International Applied Computational Electromagnetics Society Symposium – ACES-China*. 2022. DOI: <http://doi.org/10.1109/ACES-China56081.2022.10064915>.
28. Xiuxiang Zhong et al. Circularly Polarized Filtering Antenna Based on Integrated Substrate Gap Waveguide. *IEEE MTT-S International Wireless Symposium – IWS*. 2021. DOI: <http://doi.org/10.1109/IWS52775.2021.9499709>.
29. Zheng Y. Y. et al. A Compact Full-metal Circularly Polarized Filtering Antenna with an Embedded Metasurface. *IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications – IMWS-AMP*. 2022. DOI: <http://doi.org/10.1109/IMWS-AMP54652.2022.10106855>.
30. Zhou Jian et al. A Wideband Circularly Polarized Filtering Antenna Based on Stacked Patch Structure. *IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications – IMWS-AMP*. 2023. DOI: <http://doi.org/10.1109/IMWS-AMP57814.2023.10381421>.
31. Li Hui-Min et al. Compact Broadband Filtering Differential Dielectric Resonator Antenna for Dual-Polarized Applications. *IEEE Antennas and Wireless Propagation Letters*. 2024. 23 (1). DOI: <http://doi.org/10.1109/LAWP.2023.3325411>.
32. He Yejun et al. Dual-Polarized Filtering Antenna Array for 5G Base Station Applications. *International Conference on Microwave and Millimeter Wave Technology – ICMWT*. 2022. DOI: <http://doi.org/10.1109/ICMWT55580.2022.10022492>.
33. Zhao Weitao et al. A Millimeter-Wave Dual-Polarized Filtering Patch Antenna Array Supported by High-Order SIW Cavity. *IEEE 5th International Conference on Electronic Information and Communication Technology – ICEICT*. 2022. DOI: <http://doi.org/10.1109/ICEICT55736.2022.9909079>.
34. Yuan Hang et al. A Wideband and High Gain Dual-Polarized Filtering Antenna Based on Multiple Patches. *IEEE Transactions on Antennas and Propagation*. 2022. 70 (10). Pp. 9843–9848. DOI: <http://doi.org/10.1109/TAP.2022.3177494>.
35. Zhai Jian Hong et al. Millimeter-Wave Wideband Circularly Polarized Filtering Antenna for Satellite Communication. *16th UK-Europe-China Workshop on Millimetre Waves and Terahertz Technologies – UCMWT*. 2023. DOI: <http://doi.org/10.1109/UCMWT58116.2023.10310621>.
36. McLachlan N. *Theory and Application of Mathieu Functions*. Oxford, 1947. 402 p.
37. Gradshteyn I. S., Ryzhik I. M. *Table of Integrals, Series, and Products*. Academic Press, 2007. 1121 p.

Дата першого надходження статті до видання: 24.10.2025

Дата прийняття статті до друку після рецензування: 25.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

UDC 004.8:004.738.5:025.4.03

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-03>

ARCHITECTURE AND IMPLEMENTATION OF A MODEL CONTEXT PROTOCOL-BASED DOCUMENTATION ASSISTANT: A CASE STUDY WITH AI-POWERED CONTEXT GENERATION

Fant M. O., PhD, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine. <https://orcid.org/0000-0002-4994-8009>, fantkolja@gmail.com

Vakaliuk T. A., D.Sc., Prof., Zhytomyr Polytechnic State University, Zhytomyr, Ukraine. <https://orcid.org/0000-0001-6825-4697>, tetianavakaliuk@gmail.com

Abstract. Modern technical documentation systems face increasing complexity and volume, making it difficult for users to obtain accurate, contextually relevant answers without substantial manual navigation. Traditional keyword-based search mechanisms remain limited because they rely on lexical matching rather than understanding user intent or synthesising information spread across multiple documentation sources. As a result, users often receive fragmented results and must piece together separate sections to find complete answers.

This article presents the architecture, design principles, and implementation details of an intelligent documentation assistant based on the Model Context Protocol (MCP). MCP provides a standardised integration mechanism that enables seamless communication between documentation sources and Large Language Models (LLMs). Our proposed system extends existing documentation workflows by incorporating AI-powered context generation capable of automatically analysing, chunking, and structuring raw documentation into MCP-compatible resources. This approach ensures that LLM responses remain grounded in authoritative, up-to-date documentation and reduces the risk of hallucinations.

The architecture consists of four interconnected layers: the documentation website, the MCP server, the AI-powered context generation pipeline, and the LLM integration layer. Each layer is responsible for a distinct set of functions, enabling a clean separation of concerns and improving maintainability. Through a detailed case study, we demonstrate how the system dynamically selects relevant documentation chunks, retrieves context on demand, and produces semantically accurate answers.

Experimental results show that the MCP-based assistant significantly improves response quality compared to traditional search systems, offering better semantic relevance, higher completeness, and substantial reductions in hallucination rates. The standardised MCP ecosystem also enhances interoperability, scalability, and ease of deployment. The findings highlight MCP's potential as a foundational protocol for future intelligent documentation systems.

Key words: model context protocol, large language models, AI architecture, context generation, retrieval-augmented generation.

АРХІТЕКТУРА ТА РЕАЛІЗАЦІЯ АСИСТЕНТА ДОКУМЕНТАЦІЇ НА ОСНОВІ ПРОТОКОЛУ КОНТЕКСТУ МОДЕЛІ: ПРИКЛАД ГЕНЕРАЦІЇ КОНТЕКСТУ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

Микола Фант, к.ф.н., Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0002-4994-8009>, fantkolja@gmail.com

Тетяна Вакалюк, д.п.н., проф., Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0001-6825-4697>, tetianavakaliuk@gmail.com

Анотація. Сучасні системи технічної документації постійно зростають за обсягом і складністю, що ускладнює для користувачів отримання точних та контекстно релевантних відповідей без значних зусиль з навігації. Традиційні механізми пошуку, побудовані на ключових словах, мають суттєві обмеження, оскільки спираються на лексичну відповідність, а не на розуміння намірів користувача чи здатність синтезувати інформацію, розподілену між різними джерелами документації. У результаті користувачі стикаються з фрагментованими відповідями та змушені самотійно об'єднувати окремі частини, щоб отримати повну інформацію.

У статті представлено архітектуру, концептуальні принципи та особливості реалізації інтелектуального асистента документації, побудованого на основі протоколу контексту моделі (MCP). MCP забезпечує стандартизований спосіб інтеграції між джерелами документації та великими мовними моделями. Запропоноване рішення доповнює традиційні підходи за допомогою генерації контексту на основі ШІ, яка автоматично аналізує, сегментує та структурує сирий текст

документації у ресурси, сумісні з MCP. Такий підхід гарантує, що відповіді LLM залишаються точними, заснованими на актуальних даних і менш схильними до галюцинацій.

Архітектура складається з чотирьох взаємопов'язаних рівнів: вебсайту документації, сервера MCP, системи ШІ-генерації контексту та шару інтеграції LLM. Кожен рівень виконує чітко визначені функції, що підвищує масштабованість і спрощує підтримку системи. На основі детального прикладу продемонстровано, як система автоматично визначає релевантні фрагменти документації, отримує необхідний контекст і формує семантично точні відповіді.

Результати дослідження показують, що асистент на основі MCP суттєво підвищує якість відповідей порівняно з традиційним пошуком, забезпечує вищу семантичну релевантність, повноту інформації та значне зменшення рівня галюцинацій. Стандартизована інфраструктура MCP також покращує сумісність, гнучкість і простоту розгортання. Отримані результати підтверджують потенціал MCP як основи для майбутніх інтелектуальних систем документації.

Ключові слова: протокол контексту моделі, велика мовна модель, архітектура штучного інтелекту, генерація контексту, генерація з доповненим пошуком.

Introduction

Technical documentation has become increasingly complex and voluminous as software systems grow in sophistication. Traditional keyword-based search systems suffer from fundamental limitations in understanding user intent and context, requiring users to formulate precise queries using exact terminology found in the documentation [1]. Modern documentation often spans multiple interconnected topics, requiring users to synthesise information from various sources to answer a single question, yet traditional search returns isolated pages or sections, leaving the integration work to the user.

The emergence of LLMs presents a potential solution through their ability to understand natural language queries and synthesise information [2]. However, deploying LLMs for documentation assistance introduces new problems. LLMs are prone to “hallucinations” – generating plausible but incorrect information – particularly when asked about specific technical details not present in their training data [3; 4]. Additionally, LLMs have knowledge cutoff dates, making them unsuitable for answering questions about current software versions or recent features without access to up-to-date documentation [5; 6].

Custom implementations of LLM-documentation integration often result in tightly coupled architectures that are difficult to maintain or migrate. There is a clear need for a standardised protocol that enables seamless integration between documentation resources and LLM systems while ensuring responses are grounded in current, accurate documentation [7]. The MCP addresses this standardisation gap by providing a universal protocol for connecting LLM applications with data sources. MCP defines a precise specification for how context providers can expose their resources, tools, and prompts to LLM clients in a consistent, discoverable manner [8]. This protocol-based approach offers several advantages such as implementation complexity reduction, interoperability between different systems, and clear separation of concerns [9].

The emergence of AI-powered context generation tools, such as KapaAI, further motivates this research by offering automated solutions for extracting and structuring content from documentation. These tools can analyse documentation, identify semantic relationships, create appropriate chunking strategies, and generate structured contexts optimised for LLM consumption. The combination of MCP's standardised protocol and AI-powered context generation creates an opportunity to build documentation assistants that are both powerful and maintainable.

From a practical perspective, organisations invest significant resources in creating and maintaining technical documentation, yet often struggle to ensure users can effectively access this information. An intelligent documentation assistant would substantially improve the return on investment in documentation, reduce the support burden, accelerate developer onboarding, and enhance overall documentation usability.

The primary objective of this research is to design and implement an MCP-based architecture for an intelligent documentation assistant that delivers accurate, contextually relevant answers to user queries.

Results. The proposed intelligent documentation assistant employs a four-layer architecture designed around the MCP as the central integration mechanism. This architecture separates concerns while enabling seamless communication between components, creating a maintainable and scalable system. The architecture follows a user-initiated request-response pattern. When a user submits a natural language query, the request flows through each layer as shown in figure 1.

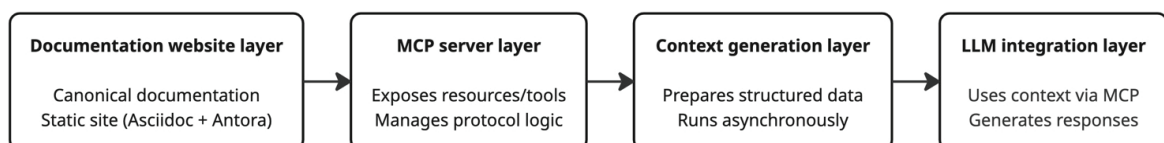


Fig. 1. Four-layer architecture flow

A key architectural principle is the separation of documentation management from intelligence capabilities. The documentation system continues to function as a traditional website, serving human readers through conventional interfaces. The MCP layer adds LLM accessibility without requiring modifications to the core documentation structure, ensuring that improvements to either component can proceed independently.

The component interaction model is built on MCP's client-server paradigm. The LLM system acts as an MCP client, discovering and accessing resources exposed by the MCP server. The server draws from pre-processed contexts generated by the AI-powered context generation system. This model provides flexibility – the context generation process can run asynchronously, updating MCP resources without interrupting active queries.

The Documentation website layer hosts the authoritative documentation content. In our implementation, this layer utilises Netlify as the hosting platform, chosen for its support for static site generation, continuous deployment, and edge network distribution. The documentation is written in AsciiDoc and built with Antora.

The website's structure adheres to standard documentation patterns, featuring clear navigation hierarchies and consistent formatting. Each page includes metadata such as title, description, and tags, which become valuable inputs for context generation. The documentation follows a 'docs-as-code' approach, where updates flow through the same review processes as code changes. The build process transforms asciidoc into static HTML and triggers the context generation pipeline, ensuring MCP resources remain synchronised.

Notably, the documentation layer remains unaware of the MCP infrastructure. No special markup is required in the documentation to support the intelligent assistant, ensuring maintenance can be performed using standard tools and workflows.

The MCP server layer implements the MCP specification, exposing documentation resources and tools to LLM clients. The server translates between the MCP protocol and the underlying documentation system, implementing the full MCP lifecycle: initialisation, resource discovery, resource access, tool invocation, and session management.

The server exposes documentation through MCP resources, which represent addressable pieces of content. Each resource corresponds to a semantically meaningful chunk – a complete page, specific section, or conceptual unit. Resources include metadata such as MIME type, URI, and annotations that help the LLM understand the resource's purpose. Resources are organised hierarchically, mirroring the documentation's logical structure.

The server also implements tools that enable interaction beyond simple retrieval. A search tool allows querying documentation using keywords or phrases, returning ranked resource identifiers. A navigation tool helps understand the documentation structure and locate related content. These tools transform documentation from a passive resource into an active, queryable knowledge base.

The implementation uses Node.js and the official MCP SDK, ensuring protocol compliance. The server maintains an in-memory index for fast discovery and implements caching to minimise redundant processing. For Netlify deployment, the server can be implemented as serverless functions or edge functions. Security mechanisms validate authorised clients, implement rate limiting, and log all access patterns.

The Context generation layer transforms raw documentation into structured, semantically optimised contexts suitable for LLM consumption. This layer employs KapaAI to automate documentation analysis, chunking, and context preparation [10]. The goal is to create contexts that maximise the LLM's ability to provide accurate answers while fitting within context window constraints.

The process begins with the ingestion of documentation, which involves reading files from the build output or source repository. It processes various content types, including markdown, HTML, and code examples, extracting metadata such as titles, headings, and cross-references.

Unlike naive approaches that split at arbitrary boundaries, semantic chunking identifies logical information units that can stand alone. KapaAI uses natural language understanding to identify semantic boundaries, ensuring chunks maintain coherence. A chunk might be a complete concept explanation, a tutorial step, or an API endpoint description with parameters and examples.

The chunking strategy balances competing goals, since chunks must be small enough to fit within context windows yet large enough to contain sufficient information. Overlapping strategies ensure that information spanning boundaries remains accessible and transparent. The system maintains metadata linking related chunks, enabling the LLM to request additional context when needed.

Each chunk is enriched with metadata, including the source documentation page, position within the hierarchy, related topics, code language, and semantic tags. This metadata becomes part of the MCP resource description, allowing informed decisions about which resources to access. The system also creates relationships between chunks, identifying prerequisites, related topics, and examples.

The output is a structured JSON dataset describing all chunks, their content, metadata, relationships, and organisation. This dataset serves as the foundation for the MCP server's resource index. The generation process is incremental, regenerating only chunks affected by documentation updates.

The LLM integration layer connects user queries to the MCP-based documentation system and generates natural language responses based on retrieved context. This layer implements the MCP client protocol, enabling it to discover and access documentation resources. The integration is designed to be LLM-agnostic, supporting various models including Sonnet 4.5, GPT-5, and others.

When a user submits a question, the system analyses the query to understand intent and identify relevant documentation topics. This analysis determines which MCP resources and tools to invoke. Prompt engineering is central to effectiveness – the system constructs prompts including the query, relevant documentation context, and instructions emphasising accuracy and citation. The prompts specify response format based on query characteristics.

Since context window management is critical, the integration layer implements strategies to maximise effective context utilisation, retrieving only the most relevant chunks, prioritising recent or frequently accessed content, and truncating peripheral information when necessary. The system maintains conversation history for multi-turn interactions, enabling coherent conversations where the LLM can reference previous answers.

Response generation includes citation and attribution. When answers are based on specific documentation sections, responses include references or links. The implementation extracts source attribution from MCP resource metadata, formatting citations naturally. Error handling ensures robust operation – if the MCP server is unavailable, the system might fall back to cached contexts or inform the user of reduced capability. Table 1 summarises the implementation details, key technologies, responsibilities, and integration aspects of each major layer in the system architecture.

Table 1
Layers description

Layer	Implementation details	Technologies	Responsibilities	Integration aspect
Documentation website	Built with Antora + Asciidoc; hosted on Netlify	Antora, Asciidoc, Netlify	Hosts canonical content; triggers context updates	Independent of MCP
MCP server	Node.js MCP SDK, serverless deployment	Node.js, MCP SDK	Exposes resources, tools, handles search, and navigation	Translates between docs and LLM
Context generation	Automated semantic chunking with KapaAI	KapaAI, JSON-RPC	Generates structured contexts & metadata	Feeds the MCP resource index
LLM integration	MCP client + prompt engineering	LLM, AI-agent	Handles query analysis, context selection, and response generation	Consumes MCP resources

The technology stack strikes a balance between protocol compatibility, deployment flexibility, performance, and ecosystem maturity. The documentation website layer employs a static site generator (Antora + Asciidoc) that transforms .adoc files into optimised HTML. Netlify serves as the hosting platform, providing global CDN distribution, automatic HTTPS, and serverless function hosting.

The MCP server layer is implemented in Node.js, running on Netlify Edge Functions, using the official MCP SDK from Anthropic. Node.js offers excellent performance for I/O-bound operations and straightforward deployment to serverless platforms. The MCP SDK provides TypeScript types and utilities ensuring protocol compliance. Netlify Edge Functions enable low-latency execution by running serverless code at the network edge, allowing the MCP server to deliver documentation resources closer to end users for improved responsiveness.

For the context generation layer, KapaAI serves as the AI-powered processing engine, providing APIs for documentation ingestion, semantic analysis, and chunking optimised for LLM applications. Integration occurs through its REST API, with data transformation between KapaAI's output and the MCP resource schema.

The implementation of LLM integration layer utilises MCP's official SDK, which includes MCP client capabilities. Development tools include Git, GitHub Actions for CI/CD, and monitoring tools like Grafana.

The server implementation provides a streamlined integration layer that exposes Hazelcast documentation retrieval capabilities to LLM systems through the Model Context Protocol. Deployed as a Netlify Edge Function, this approach leverages edge computing infrastructure to provide low-latency access to documentation context with minimal operational overhead [10].

Rather than implementing the full MCP specification from scratch, the server utilises the official MCP SDK (@modelcontextprotocol/sdk) combined with the Netlify adapter (@modelfetch/netlify) to handle the complexities of JSON-RPC communication and Server-Sent Events (SSE) streaming in edge environments [11]. This architectural decision addresses the unique constraints of edge computing platforms, where traditional Node.js HTTP handling mechanisms are unavailable.

The @modelfetch/netlify adapter implements the streamingHttp protocol, which internally uses StreamableHTTPServerTransport to manage SSE streams in serverless edge contexts. This adapter layer

performs several critical functions: it translates edge fetch Request / Response objects to Node-style HTTP transports expected by the MCP SDK, parses incoming JSON-RPC payloads, routes protocol methods (initialise, tool:discover, tool:invoke) to registered tools, manages SSE streaming for real-time responses, and handles error formatting according to JSON-RPC specifications [12].

Server initialisation loads configuration from environment variables, including the KapaAI API key, project ID, and integration ID. These credentials enable authenticated access to the pre-processed documentation contexts that KapaAI has generated from the Hazelcast documentation site. The server is configured with metadata identifying it as “Hazelcast Docs MCP” with version tracking for compatibility management:

```
const server = new McpServer({
  name: 'Hazelcast Docs MCP',
  version: SERVER_VERSION,
})
```

The server exposes a single tool called “ask_hazelcast_docs” that serves as the primary interface for documentation queries. This tool is registered with comprehensive metadata that helps LLM clients understand when and how to invoke it.

The implementation includes sophisticated request handling logic to distinguish between browser requests and MCP client requests. When browser traffic is detected (through User-Agent headers and Accept headers indicating text / html), the server performs a redirect to the documentation homepage rather than returning raw JSON-RPC responses. This ensures a better user experience when the endpoint is accidentally accessed through a web browser.

For legitimate MCP client requests, the server enforces protocol requirements by patching request headers to include both application/json and text/event-stream in the Accept header, as required by the MCP specification. The server also enforces the POST method for the /mcp endpoint, returning a 405 Method Not Allowed response for other HTTP methods.

The Edge Function configuration includes rate limiting to protect against abuse and prevent unexpected usage spikes that could lead to cost overruns. The rate limit is set to 20 requests per 120-second window, aggregated by both IP address and domain. This balanced approach allows legitimate usage while preventing denial-of-service scenarios and controlling API costs:

```
export const config = {
  path: '/mcp',
  rateLimit: {
    windowLimit: 20,
    windowSize: 120,
    aggregateBy: ["ip", "domain"],
  }
}
```

To facilitate seamless interaction with the MCP server, users can connect using either the MCP Inspector tool or directly from Visual Studio Code. Below are brief guides for each method, with placeholders for corresponding figures.

To initiate a connection from the MCP Inspector, open the tool and navigate to the connection settings panel (see figure 2). Enter the server's endpoint URL, ensuring you use the correct protocol (such as HTTPS) and port. Then, click "Connect" to establish communication. Once connected, you can inspect requests and responses in real time.

Within Visual Studio Code, install the MCP extension if it is not already present. Then, open the command palette and select "MCP: Connect to Server" (see figure 3). Enter the server's endpoint when prompted, and the extension will handle authentication and establish the connection. A successful connection will enable you to interact directly with the server from your code editor.

Implementation and evaluation revealed several key findings about the MCP-based architecture and its effectiveness in providing documentation assistance. The architecture successfully decouples documentation management from AI capabilities, allowing for the independent evolution of each component. Documentation teams can update content without affecting the assistant, and improvements to the LLM integration do not require changes to the documentation.

AI-powered context generation through KapaAI proved highly effective at creating semantically coherent chunks. AI-generated chunks preserved conceptual boundaries better than rule-based approaches, recognising that code examples and explanations form units that should remain together. The impact of the key findings is shown in Table 2.

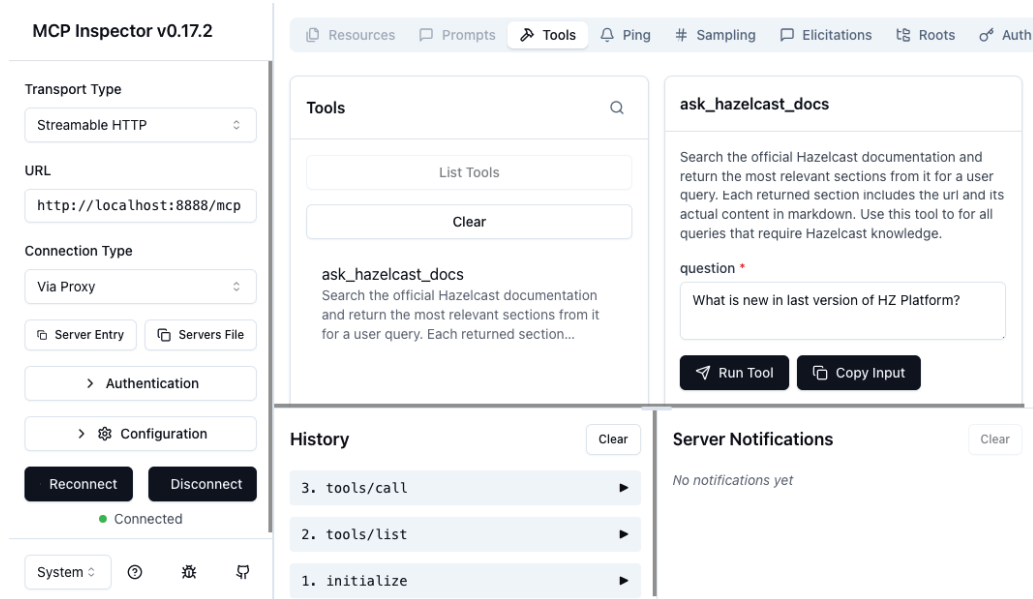


Fig. 2. Connecting from the MCP Inspector

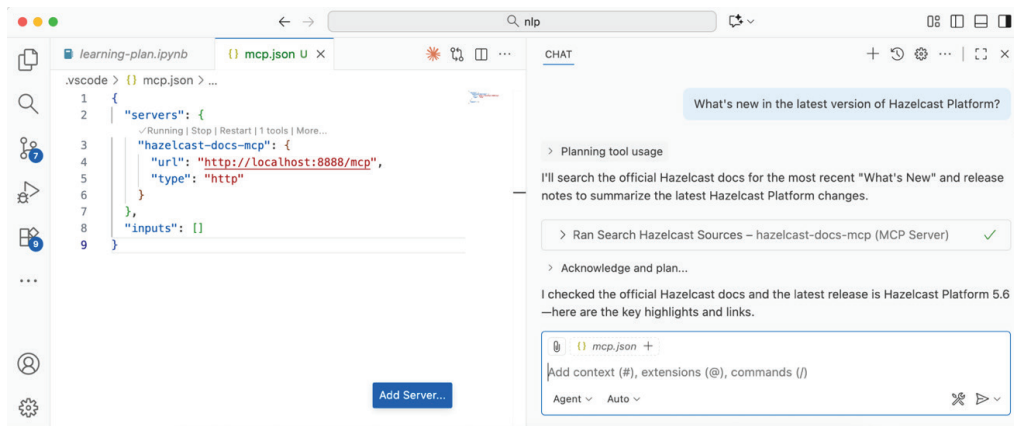


Fig. 3. Connecting from the MCP Inspector

Table 2
Key findings

Aspect	Finding	Impact
Hallucination Reduction	Reduction vs. ungrounded LLM	Significantly improved answer accuracy
Optimal Context Size	5,000-10,000 tokens	Best balance of relevance and performance
Answer Completeness	Multi-source synthesis	Better than traditional keyword search

Context window management emerged as critical. Even with large context windows, careful chunk selection is essential – including too much context can degrade response quality by diluting attention. The optimal strategy retrieves only the most relevant chunks rather than maximising context usage.

The layered architecture, with clear separation of concerns, proved highly maintainable, as each layer was developed and tested independently. The REST API wrapper approach (rather than full MCP implementation) provided the necessary functionality without excessive complexity, demonstrating that pragmatic solutions often outperform over-engineered ones.

Conclusions. This article presented the architecture and implementation of an intelligent documentation assistant built on the MCP, demonstrating a practical approach to integrating LLM with technical documentation systems. The proposed four-layer architecture – comprising a documentation website, an MCP server, AI-powered context generation, and LLM integration – successfully separates concerns while enabling seamless information flow from documentation sources to user responses.

Implementation findings validate the architectural approach while revealing important practical considerations. The MCP protocol introduces minimal overhead while providing significant benefits in

standardisation, maintainability, and flexibility. Response quality improvements over both traditional search and ungrounded LLM queries are substantial, with marked reductions in hallucination rates and increases in answer completeness. However, success depends critically on the quality of documentation – the intelligent assistant amplifies both the strengths and weaknesses of existing documentation.

The system's practical implications extend beyond immediate documentation assistance. The patterns demonstrated – utilising MCP as an integration layer, employing AI for automated context preparation, and combining retrieval with generation – apply to a wide range of knowledge-intensive applications. Organisations implementing similar systems should anticipate shifts in documentation workflows, support patterns, and user behaviour.

The MCP represents an important step toward standardised LLM integration patterns. By adopting and contributing to such open protocols, the technical community can avoid fragmentation and proprietary lock-in while enabling innovation and interoperability. The documentation assistant use case serves as a compelling demonstration of MCP's value and a template for future applications, successfully achieving the research objectives of designing an effective MCP-based system, investigating integration patterns, evaluating AI-powered context generation, and providing practical guidance for practitioners.

References

1. Huang L., Yu W., Ma W., Zhong W., Feng Z., Wang H., Chen Q., Peng W., Feng X., Qin B., Liu T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv preprint arXiv:2311.05232.
2. Gao Y., Xiong Y., Gao X., Jia K., Pan J., Bi Y., Dai Y., Sun J., Wang M., Wang H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.
3. Xu Z., Jain S., Kankanhalli M. (2024). Hallucination is inevitable: An innate limitation of large language models. arXiv preprint arXiv:2401.11817.
4. Tonmoy S. M. T. I., Zaman S. M., Jain V., Rani A., Rawte V., Chadha A., Das A. (2024). A comprehensive survey of hallucination mitigation techniques in large language models. arXiv preprint arXiv:2401.01313.
5. Cheng J., Liu Q., Liu T., Wang J., Zhou A., Sun X. (2024). Dated data: Tracing knowledge cutoffs in large language models. arXiv preprint arXiv:2403.12958.
6. Kasai J., Sakaguchi K., Takemoto Y., Lee K., Bisk Y. (2023). Evaluating GPT-3.5 and GPT-4 on diverse factual text generation tasks. arXiv preprint arXiv:2305.14251.
7. Anthropic. (2024). Model Context Protocol specification. URL: <https://modelcontextprotocol.io/>
8. Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., Riedel S., Kiela D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*. 33. 9459–9474.
9. Ram O., Levine Y., Dalmedigos I., Muhlgay D., Shashua A., Leyton-Brown K., Shoham Y. (2023). In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*. 11. 1316–1331.
10. kapa.ai. (2025, June 27). Build an MCP Server with kapa.ai. *kapa.ai Blog*. URL: <https://www.kapa.ai/blog/build-an-mcp-server-with-kapa-ai>
11. Netlify. (2024). Writing MCPs on Netlify: A guide to implementing Model Context Protocol servers on Netlify Edge Functions. *Netlify Developer Documentation*. URL: <https://developers.netlify.com/guides/write-mcps-on-netlify/>
12. Model Context Protocol. (2024). @modelfetch/netlify: Netlify adapter for Model Context Protocol servers (Version 1.0.0) [Computer software]. *npm*. URL: <https://www.npmjs.com/package/@modelfetch/netlify>

Дата першого надходження статті до видання: 23.10.2025

Дата прийняття статті до друку після рецензування: 25.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 004.4:004.72

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-04>ЕВОЛЮЦІЯ АРХІТЕКТУР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ
ОБЧИСЛЮВАЛЬНИХ СИСТЕМ: ВІД МОНОЛІТНИХ ДО МІКРОСЕРВІСНИХ

Бабич С. М., к.т.н., доц., Рівненський державний гуманітарний університет, Рівне, Україна.
<https://orcid.org/0000-0003-2145-6392>, stepaniia.babych@rshu.edu.ua

Ляшук Т. Г., к.ф.-м.н., Рівненський державний гуманітарний університет, Рівне, Україна.
<https://orcid.org/0000-0002-2242-7537>, taras.liashuk@rshu.edu.ua

Анотація. Проведено дослідження історичної еволюції архітектур програмного забезпечення обчислювальних систем – від перших монолітних структур до сучасних мікросервісних рішень. Розглянуто ключові етапи розвитку: ранні обчислювальні системи з простими програмними структурами (1940–1960-ті роки), становлення монолітної архітектури як фундаментально-го підходу до побудови цілісних систем, перехід до клієнт-серверних та багаторівневих (n-tier) моделей, появу розподілених об'єктних систем (CORBA, RMI) та web-сервісів (SOAP), еволюцію REST як легковагової альтернативи та формування сервіс-орієнтованої архітектури (SOA), що підготувала ґрунт для мікросервісної парадигми. Особливу увагу приділено детальному аналізу мікросервісної архітектури як сучасного стандарту побудови масштабованих, гнучких та відмовостійких програмних систем, підтримуваних екосистемою технологій контейнеризації (Docker), оркестрації (Kubernetes) та Service Mesh.

Здійснено систематичне порівняння монолітних, перехідних (клієнт-серверних, n-tier, SOA) та мікросервісних архітектур за критеріями масштабованості, гнучкості, складності підтримки, економічної ефективності та безпеки. Виявлено їхні переваги й обмеження та визначено закономірності еволюційного переходу між ними. Показано, що мікросервісний підхід є природним етапом розвитку архітектур програмного забезпечення, який забезпечує високий рівень автономності команд розробки, горизонтальну масштабованість компонентів, технологічну незалежність (поліглотне програмування) та інтеграцію з практиками DevOps і CI/CD. Водночас окреслено виклики мікросервісної архітектури – ускладнення інфраструктури, проблеми управління узгодженістю даних у розподілених системах і забезпечення безпеки міжсервісної комунікації.

Окреслено перспективні напрями розвитку архітектур програмного забезпечення: поширення serverless-підходів та event-driven моделей для реактивних систем, впровадження Service Mesh технологій для надійної комунікації, інтеграцію edge computing для обробки даних на периферії мережі, автоматизацію через Infrastructure as Code (IaC) та використання штучного інтелекту (AIOps) для інтелектуальної оптимізації, моніторингу й створення self-healing систем. Результати дослідження є корисними для архітекторів програмного забезпечення, інженерів і дослідників у галузі розподілених систем.

Ключові слова: архітектура програмного забезпечення, монолітна архітектура, мікросервіси, сервіс-орієнтована архітектура (SOA), клієнт-серверна модель, масштабованість, DevOps, контейнеризація, Kubernetes, Infrastructure as Code, еволюція програмних систем.

THE EVOLUTION OF COMPUTING SYSTEM SOFTWARE ARCHITECTURES:
FROM MONOLITHIC TO MICROSERVICES

Stepaniia Babych, Ph.D., Associate Professor, Rivne State University of the Humanities, Rivne, Ukraine.
<https://orcid.org/0000-0003-2145-6392>, stepaniia.babych@rshu.edu.ua

Taras Liashuk, Ph.D., Assistant Professor, Rivne State University of the Humanities, Rivne, Ukraine.
<https://orcid.org/0000-0002-2242-7537>, taras.liashuk@rshu.edu.ua

Abstract. The study explores the historical evolution of software architecture in computing systems – from early monolithic structures to modern microservice-based solutions. The research highlights key development stages: early computing systems with simple software structures (1940s–1960s), the establishment of monolithic architecture as a fundamental approach to building integrated systems, the transition to client–server and multi-tier (n-tier) models, the emergence of distributed object systems

(CORBA, RMI) and web services (SOAP), the evolution of REST as a lightweight alternative, and the rise of service-oriented architecture (SOA) that laid the groundwork for the microservice paradigm. Particular attention is given to the in-depth analysis of microservice architecture as the modern standard for building scalable, flexible, and fault-tolerant software systems supported by containerization (Docker), orchestration (Kubernetes), and Service Mesh technologies.

A systematic comparison of monolithic, transitional (client-server, n-tier, SOA), and microservice architectures was conducted according to scalability, flexibility, maintenance complexity, cost efficiency, and security criteria. Their advantages and limitations were identified, and evolutionary transition patterns between architectural models were defined. The study shows that the microservice approach represents a natural stage in the evolution of software architectures, providing a high degree of team autonomy, horizontal scalability of components, technological independence (polyglot programming), and integration with DevOps and CI/CD practices. At the same time, it emphasizes the challenges of microservices – infrastructure complexity, data consistency in distributed systems, and interservice communication security.

The prospective directions for the development of software architectures are outlined, including the expansion of serverless approaches and event-driven models for reactive systems, the implementation of Service Mesh technologies for reliable communication, the integration of edge computing for data processing at the network periphery, automation through Infrastructure as Code (IaC), and the use of artificial intelligence (AIOps) for intelligent optimization, monitoring, and the creation of self-healing systems. The research findings are valuable for software architects, engineers, and researchers in the field of distributed systems.

Key words: software architecture, monolithic architecture, microservices, service-oriented architecture (SOA), client-server model, scalability, DevOps, containerization, Kubernetes, Infrastructure as Code, software systems evolution.

Вступ

Сучасні обчислювальні системи відіграють ключову роль у розвитку цифрової економіки, забезпечуючи ефективне функціонування бізнес-процесів, державного управління та соціальних сервісів. Зростання обсягів даних, підвищення вимог до масштабованості, продуктивності та надійності інформаційних рішень актуалізували потребу в нових архітектурних підходах до проектування програмного забезпечення. Традиційна монолітна архітектура, яка домінувала в перших поколіннях програмних систем, забезпечувала простоту створення й розгортання, проте виявила низку обмежень за умов високої динамічності IT-ринку, глобалізації та зростання складності обчислювальних процесів.

У відповідь на зазначені виклики відбувся поступовий перехід до клієнт-серверних і багаторівневих архітектур, що відобразив тенденцію до підвищення модульності, розподілу відповідальності та незалежності компонентів системи. Подальший розвиток розподілених об'єктних технологій, web-сервісів і REST-підходів сприяв стандартизації міжкомпонентної взаємодії та уніфікації механізмів обміну даними. Формування сервіс-орієнтованої архітектури (SOA) створило концептуальне підґрунтя для становлення сучасної мікросервісної парадигми, яка поєднує автономність сервісів із можливостями контейнеризації, оркестрації та автоматизації процесів розгортання.

У цьому контексті актуальною є наукова проблема, яка полягає в систематизації знань про еволюцію архітектур програмних систем та розробленні критеріїв вибору оптимальних архітектурних рішень для різних типів обчислювальних систем. Незважаючи на активне впровадження мікросервісних підходів у промисловій розробці, залишається недостатньо дослідженим питання системного аналізу чинників, які зумовили перехід від монолітних до розподілених архітектур, а також узагальнення переваг і обмежень кожного етапу цієї еволюції.

Питання еволюції архітектур програмного забезпечення активно досліджується як в українській, так і в зарубіжній науковій літературі. У праці А. В. Гавриляка та К. І. Ліщук [1] систематизовано основні архітектурні шаблони та стилі проектування програмних систем, зокрема клієнт/сервер, компонентну, багатошарову, N-рівневу архітектури, архітектуру на базі шини повідомлень, об'єктно-орієнтовану та сервісно-орієнтовану архітектури (SOA). Для кожного стилю описано принципи проектування, переваги та сфери застосування. Дослідники підкреслюють важливість правильного вибору архітектури для забезпечення масштабованості, гнучкості та зручності обслуговування програмних систем.

У роботі В. Киселевича та співавторів [2] проаналізовано переваги та недоліки мікросервісної архітектури на прикладах конкретних систем різних масштабів. Автори демонструють, що мікросервісна архітектура не є універсальним рішенням і може бути як ефективним підходом, так і джерелом значних проблем. Особливу увагу приділено випадкам нераціонального застосування мікросервісів у системах середньої складності. Дослідники наголошують на необхідності ґрунтовного аналізу вимог до масштабованості, складності управління та витрат на обслуговування перед прийняттям рішення про вибір архітектурного підходу.

Серед зарубіжних досліджень важливе місце займають фундаментальні праці С. Newman [3] та С. Richardson [4], які систематизують принципи проектування мікросервісних систем і пропонують практичні шаблони їх реалізації. Значний внесок у розуміння питань масштабованості, надійності

та обробки великих обсягів даних зроблено в [5], де аналізуються архітектурні рішення для побудови розподілених систем, орієнтованих на стійкість до збоїв.

Автори в роботі [6] узагальнюють підходи до побудови програмних архітектур, розкриваючи взаємозв'язок між архітектурними рішеннями, якістю програмного продукту та бізнес-вимогами. Дослідження [7] акцентує увагу на інтеграції мікросервісної архітектури з DevOps-підходами, що дозволяє забезпечити безперервну доставку та високу надійність сервісів. У свою чергу, у [8] систематизуються антипатерни мікросервісних систем і пропонується таксономія типових архітектурних помилок, що мають практичну цінність для уникнення ризиків під час реалізації розподілених систем.

Отже, проведений аналіз свідчить, що сучасні дослідження зосереджуються переважно на практичних аспектах мікросервісних архітектур, проте питання комплексного порівняння еволюційних етапів розвитку архітектур програмних систем, а також виявлення закономірностей переходу від монолітних до мікросервісних підходів залишаються недостатньо розробленими. Це зумовлює потребу в подальшій систематизації наукових положень і критеріях оцінювання ефективності архітектурних рішень.

Виконання досліджень

Історичні передумови розвитку архітектур програмного забезпечення.

Ранні етапи розвитку обчислювальних систем (1940–1960-ті роки) характеризувалися обмеженими апаратними ресурсами та відсутністю розвинених засобів програмування. Програмні продукти мали лінійну або процедурну структуру, оскільки основна увага зосереджувалася на ефективному виконанні обчислень за умов мінімальної пам'яті та обчислювальної потужності. Зі зростанням складності задач виникла потреба в структуризації коду, що започаткувало концепцію модульності та поклато початок формуванню архітектурних підходів [1].

Подальший розвиток апаратного забезпечення та мов програмування в 1970–1980-х роках зумовив перехід від неструктурованого коду до організованих програмних систем. Першим системним етапом цього переходу стала поява монолітної архітектури, яка передбачала створення програми як цілісного набору взаємозалежних модулів, що виконувалися в одному процесі і розгорталися як єдине ціле. Такий підхід забезпечував централізоване управління бізнес-логікою та спрощував інтеграцію компонентів у єдину систему, що було ефективним у контексті обмежених обчислювальних середовищ того часу.

Монолітна архітектура характеризується низкою ключових рис, що визначають її внутрішню структуру та принципи організації програмного забезпечення. Основними структурними особливостями моноліту є єдина кодова база, централізоване управління ресурсами та простота розгортання системи як єдиного модуля. Хоча ці характеристики забезпечували швидкий старт і простоту розробки, з часом вони породили суттєві обмеження в масштабованості, супроводі та гнучкості, особливо для великих і динамічних систем. Ці обмеження стали поштовхом до пошуку нових архітектурних рішень, орієнтованих на розподіл функціональності та підвищення адаптивності програмного забезпечення.

Обмеження та проблеми монолітної архітектури

Високий рівень інтеграції компонентів і значний обсяг кодової бази зумовили низку проблем, пов'язаних із підтримкою, супроводом та масштабуванням монолітних систем. У межах такої архітектури всі компоненти об'єднані в єдиний програмний модуль, тому будь-які зміни в одному елементі можуть впливати на роботу інших. Це ускладнює внесення нових функцій, виправлення помилок і проведення тестування, що поступово знижує гнучкість та керованість програмного продукту.

Зі зростанням складності систем і кількості користувачів ці обмеження ставали дедалі помітнішими, породжуючи технічний борг і зменшуючи ефективність розробки. Обмежена масштабованість монолітної архітектури означає, що систему можна масштабувати лише цілком, без можливості гнучкого розподілу навантаження між окремими компонентами. Це призводить до нераціонального використання ресурсів і збільшення витрат на інфраструктуру.

З економічної точки зору, монолітна архітектура демонструє порівняно низькі витрати на початкових етапах, оскільки вся система створюється як єдине ціле без складної інфраструктури. Однак у довгостроковій перспективі витрати на підтримку та оновлення зростають, адже кожна зміна вимагає повторного тестування всього продукту, а накопичення технічного боргу потребує додаткових ресурсів для його усунення.

У результаті монолітна архітектура, яка свого часу забезпечувала простоту створення та розгортання, поступово втратила ефективність у динамічному середовищі сучасних обчислювальних систем. Це зумовило перехід до нових архітектурних підходів, орієнтованих на розподіл функцій між компонентами, – передусім до клієнт-серверної моделі.

Перехідні архітектури

Проблеми монолітних систем, пов'язані з масштабуванням і підтримкою, стимулювали перехід до більш гнучких архітектурних рішень. Клієнт-серверна архітектура стала першим кроком у цьому напрямі, розділивши функціонал на два компоненти: клієнт надсилає запити, а сервер обробляє їх, виконує бізнес-логіку та забезпечує збереження даних [6]. Такий підхід відкрив шлях від централізованих монолітів до розподілених систем, що дало змогу ізолювати частину навантаження та підвищити ефективність використання ресурсів.

Подальший розвиток розподілених систем у 1990-х роках привів до появи розподілених об'єктних технологій, зокрема CORBA та RMI, які забезпечили міжплатформну взаємодію об'єктів у різних середовищах. З поширенням інтернету відбувся перехід до Web-сервісів, де для стандартизованої взаємодії використовувався протокол SOAP – обмін XML-повідомленнями з суворою типізацією та вбудованими механізмами безпеки. Проте складність SOAP і громіздкість XML обмежували його гнучкість.

Це стимулювало розвиток багаторівневих (n-tier) архітектур, які розділили систему на рівні представлення, бізнес-логіки та даних. Така структура підвищила модульність, спростила супровід і створила основу для незалежного масштабування окремих рівнів.

Прагнення до спрощення комунікацій призвело до появи REST-підходу, запропонованого Роем Філдіном у 2000 році. REST використовує стандартні можливості протоколу HTTP (GET, POST, PUT, DELETE) і компактні формати обміну даними (переважно JSON), що забезпечує простоту реалізації, високу продуктивність і масштабованість завдяки stateless-взаємодії.

Інтеграція зазначених ідей відбулася в сервіс-орієнтованій архітектурі (SOA), де програмні компоненти реалізуються як незалежні сервіси, що взаємодіють через стандартизовані інтерфейси. SOA забезпечила інтеграцію великомасштабних систем через єдину шину сервісів (ESB) і сприяла повторному використанню компонентів. Водночас значний розмір сервісів і спільне використання ресурсів залишали певні обмеження щодо гнучкості та автономності.

Мікросервісна архітектура стала наступним етапом еволюції, усунувши ці обмеження. Вона розбиває систему на дрібні, автономні сервіси, кожен з яких виконує окрему бізнес-функцію. Мікросервіси використовують легковагові протоколи взаємодії (HTTP/REST, gRPC), мають власні бази даних (підхід database-per-service) і позбавляються централізованих компонентів на кшталт ESB. Це забезпечує гнучкість, прискорює впровадження змін та підвищує автономність команд розробки.

Таким чином, еволюція архітектур програмного забезпечення – від клієнт-серверної моделі до сервіс-орієнтованих підходів – відображає послідовний рух у напрямі підвищення автономності, масштабованості та адаптивності систем. Кульмінацією цього процесу стало формування мікросервісної архітектури, яка поєднує принципи розподіленості з гнучкістю сучасних технологічних рішень.

Мікросервісна архітектура: суть та принципи

Послідовна еволюція архітектур програмного забезпечення привела до появи мікросервісної парадигми, яка поєднує переваги розподілених систем і принципів модульності. Мікросервісна архітектура ґрунтується на ідеї поділу складної системи на набір невеликих, незалежних сервісів, кожен з яких виконує окрему бізнес-функцію згідно з принципом єдиного обов'язку [3]. Такий підхід забезпечує ізоляцію змін, спрощує тестування та розгортання, підвищуючи загальну гнучкість системи.

Основою мікросервісної архітектури є декомпозиція, що передбачає розподіл за бізнес-можливостями (управління замовленнями, оплати, профілі користувачів) або за предметними областями в межах Domain-Driven Design. Використання підходу bounded contexts дозволяє чітко визначити межі відповідальності кожного сервісу. Важливу роль відіграє також подієво-орієнтована взаємодія, що забезпечує асинхронний обмін повідомленнями та мінімізує залежності між компонентами.

Переваги мікросервісів полягають у підвищеній відмовостійкості, можливості незалежного розгортання та швидкому впровадженні оновлень. Разом з тим, така декомпозиція потребує ретельного планування архітектури. Використання шаблонів проектування – Circuit Breaker, Event Sourcing, CQRS – сприяє забезпеченню надійності та узгодженості розподіленої системи [4].

Ключовим елементом взаємодії між сервісами є інтерфейси програмування застосунків (API). Вони забезпечують стандартизований спосіб комунікації між сервісами, дозволяють реалізувати слабке зв'язування та приховують внутрішню реалізацію компонентів. Найчастіше застосовують RESTful API або gRPC, які забезпечують ефективну передачу даних і підтримують версіонування для сумісності. Безпека комунікації реалізується за допомогою OAuth 2.0, JWT і механізмів обмеження частоти запитів.

Для управління взаємодією на рівні всієї системи використовується API Gateway, який виконує функції балансування навантаження, контролю доступу, маршрутизації запитів і агрегації відповідей. Це спрощує інтеграцію клієнтів і приховує внутрішню складність мікросервісної інфраструктури.

Ефективне розгортання множини мікросервісів потребує застосування контейнеризації та оркестрації. Контейнери, зокрема Docker, забезпечують стандартизовану упаковку сервісів разом із залежностями, що підвищує портативність і швидкість розгортання. Платформа Kubernetes автоматизує керування життєвим циклом контейнерів, включаючи масштабування, балансування навантаження, самовідновлення при збоях та оновлення без простоїв.

У поєднанні технології Docker і Kubernetes формують основу сучасного середовища для розробки й експлуатації мікросервісних систем. Вони забезпечують автономність компонентів, підтримку CI/CD-процесів та високу доступність, реалізуючи на практиці ключові переваги мікросервісної архітектури.

Отже, мікросервісна архітектура формує сучасну парадигму побудови програмних систем, у якій поєднано модульність, розподіленість та гнучкість. Завдяки використанню контейнеризації, оркестрації й стандартизованих API забезпечується висока автономність компонентів, стабільність

і масштабованість системи. Ці характеристики визначають ключові переваги мікросервісного підходу, які стали підґрунтям його широкого впровадження у промислових програмних рішеннях.

Переваги мікросервісів

Мікросервісна архітектура забезпечує високу гнучкість та адаптивність завдяки розподіленню системи на автономні сервіси, кожен із яких можна розробляти, тестувати й оновлювати незалежно [2]. Це дає змогу швидко впроваджувати нові функції, виправляти помилки без зупинки всієї системи та знижувати ризики під час оновлень. Такий підхід суттєво скорочує час виведення продукту на ринок і підвищує стійкість до змін бізнес-вимог. Використання поліглотного програмування дозволяє обирати оптимальний технологічний стек для кожної задачі – наприклад, Python для аналітики та машинного навчання, Go для високопродуктивних сервісів, Node.js для веб-застосунків. Автоматизація процесів через CI/CD та використання контейнерних технологій (Docker, Kubernetes) прискорюють розгортання, а стратегії blue-green deployment і canary releases мінімізують простой, забезпечуючи можливість безпечного відкату змін.

Однією з ключових переваг є гнучка масштабованість. Кожен мікросервіс може масштабуватися окремо, відповідно до рівня навантаження, що забезпечує ефективне горизонтальне масштабування (scale-out). Наприклад, при зростанні кількості транзакцій можна масштабувати лише сервіс обробки платежів, оптимізуючи використання ресурсів та витрати на інфраструктуру. Платформа Kubernetes підтримує динамічне балансування навантаження завдяки автоматичному масштабуванню (Horizontal Pod Autoscaler), що гарантує еластичність системи. Додатково, географічно розподілене розгортання забезпечує низьку затримку обробки запитів та високу доступність навіть у разі регіональних збоїв.

Мікросервісна архітектура також гармонійно поєднується з практиками DevOps [7], сприяючи інтеграції розробки, тестування та експлуатації в єдиний безперервний процес. CI/CD автоматизує всі етапи життєвого циклу програмного забезпечення, що дозволяє впроваджувати зміни за години замість тижнів. Кожен сервіс має власний pipeline розгортання, що підвищує стабільність і передбачуваність оновлень. Використання підходу Infrastructure as Code (Terraform, Ansible) забезпечує відтворюваність і консистентність конфігурацій у різних середовищах. Формування культури “you build it – you run it” стимулює відповідальність команд і сприяє створенню більш надійного та якісного коду.

Попри значні переваги, мікросервісна архітектура має і низку специфічних викликів, пов'язаних із підвищеною складністю управління, координацією сервісів та забезпеченням їхньої узгодженої роботи в розподіленому середовищі.

Виклики та обмеження мікросервісної архітектури

Разом із перевагами мікросервісна архітектура супроводжується суттєвими інфраструктурними та організаційними викликами. На відміну від монолітних систем, мікросервісний підхід передбачає наявність множини незалежних сервісів, кожен із яких необхідно окремо розгортати, моніторити та масштабувати. Це ускладнює загальне управління системою, потребує стандартизації процесів, централізованого логування, розподіленого трасування та підтримки численних репозиторіїв коду. Додаткову складність створює різноманітність технологічних стеків і версій фреймворків, що впливає на сумісність компонентів та процес інтеграції.

Ефективне управління розподіленими системами потребує використання оркестраційних платформ (Kubernetes, Helm, Istio), брокерів повідомлень (Kafka, RabbitMQ) і технологій service mesh. Мережева взаємодія між сервісами піддається ризикам затримок (latency) або тимчасової недоступності, що може спричинити каскадні збої. Тестування множини взаємозалежних сервісів у локальному середовищі є складним завданням, а перехід до мікросервісної архітектури вимагає значних організаційних та технологічних ресурсів. Ці фактори обумовлюють необхідність ретельного планування переходу й інвестицій у навчання персоналу [8]. Серед технічних викликів особливе місце посідає проблема управління даними.

Управління даними є одним із ключових викликів мікросервісної архітектури через її розподілений характер [5]. Кожен сервіс має власну базу даних (database-per-service), що забезпечує автономність, але ускладнює підтримку узгодженості та транзакційної цілісності. Якщо в монолітних системах використовуються ACID-транзакції, то в мікросервісах частіше застосовується підхід eventual consistency – відкладеної узгодженості даних. Для мінімізації ризиків та підтримки консистентності використовуються архітектурні патерни Event Sourcing, CQRS (Command Query Responsibility Segregation) і Saga, які дозволяють координувати розподілені транзакції між сервісами.

Проблемним є також виконання аналітичних запитів, які потребують об'єднання даних із кількох сервісів, оскільки традиційні SQL-операції (JOIN) у такому середовищі непридатні. Це вимагає побудови аналітичних сховищ або реалізації агрегаційних шарів через API. Серед додаткових викликів – забезпечення продуктивності (через мережеві накладні витрати), організація резервного копіювання великої кількості баз даних, дотримання вимог законодавства (наприклад, GDPR) та реалізація наскрізного шифрування.

Ще одним критичним аспектом є безпека мікросервісного середовища. Оскільки кожен сервіс взаємодіє через відкриті мережеві інтерфейси, розширюється поверхня потенційних атак. Основни-

ми завданнями є автентифікація, авторизація та шифрування даних. Для цього застосовуються протоколи OAuth 2.0, JWT, а також mutual TLS (mTLS) для двосторонньої автентифікації. Технології Service Mesh (Istio, Linkerd) забезпечують автоматизацію безпечної маршрутизації трафіку через sidecar proxies.

Забезпечення спостережуваності та відмовостійкості є ще одним важливим напрямом. Централізоване логування із використанням correlation ID дозволяє відстежувати транзакції та діагностувати аномалії за допомогою інструментів типу ELK Stack або Splunk. Підвищення надійності систем досягається впровадженням патернів Circuit Breaker, Retry та Timeout, які мінімізують наслідки збоїв. Платформа Kubernetes додатково забезпечує мережеву ізоляцію через Network Policies, автоматичне відновлення сервісів після збоїв і безпечне управління секретами.

Усвідомлення цих викликів і обмежень стимулює розвиток нових технологій, методологій DevSecOps та архітектурних підходів, спрямованих на підвищення стабільності, безпеки й керованості розподілених систем.

Сучасні тенденції та перспективи розвитку

У 2025 році serverless-архітектури посіли провідне місце серед хмарних технологій, ставши ключовим етапом еволюції розподілених обчислень. У цій моделі інфраструктура повністю управляється хмарним провайдером, тоді як розробники зосереджуються виключно на бізнес-логіці, реалізованій у вигляді функцій (Function as a Service, FaaS). Провідні платформи – AWS Lambda, Azure Functions та Google Cloud Functions – забезпечують автоматичне масштабування, високу доступність і модель оплати pay-per-use. Інтеграція serverless із edge computing сприяє розгортанню функцій ближче до користувача (наприклад, Cloudflare Workers, Lambda@Edge), що мінімізує затримки обробки даних і підвищує ефективність для IoT-та інтерактивних застосунків.

Event-driven архітектури (EDA) поступово стають стандартом для побудови слабо зв'язаних систем, у яких взаємодія компонентів здійснюється через обмін подіями у event bus. Сучасні реалізації базуються на технологіях Apache Kafka, RabbitMQ, а також на хмарних сервісах – AWS EventBridge, Azure Event Grid. Поєднання EDA із патернами Event Sourcing, CQRS та Saga забезпечує масштабованість та стійкість до збоїв, однак породжує виклики, пов'язані з налагодженням систем і підтриманням eventual consistency. Ефективна реалізація таких архітектур потребує стабільної інфраструктури для управління потоками подій і гарантій доставки повідомлень.

Service Mesh формується як окремий інфраструктурний шар, що забезпечує безпечну та керовану комунікацію між мікросервісами за допомогою sidecar проху. Провідні рішення – Istio (з компонентами Envoy, traffic management, mTLS, телеметрією), Linkerd (легковагове рішення на Rust) та Consul Connect, який підтримує як Kubernetes, так і віртуальні машини. Ключові можливості Service Mesh охоплюють інтелектуальну маршрутизацію, автоматичне шифрування (mTLS), збирання метрик і реалізацію механізмів відмовостійкості. На сьогодні ця технологія стала стандартом для production-ready систем із підвищеними вимогами до безпеки та керованості.

Edge computing активно інтегрується у сферу IoT, автономних транспортних систем, AR/VR та промислової автоматизації. Його метою є обробка даних безпосередньо на периферії мережі, що знижує затримки та зменшує навантаження на центральні сервери. Основні моделі включають Fog Computing, Mobile Edge Computing (5G), CDN Edge і On-Premise Edge. Інтеграція з мікросервісною архітектурою через KubeEdge, AWS Greengrass, Azure IoT Edge сприяє розвитку концепцій Industry 4.0 та розумних міст (Smart Cities).

Паралельно розвивається підхід Infrastructure as Code (IaC), який досяг зрілості завдяки інструментам Terraform, Ansible, Pulumi. Методологія GitOps (ArgoCD, FluxCD) використовує Git як єдине джерело істини (single source of truth), тоді як Policy as Code (OPA, Sentinel) забезпечує автоматичну перевірку дотримання стандартів безпеки та комплаєнсу. Новим напрямом є впровадження штучного інтелекту у сферу DevOps – AIOps – для аналізу операційних даних, прогнозування збоїв і автоматичного масштабування систем. Використання Deep Reinforcement Learning та Graph Neural Networks оптимізує топологію мікросервісів, тоді як машинне навчання у моніторингу дає змогу виявляти аномалії в реальному часі. Великі мовні моделі (наприклад, GitHub Copilot) підвищують продуктивність команд, сприяючи створенню self-healing систем і автоматизації тестування.

Висновок

Аналіз еволюції архітектур програмного забезпечення засвідчив закономірний рух від централізованих та жорстко інтегрованих систем до гнучких, масштабованих і динамічних моделей. Монолітна архітектура, що колись забезпечувала простоту та надійність, у сучасних умовах виявила низку обмежень, які були поступово подолані завдяки появі клієнт-серверних, багаторівневих, розподілених об'єктних систем, web-сервісів, REST та сервіс-орієнтованих підходів. Кожен етап еволюції сприяв формуванню концепцій незалежності компонентів, стандартизації інтерфейсів та розподіленої обробки даних.

Мікросервісна архітектура стала відповіддю на виклики цифрової трансформації, оскільки поєднує незалежність компонентів, можливість гнучкого масштабування та інтеграцію з сучасними практиками розробки (DevOps, CI/CD). Підтримка контейнеризації, оркестрації та Service Mesh технологій

забезпечує ефективне управління складними розподіленими системами. Разом із тим мікросервісна архітектура породжує нові виклики, пов'язані з управлінням інфраструктурою, узгодженістю даних та безпекою міжсервісної комунікації. Проте перехід до мікросервісів забезпечує організаціям конкурентні переваги через швидкість інновацій, автономність команд та здатність до оперативного масштабування відповідно до бізнес-потреб.

Подальший розвиток архітектур програмного забезпечення пов'язаний із поширенням serverless-підходів та event-driven моделей, впровадженням Service Mesh для надійної комунікації, інтеграцією edge computing для обробки даних на периферії мережі, автоматизацією інфраструктури через Infrastructure as Code та застосуванням штучного інтелекту для інтелектуальної оптимізації систем (AIOps). Ці тенденції відкривають нові горизонти для створення високопродуктивних, надійних та адаптивних обчислювальних систем із властивостями self-healing, здатних автономно реагувати на зміни навантаження та вимоги сучасного цифрового бізнес-середовища.

Література

1. Гавриляк А. В., Ліщук К. І. Архітектурні шаблони та стилі програмного забезпечення. *Вісник НТУУ «КПІ». Серія: Інформатика та обчислювальна техніка*. 2020. № 73. С. 47–54.
2. Киселевич В., Усата О., Сікора Я., Вербівський Д., Іванов Д. Мікросервісна архітектура: переваги та недоліки її практичного застосування. *Інформаційні технології та комп'ютерна інженерія*. 2024. № 2 (63). С. 45–54. DOI: <https://doi.org/10.32782/IT/2024-2-7>.
3. Newman S. *Building Microservices: Designing Fine-Grained Systems*. 2nd ed. Sebastopol, CA : O'Reilly Media, 2021. 612 p.
4. Richardson C. *Microservices Patterns: With Examples in Java*. Shelter Island, NY : Manning Publications, 2018. 520 p.
5. Kleppmann M. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. Sebastopol, CA : O'Reilly Media, 2017. 616 p.
6. Bass L., Clements P., Kazman R. *Software Architecture in Practice*. 4th ed. Boston, MA : Addison-Wesley Professional, 2021. 464 p.
7. Waseem M., Liang P., Shahin M. A Systematic Mapping Study on Microservices Architecture in DevOps. *Journal of Systems and Software*. 2020. Vol. 170. Article 110798. DOI: <https://doi.org/10.1016/j.jss.2020.110798>.
8. Taibi D., Lenarduzzi V., Pahl C. Microservices Anti-patterns: A Taxonomy. In: *Microservices: Science and Engineering*. Cham : Springer, 2020. P. 111–128. DOI: https://doi.org/10.1007/978-3-030-31646-4_5.

References

1. Havryliak A. V., Lishchuk K. I. (2020). Arkhitekturni shablony ta styli prohramnoho zabezpechennia [Architectural patterns and software design styles]. *Visnyk NTUU "KPI". Informatyka ta obchysliuvalna tekhnika [Bulletin of NTUU "KPI". Informatics and Computer Engineering]*. (73). 47–54 [in Ukrainian].
2. Kyselevych V., Usata O., Sikora Ya., Verbivskyi D., Ivanov D. (2024). Mikroservisna arkhitektura: perevahy ta nedoliky yii praktychnoho zastosuvannia [Microservice architecture: Advantages and disadvantages of its practical implementation]. *Informatsiini tekhnolohii ta komp'uterna inzheneriia [Information Technologies and Computer Engineering]*. 2 (63). 45–54. DOI: <https://doi.org/10.32782/IT/2024-2-7> [in Ukrainian].
3. Newman S. (2021). *Building microservices: Designing fine-grained systems* (2nd ed.). Sebastopol, CA : O'Reilly Media.
4. Richardson C. (2018). *Microservices patterns: With examples in Java*. Shelter Island, NY : Manning Publications.
5. Kleppmann M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. Sebastopol, CA : O'Reilly Media.
6. Bass L., Clements P., Kazman R. (2021). *Software architecture in practice* (4th ed.). Boston, MA : Addison-Wesley Professional.
7. Waseem M., Liang P., Shahin M. (2020). A systematic mapping study on microservices architecture in DevOps. *Journal of Systems and Software*. 170. 110798. DOI: <https://doi.org/10.1016/j.jss.2020.110798>.
8. Taibi D., Lenarduzzi V., Pahl C. (2020). Microservices anti-patterns: A taxonomy. In *Microservices: Science and engineering* (pp. 111–128). Cham : Springer. DOI: https://doi.org/10.1007/978-3-030-31646-4_5.

Дата першого надходження статті до видання: 20.10.2025
 Дата прийняття статті до друку після рецензування: 21.11.2025
 Дата публікації (оприлюднення) статті: 26.12.2025

УДК 621.396:002.53

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-05>

ЗАХИСТ ВІД АКТИВНИХ ЗАВАД БЕЗДРОТОВИХ СИСТЕМ ЗВ'ЯЗКУ З АДАПТИВНИМИ ФАЗОВАНИМИ АНТЕННИМИ РЕШІТКАМИ

Безрук В. М., д.т.н., проф., завідувач кафедри інформаційно-мережної інженерії Факультет інфокомунікацій, Харківський національний університет радіоелектроніки, Харків, Україна. <https://orcid.org/0000-0002-5482-9960>, valerii.bezruk@nure.ua

Голобородько Ю. М., к.т.н., доцент кафедри безпеки інформаційних технологій Факультет комп'ютерної інженерії та управління Харківський національний університет радіоелектроніки, Харків, Україна. <https://orcid.org/0009-0007-4040-9467>, yurii.holoborodko@nure.ua

Заболотний В. І., к.т.н., доцент, професор кафедри безпеки інформаційних технологій Факультет комп'ютерної інженерії та управління Харківський національний університет радіоелектроніки, Харків, Україна. <https://orcid.org/0000-0003-3258-8489>, volodymyr.zabolotnyi@nure.ua

Анотація. Актуальність та проблема. Надійність зв'язку є критично важливою для військових та силових структур, але бездротові системи вразливі до активних завад (глушіння, імітація). Ці завади мають ворожу природу, динамічно адаптуються та становлять пряму загрозу безпеці. Аналіз загроз показує, що активні перешкоди, на відміну від пасивних (тепловий шум, атмосферні), вимагають адаптивних рішень для збереження стійкості комунікацій у сучасному, насиченому радіоелектронному середовищі.

Мета роботи. Обґрунтування та дослідження методології захисту бездротових систем зв'язку від активних завад шляхом інтеграції адаптивних фазованих антенних решіток (АФАР) та засобів радіо- та радіотехнічної розвідки (РПТР). Дослідження зосереджено на використанні АФАР для динамічного керування діаграмою спрямованості (ДС) та ефективного пригнічення перешкод.

Методи дослідження. Запропонована методологія ґрунтується на просторово-частотній обробці сигналу з використанням АФАР. Підхід передбачає два ключові етапи: 1) використання засобів РПТР для оперативного виявлення, розпізнавання та пеленгування джерел активних завад; 2) адаптоване керування ДС АФАР (як приймальних, так і передавальних) для формування глибоких провалів («нулів») у напрямку перешкод та максимізації енергії корисного сигналу у напрямку кореспондента. Розглянуто різні алгоритми керування, включаючи методи на основі кореляційної матриці та градієнтні методи (математична оптимізація), які є найбільш ефективними для систем реального часу.

Основні результати та практичне значення. Доведено, що інтеграція АФАР із системами зв'язку є ефективним механізмом підвищення стійкості, що дозволяє зменшити вплив потужних перешкод ще до етапу демодуляції. Сумарний виграш від використання АФАР у порівнянні з всеспрямованою антеною може сягати 20–30 дБ, що складається з конструктивного виграшу (6–15 дБ) та виграшу від активного подавлення перешкод (10–20 дБ). Цей значний приріст у співвідношенні сигнал/перешкода забезпечує надійний зв'язок в умовах високої щільності завад. Розроблені рішення мають стратегічне значення для підвищення захищеності критично важливих комунікаційних систем.

Ключові слова: адаптивна фазована антенна решітка, активні завади, радіоелектронна боротьба (РЕБ), радіотехнічна розвідка (РПТР), динамічне керування, просторово-частотна обробка.

PROTECTION AGAINST ACTIVE INTERFERENCE OF WIRELESS COMMUNICATION SYSTEMS USING ADAPTIVE PHASED ARRAY ANTENNAS

Valerii Bezruk, Dr. habil, Prof., Head of Department of Information and Network Engineering of Faculty of Infocommunications of Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. <https://orcid.org/0000-0002-5482-9960>, valerii.bezruk@nure.ua

Yurii Holoborodko, Ph.D., Ass. Prof of Department of Information Technology Security of Faculty of Computer Engineering and Control of Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. <https://orcid.org/0009-0007-4040-9467>, yurii.holoborodko@nure.ua

Volodymyr Zabolotnyi, Ph.D., Ass. Prof., Prof of Department of Information Technology Security of Faculty of Computer Engineering and Control of Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. <https://orcid.org/0000-0003-3258-8489>, volodymyr.zabolotnyi@nure.ua

Abstract. Relevance and Problem. Communication reliability is vital for military and security operations, but wireless systems are susceptible to active interference (jamming, spoofing). Such interference is hostile, dynamically adapts, and poses a direct security threat. The analysis shows that active interference, unlike

passive noise, necessitates adaptive solutions to maintain communication resilience in today's saturated radio-electronic environment.

Objective. To substantiate and investigate a methodology for protecting wireless communication systems against active interference by integrating Adaptive Phased Array Antennas (APAAs) with Radio and Radiotechnical Reconnaissance (RRTR) assets. The study focuses on using APAAs for dynamic control of the radiation pattern (RP) to effectively suppress interference.

Methods. The proposed methodology is based on spatial-frequency signal processing using APAAs. The approach involves two key stages: 1) utilization of RRTR assets for operational detection, recognition, and direction finding (DF) of active interference sources; 2) adaptive control of the APAA's RP (for both transmitting and receiving) to form deep nulls in the direction of interference and maximize the energy of the useful signal towards the correspondent. Various control algorithms are discussed, including correlation matrix-based methods and gradient methods (mathematical optimization), which are most efficient for real-time systems.

Main Results and Practical Significance. The integration of APAAs into communication systems is proven to be an effective mechanism for enhancing resilience, allowing for the mitigation of powerful interference before the demodulation stage. The total gain from using APAAs, compared to an omnidirectional antenna, can reach 20–30 dB, comprising the constructive gain (6–15 dB) and the gain from active interference suppression (10–20 dB). This significant increase in the signal-to-interference ratio ensures reliable communication under high interference density. The developed solutions are of strategic importance for strengthening the security of critical communication systems.

Key words: adaptive phased array antenna, active interference, electronic warfare (EW), radio-technical reconnaissance (RRTR), dynamic control, spatial-frequency processing.

Вступ

Надійний зв'язок є критично важливим для успіху військових операцій та функціонування силових структур, забезпечуючи синхронізацію дій, оперативне виявлення загроз та стратегічну координацію. З розвитком сучасних технологій та широким впровадженням цифрових систем, залежність озброєння та бойової техніки від сенсорів та обміну інформацією постійно зростає, що підкреслює вирішальну роль радіоелектронної боротьби (РЕБ). У цьому контексті, ефективність армії майбутнього значною мірою визначається її здатністю гнучко адаптувати доктрини, структури та методи підготовки до розвитку РЕБ [1; 2].

Особливе значення в РЕБ має радіо- та радіотехнічна розвідка (РПТР), яка відповідає за виявлення, пеленгування та аналіз джерел електромагнітного випромінювання противника [3]. Якщо стратегічна РПТР зосереджена на підготовці до операцій, то тактична РПТР ведеться безперервно під час бойових дій, надаючи критично важливу інформацію для управління військами та озброєнням. Однак бездротові системи зв'язку, попри свої переваги (передача інформації на великі відстані, обмін даними через важкодоступну місцевість, швидке встановлення зв'язку та мовлення багатьом кореспондентам), є вразливими до низки факторів. До них належать атмосферні перешкоди, можливість пеленгування та перехоплення інформації противником, а головне – подавлення радіозв'язку засобами РЕБ противника, що призводить до низької якості та пропускну здатності каналів зв'язку.

Декаметровий (короткохвильовий, КХ/ВЧ, 3,0–30,0 МГц) радіозв'язок, що забезпечує зв'язок на значні відстані завдяки відбиттю від іоносфери, є особливо чутливим до нестабільності середовища поширення та активних завад [3; 4]. Сучасні радіосистеми поділяються на неадаптивні, що вимагають постійного ручного налаштування оператором, та адаптивні, які автоматизують процеси встановлення та ведення зв'язку, що значно підвищує їхню ефективність. Попри переваги адаптивних систем, неадаптивні системи все ще використовуються, що створює потребу в універсальних рішеннях для підвищення стійкості до завад.

Проблема захисту бездротових систем зв'язку від активних завад є критично важливою не лише для військових, але й для правоохоронних органів (СБУ, МВС) та регуляторних органів (НКРЕ), які стикаються з викликами контролю за радіочастотним спектром, протидії злочинним угрупованням та забезпечення чистоти урядових каналів зв'язку. У зв'язку з цим, методи захисту, що ґрунтуються на використанні адаптивних фазованих антенних решіток (АФАР), набувають особливої актуальності. Ці технології дозволяють динамічно керувати діаграмою спрямованості антени, максимізуючи прийом корисного сигналу та мінімізуючи вплив активних завад, що є ключовим для підвищення надійності та безпеки бездротових комунікацій в умовах сучасного інформаційного середовища.

Виконання досліджень

Одним із найважливіших чинників, що впливають на якість та надійність бездротових телекомунікаційних систем, є вплив шумових завад. У контексті рухомих об'єктів ці фактори стають ще критичнішими, адже зв'язок має підтримуватися в умовах змінного середовища, різної швидкості об'єкта

та можливого навмисного зовнішнього втручання. Загалом за характером джерела всі перешкоди поділяються на пасивні та активні.

Пасивні перешкоди виникають внаслідок природних або стохастичних процесів, без спрямованого наміру порушити роботу системи. Вони є невід'ємною частиною будь-якого реального фізичного каналу зв'язку. До основних типів пасивних перешкод належать:

1. Тепловий шум (Johnson-Nyquist noise): Виникає внаслідок теплового руху носіїв заряду в провіднику або напівпровіднику. Характеризується рівномірним спектром і присутній у будь-якому електронному колі. Його потужність прямо пропорційна температурі та смузі пропускання.

2. Адитивний білий гаусівський шум (AWGN): Теоретична модель випадкового шуму, який нормально розподілений і широко використовується для тестування телекомунікаційних систем як наближення до природного фону.

3. Фононний або фотонний шум: Характерний для оптоелектронних і квантових пристроїв, з'являється у високочастотних підсистемах, де теплові флуктуації взаємодіють з фотонними процесами (наприклад, у сенсорах, лазерах).

4. Атмосферні перешкоди: Включають перешкоди від блискавок, сонячної активності, а також радіовипромінювання земного та космічного походження.

Такі перешкоди не є ворожими, проте здатні призводити до зниження відношення сигнал / шум (SNR), втрати пакетів та збільшення помилок, особливо за низької потужності сигналу. У русі цей ефект посилюється через зміну кута падіння хвилі, доплерівський зсув та багатопроменевість.

Активні перешкоди – це навмисне створення завад з метою впливу на канал зв'язку. Їх джерелом може бути як інша техніка, так і зловмисник, який має на меті порушити передачу, заглушити сигнал, перехопити або змінити дані. Основні типи активних шумових перешкод:

1. Шум випромінювання: Електромагнітне випромінювання пристроїв або спеціально створене джерело перешкоди, що працює у тому ж частотному діапазоні (наприклад, використання генератора білого шуму для придушення Wi-Fi сигналу).

2. Імпульсні перешкоди: Короточасні інтенсивні сплески напруги або випромінювання. Можуть виникати при запуску двигунів, переключенні реле або цілеспрямованому електромагнітному імпульсі (EMI).

3. Глушіння (Jamming): Навмисне створення потужного сигналу на робочій частоті каналу для його блокування. Існують різні типи: безперервне, псевдовипадкове, імпульсне, направлене, адаптивне глушіння.

4. Імітаційні перешкоди (Spoofing): Створення сигналу, що імітує легітимне джерело, з метою обману системи виявлення або навігації (наприклад, GPS spoofing).

Активні перешкоди становлять реальну загрозу інформаційній безпеці, оскільки призводять не лише до завад, а й до порушення функцій зв'язку – зниження доступності, збільшення затримки, втрати керованості. У випадку безпілотних систем або автономних транспортних засобів це може призвести до аварій або порушень навігації.

Радіоелектронна боротьба (РЕБ) – це комплекс узгоджених дій, спрямованих на використання електромагнітної енергії для управління спектром, впливу на радіоелектронні системи противника та захисту власних засобів. Згідно з керівними документами Армії США, радіоелектронна війна (РЕВ) охоплює радіоелектронну атаку, радіоелектронний захист та забезпечення ведення РЕВ, підкреслюючи її роль як елемента інформаційних операцій. У Збройних Силах України термін РЕБ [3] включає організацію виконання завдань щодо планування кібероборони України, планування та ведення радіоелектронної боротьби та кіберборотьби, управління радіочастотним спектром (далі – РЧС), розвитком відповідних спроможностей.

РРТР, як невід'ємна частина РЕБ, відіграє ключову роль у виявленні, розпізнаванні та визначенні місцезнаходження джерел активних завад. Вона забезпечує збір оперативних даних про ворожі радіовипромінювання, що є основою для подальших дій РЕБ. Радіоелектронне подавлення, своєю чергою, передбачає безпосередній вплив на радіоелектронні системи противника шляхом електромагнітного випромінювання, застосування оманних цілей, дезінформуючих повідомлень, імітації роботи або зміни умов поширення радіохвиль.

Висновки з аналізу загроз: Проведений аналіз показує, що пасивні перешкоди, хоч і є невід'ємною частиною будь-якого каналу зв'язку, здебільшого піддаються математичному моделюванню та компенсації. Активні перешкоди, навпаки, мають ворожу природу і можуть динамічно адаптуватися до параметрів сигналу, що робить їх надзвичайно небезпечними, особливо для систем з рухомими об'єктами. Традиційні методи боротьби з такими перешкодами мають обмежену ефективність [6]. Це обґрунтовує нагальну потребу у використанні адаптивних фазованих антенних решіток (АФАР), які здатні динамічно змінювати напрямок прийому та випромінювання, формувати провали в діаграмі спрямованості на джерело завади та відновлювати надійний зв'язок навіть у складних умовах.

Для подолання проблем, пов'язаних із частотним дефіцитом, впливом завад та високою щільністю користувачів у мобільних мережах, перспективним напрямом є використання адаптивних антен-

них систем. Ці системи забезпечують динамічне керування діаграмою спрямованості (ДС) антенних решіток залежно від місцеположення джерел сигналу та завад, потужності сигналу та поточних радіоумов. Вони дозволяють формувати вузькі направлені промені у бік абонента, одночасно мінімізуючи випромінювання в напрямку джерел завад. Такий підхід забезпечує просторове розділення користувачів, зменшує рівень міжканальних інтерференцій та безпосередньо підвищує якість сервісу й надійність передачі даних [5; 7; 8].

З точки зору захисту інформації та зменшення вразливості до завад, адаптивні антени дозволяють не лише фокусувати сигнал у потрібному напрямку, а й активно пригнічувати небажані сигнали шляхом формування «нулів» у діаграмі спрямованості у визначених напрямках. Це створює просторовий бар'єр для потенційних злоумисників, відкриваючи широкі можливості для застосування таких систем у військових, критично важливих та спеціалізованих комунікаційних системах, де питання безпеки та захищеності є пріоритетними.

У запропонованому дослідженні розглядається методологія захисту, що базується на використанні просторово-частотної обробки сигналу з АФАР. Підхід передбачає поєднання таких ключових технік:

1) використання засобів РРТР для виявлення, розпізнавання та визначення напрямків на джерела пасивних та активних перешкод;

2) адаптоване керування діаграмою спрямованості антенних решіток засобів зв'язку (як приймальних, так і передавальних) для придушення високопотужних сигналів перешкод та формування максимуму випромінювання сигналу у бік розташування приймальної сторони.

Такий підхід дозволяє зменшити вплив потужних перешкод ще до етапу демодуляції та одночасно збільшити ефективну потужність випромінювання корисного сигналу. Розвиток цифрової техніки обробки сигналів дає змогу будувати складні антенні системи, вихідний сигнал яких формується в результаті обробки сигналів, прийнятих кожним елементом антени [7; 8]. Ця обробка покращує характеристики антени та співвідношення сигнал/перешкода порівняно з антенами без такої обробки. Зокрема, алгоритми обробки дозволяють автоматично пригнічувати перешкоди, що діють у ділянці бічних пелюсток ДС антени, формуючи «провали» у їхньому напрямку. Найважливішими якісними показниками для алгоритмів придушення бічних пелюсток є швидкість адаптації та ступінь придушення перешкод. Отриманий вииграш у співвідношенні сигнал/перешкода залежить від потужностей сигналу та перешкод, їхньої кількості та просторового розташування.

На рис. 1 зображено структурну схему системи зв'язку з використанням АФАР з адаптивним керуванням діаграмами спрямованості.

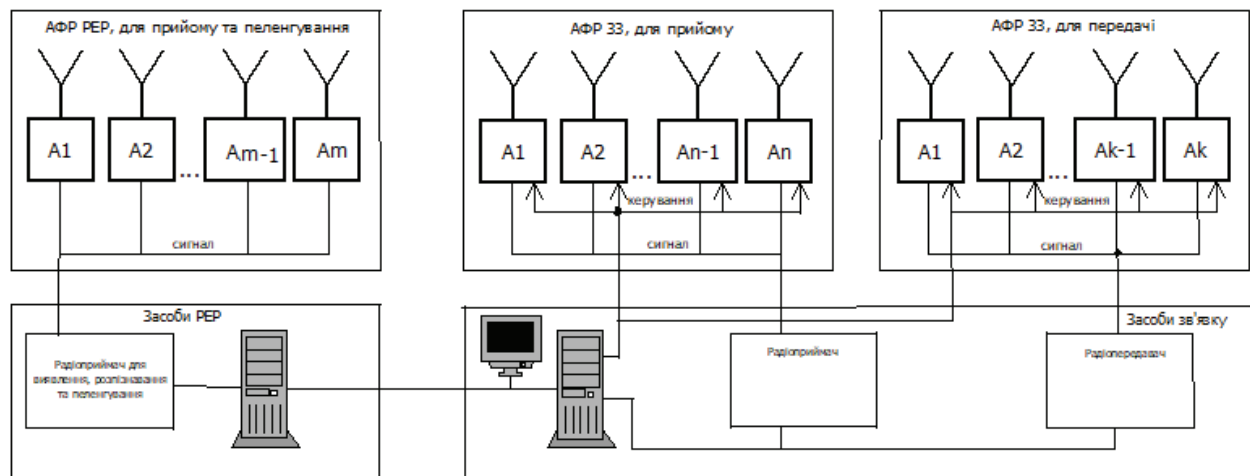


Рис. 1. Структурна схема системи зв'язку з використанням АФАР з адаптивним керуванням діаграм спрямованості

Система зв'язку складається з:

- системи РРТР (радіоелектронної розвідки);
- засобів зв'язку (ЗЗ).

Система РРТР включає:

- антенну систему, яка може бути реалізована як антенною решіткою з елементів $A_1 \dots A_m$, так і просто пеленгаційною антеною;
- радіоприймач, що сканує та виконує функції пеленгування [3];
- систему обробки та керування, призначену для виявлення, розпізнавання та визначення напрямків на джерела активних перешкод, а також передачі цих даних до засобів зв'язку.

Засоби зв'язку включають:

- АФАР з елементів $A_1 \dots A_n$ для прийому;
- АФАР з елементів $A_1 \dots A_k$ для передачі;
- радіоприймач;
- радіопередавач;

– систему обробки та керування, яка отримує дані про джерела активних перешкод, розраховує за цими даними оптимальні діаграми спрямованості АФАР, керує фазообертачами елементів АФАР для формування необхідних ДС, а також управляє радіоприймачем та радіопередавачем.

Принцип дії АФАР полягає в динамічному керуванні характеристиками випромінювання окремих активних антенних елементів. Головна мета – забезпечити максимальну чутливість антени у напрямку корисного сигналу, одночасно мінімізуючи рівень прийому завадових сигналів з інших напрямків. В ідеалі, у напрямках перешкод діаграма спрямованості повинна мати нульове значення, що відповідає повному пригніченню завад. Хоча на практиці досягти абсолютного нуля неможливо, у технічній термінології використовується поняття «нульова характеристика» у напрямку джерел шуму та перешкод.

Для реалізації такого керування характеристиками адаптивного антенного масиву застосовуються спеціальні алгоритми, які можна розділити на кілька основних категорій, зазначених нижче.

1. Алгоритми на основі кореляційної матриці.

Кожен активний елемент антенного масиву оснащений власним приймачем. На основі сигналів, прийнятих усіма елементами, формується кореляційна матриця, за якою розраховуються оптимальні вагові коефіцієнти, що максимізують співвідношення корисного сигналу до сумарних перешкод і шуму (SINR). Цей підхід є одним із найефективніших, оскільки використовує повну інформацію про прийняті сигнали, проте вимагає більшої кількості приймачів та регулярної калібровки.

2. Алгоритми, засновані на апріорній інформації про джерело перешкод.

Базуються на точних відомостях про напрямок джерела перешкод та характеристики елементів антени (атенюаторів, фазообертачів). На основі цієї інформації система математично визначає ваги, які формують «нуль» у напрямку завади. Цей метод ефективний за наявності достовірних даних про джерело перешкоди, для отримання яких пропонується використовувати засоби РРТР.

3. Алгоритми випадкового перебору.

Передбачають випадковий перебір різних варіантів налаштування вагових коефіцієнтів для антенного масиву. Хоча теоретично цей метод може призвести до оптимального рішення, він вимагає значного часу на обчислення, що робить його непридатним для систем реального часу.

4. Алгоритми математичної оптимізації (градієнтні методи).

Вагові коефіцієнти елементів антени обираються так, щоб мінімізувати загальну потужність на виході адаптивної системи. Зазвичай для цього формується вектор градієнта, який послідовно підбирає вагові коефіцієнти для досягнення мінімуму потужності, знижуючи рівень шуму та перешкод.

Таким чином, адаптивні алгоритми забезпечують ефективне управління характеристиками антенного масиву, підвищуючи якість прийому корисного сигналу навіть у складних умовах із наявністю значних перешкод.

Висновки

Проведене дослідження підтверджує, що використання адаптивних фазованих антенних решіток (АФАР) є ефективним методом для значного підвищення стійкості бездротових систем зв'язку до впливу активних завад. Завдяки здатності до динамічного керування діаграмою спрямованості, АФАР посилює корисний сигнал та ефективно пригнічує перешкоди, що є особливо критичним в умовах швидкоплинних змін у радіоелектронному середовищі.

Як показують попередні дослідження (або: за результатами проведеного аналізу /моделювання), сумарний виграш від адаптивних фазованих антенних решіток у системах мобільного чи супутникового зв'язку може сягати:

- 6–15 дБ за рахунок спрямованості (конструктивний виграш);
- 10–20 дБ за рахунок активного подавлення перешкод.

Таким чином, загальний виграш у системі може становити 20–30 дБ при використанні АФАР у порівнянні з всеспрямованою антеною. Це суттєво покращує стійкість до завад, ефективність передачі та захист від цілеспрямованих атак. Інтеграція АФАР із системами РРТР для оперативного виявлення та пеленгування джерел завад є ключовим фактором для реалізації їх повного потенціалу.

Перспективи подальших досліджень можуть включати наступне.

Проведення додаткових експериментів для оцінки ефективності запропонованих захисних механізмів у різних реальних сценаріях.

Удосконалення програмних моделей та алгоритмів керування адаптивними фазованими антенними решітками з метою підвищення їх продуктивності та швидкості адаптації.

Придільнення особливої уваги інтеграції захисних рішень на фізичному та програмному рівнях для створення комплексних систем захисту з підвищеним рівнем безпеки та надійності.

Загалом, подальший розвиток цієї галузі сприятиме значному посиленню захищеності бездротових систем зв'язку та підвищенню їх ефективності в умовах сучасного, насиченого інформаційного середовища.

Література

1. Головне управління радіоелектронної та кіберборотьби. *Офіційний веб-сайт МО*. URL: <https://www.mil.gov.ua/ministry/struktura-generalnogo-shtabu/golovne-upravlinnya-radioelektronnoi-ta-kiberborotbi.html>
2. Довідник військового зв'язківця. Засоби радіоелектронної боротьби та розвідки, які використовуються російською федерацією. Київ : ЦУЛ, 2024. 64 с.
3. Голобородько Ю. М., Радіоконтроль діапазону Short Waves, Research in Science, Technology and Economics : collection of Scientific Papers. *International Scientific Unity* : Proceedings of the 2nd International Scientific and Practical Conference (March 5–7, 2025. Luxembourg). Luxembourg. 397 p. Pp. 275–278. ISBN 979-8-89704-985-1 (series). DOI: <https://doi.org/10.70286/ISU-05.03.2025>.
4. Короткохвильовий радіозв'язок : настанова. Київ : ЦУЛ, 2024. 83 с.
5. Omelchenko V. A. Balabanov V. V. Bezruk V. M. Goloborod'ko Ju. N. Detection of Changes in the Random Signal Properties by Spectral Methods Under the Conditions of A Priori Uncertainty. *Telecommunications and Radio Engineering*. 1999. 53 (9–10). P. 1–10.
6. Заславець В. П., Долина М. П., Чечуй О. В. Особливості розрахунку завадозахищеності ліній радіозв'язку в умовах радіоподавлення (радіоелектронного конфлікту). DOI: <https://doi.org/10.30748/soivt.2020.61.01>.
7. Пресняков І. М., Ситник О. В. Алгоритм адаптивного формування діаграми напрямленості антенної решітки в умовах інтенсивних сигналів і перешкод. *Радіотехніка* : міжвідомчий науково-технічний збірник. 1988. Вип. 87. С. 71–78.
8. Плющ О. Г., Рибидайло А. А. Алгоритм налаштування адаптивної антенної решітки, який не потребує присутності опорного сигналу. *Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняховського*. 2019. № 2 (66). С. 108–114.

References

1. Holovne upravlinnia radioelektronnoi ta kiberborotby. *Ofitsiyni veb-sait MO* [Main Directorate of Radio Electronic and Cyber Warfare. *Official website of the Ministry of Defense*]. URL: <https://www.mil.gov.ua/ministry/struktura-generalnogo-shtabu/golovne-upravlinnya-radioelektronnoi-ta-kiberborotbi.html> [in Ukrainian].
2. Dovidnyk viiskovoho zv'iazkivtsia. Zasoby radioelektronnoi borotby ta rozvidky, yaki vykorystovuiutsia rosiiskoiu federatsiiei. Kyiv : "TsUL", [Military Signaller's Handbook. Electronic Warfare and Intelligence Means Used by the Russian Federation. Kyiv : "TsUL"], 2024. 64 p. [in Ukrainian].
3. Holoborodko Y. M., Radiokontrol' diapazonu Short Waves. [Short Wave Radio Monitoring]. Research in Science, Technology and Economics : collection of Scientific Papers. *International Scientific Unity* : Proceedings of the 2nd International Scientific and Practical Conference (March 5–7, 2025. Luxembourg). Luxembourg. 397 p. Pp. 275–278. ISBN 979-8-89704-985-1 (series). DOI: <https://doi.org/10.70286/ISU-05.03.2025>.
4. Korotkikhvyl'nyi radiozviazok: nastanova. Kyiv : "TsUL", [Shortwave radio communication: a guide. Kyiv : "TsUL"] 2024. 83 p. [in Ukrainian].
5. Omelchenko V. A. Balabanov V. V. Bezruk V. M. Goloborod'ko Ju. N., Detection of Changes in the Random Signal Properties by Spectral Methods Under the Conditions of A Priori Uncertainty. *Telecommunications and Radio Engineering*. 1999. 53 (9–10). Pp. 1–10.
6. Zaslavets V. P., Dolyna M. P., Chechui O. V. Osoblyvosti rozrakhunku zavadozakhyshchenosti linii radiozviazku v umovakh radiopodavleniia (radioelektronnoho konfliktu) [Peculiarities of calculating the interference immunity of radio communication lines in conditions of radio suppression (radioelectronic conflict)]. URL: <https://doi.org/10.30748/soivt.2020.61.01> [in Ukrainian].
7. Presniakov I. M., Sytnyk O. V. Alhorytm adaptynoho formuvannia diahramy napriamlenosti antennoi reshitky v umovakh intensyvnykh syhnaliv i pereshkod [Algorithm for adaptive formation of the antenna array directivity diagram in conditions of intense signals and interference]. *Radio engineering, interdepartmental scientific and technical collection*. 1988. 87. P. 71–78 [in Ukrainian].
8. Pliushch O. H., Rybydailo A. A. Alhorytm nalashtuvannia adaptynoi antennoi reshitky, yakyi ne potrebuie prysutnosti opornoho syhnalu [Adaptive antenna array tuning algorithm that does not require the presence of a reference signal]. *Collection of scientific papers of the Center for Military and Strategic Studies of the Ivan Chernyakhovsky National Defense University of Ukraine* 2019. № 2 (66). P. 108–114 [in Ukrainian].

Дата першого надходження статті до видання: 27.10.2025
 Дата прийняття статті до друку після рецензування: 26.11.2025
 Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.42:378.147:005.7

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-06>

ПРОЄКТУВАННЯ ПРОГРАМНОЇ СИСТЕМИ ПЛАНУВАННЯ ЧАСУ НА ОСНОВІ КРОСПЛАТФОРМНИХ ТЕХНОЛОГІЙ

Ващіліна О. В., к. ф.-м. н., доц., Київський національний університет імені Тараса Шевченка, Київ, Україна. <https://orcid.org/0000-0001-6867-6216>, olenavashchilina@knu.ua

Голотюк В. М., бакалавр, Київський національний університет імені Тараса Шевченка, Київ, Україна. <https://orcid.org/0009-0008-3947-4613>, viktoriia_holotiuk@knu.ua

Анотація. У статті розглянуто процес проєктування навчально орієнтованої програмної системи для автономного планування часу, що відповідає сучасним вимогам кроссплатформності, адаптивності та конфіденційності. Актуальність дослідження зумовлена потребою в гнучких цифрових інструментах тайм-менеджменту, які можуть бути модифіковані відповідно до індивідуальних потреб користувача. Попри наявність широкого спектра готових рішень, зокрема Todoist, Microsoft To Do та Any.do, актуальним залишається завдання створення відкритої архітектури, орієнтованої на освітнє середовище та практичне засвоєння принципів проєктування програмного забезпечення. У межах дослідження здійснено порівняльний аналіз згаданих цифрових рішень, сформульовано функціональні та нефункціональні вимоги до нової програмної системи, а також обґрунтовано вибір архітектурного шаблону MVVM з урахуванням принципів clean architecture, що сприяють чіткому розмежуванню функцій між компонентами. Побудовано низку UML-діаграм, що відображають функціональну та структурну модель системи: діаграма варіантів використання окреслює основні сценарії взаємодії користувача, діаграма активностей демонструє логіку виконання дій, діаграма послідовності ілюструє часову взаємодію між компонентами, а діаграма класів формалізує структуру сутностей та їх зв'язки. Реалізація системи не входить до складу цієї роботи та розглядається як наступний етап, для якого запропоновано узагальнені рекомендації щодо вибору технологічного стеку, здатного забезпечити кроссплатформну розробку, автономність та зручність впровадження. Результати роботи можуть бути використані в освітньому процесі для формування навичок архітектурного моделювання, системного мислення та підготовки до реалізації повноцінних програмних продуктів. Запропонована модель має потенціал для подальшого розширення функціоналу та інтеграції з хмарними сервісами.

Ключові слова: тайм-менеджмент; проєктування програмної системи; архітектура MVVM; UML-діаграми; кроссплатформність.

DESIGN OF TIME MANAGEMENT SOFTWARE SYSTEM BASED ON CROSS-PLATFORM TECHNOLOGIES

Olena Vashchilina, Ph.D., Ass. Prof., Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. <https://orcid.org/0000-0001-6867-6216>, olenavashchilina@knu.ua

Viktoriia Holotiuk, Bachelor, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. <https://orcid.org/0009-0008-3947-4613>, viktoriia_holotiuk@knu.ua

Abstract. The article explores the design process of an educationally oriented software system for autonomous time planning that meets modern requirements of cross-platform compatibility, adaptability, and confidentiality. The relevance of the study is driven by the need for flexible digital time-management tools that can be modified to suit individual user needs. Despite the wide range of available solutions, in particular Todoist, Microsoft To Do, and Any.do, the task of creating an open architecture focused on educational environments and the practical assimilation of software design principles remains relevant. The study includes a comparative analysis of the aforementioned digital solutions, the formulation of functional and non-functional requirements for a new software system, and the justification for selecting the MVVM architectural pattern, taking into account the principles of clean architecture, which promote a clear separation of concerns among components. A set of UML diagrams was developed to represent the system's functional and structural model: the use case diagram outlines key user interaction scenarios; the activity diagram illustrates the logic of action execution; the sequence diagram shows the temporal interaction between components; and the class diagram formalizes the structure of entities and their relationships. System implementation is beyond the scope of this work and is considered a subsequent stage, for which general recommendations are provided regarding the choice of a technology stack capable of supporting cross-platform development, offline functionality, and ease of deployment. The results of the study can be used

in the educational process to develop skills in architectural modeling, systems thinking, and preparation for the implementation of full-fledged software products. The proposed model has the potential for further functional expansion and integration with cloud services.

Key words: time management; software system design; MVVM architecture; UML diagrams; cross-platform development.

Вступ

У сучасних умовах цифрової трансформації та зростаючої багатозадачності ефективно управління часом набуває особливого значення як у професійній, так і в освітній діяльності. Тайм-менеджмент розглядається як ключовий компонент особистої продуктивності, що потребує гнучких, інтуїтивно зрозумілих та адаптивних інструментів [1]. Попри наявність широкого спектра готових програмних рішень, таких як Todoist, Microsoft To Do, Any.do [2–4], актуальним залишається завдання створення автономних та навчально орієнтованих систем, які можуть бути модифіковані відповідно до специфічних потреб користувача.

Згідно з сучасними підходами до підготовки фахівців у галузі інформаційних технологій, проектування власного застосунку в межах студентського проекту сприяє формуванню інженерного мислення, навичок архітектурного моделювання та практичного застосування сучасних технологій [5]. Такий підхід дозволяє не лише сформувати концепцію функціонального програмного продукту, а й забезпечити глибоке розуміння процесів аналізу вимог, побудови архітектури та логічної структури програмної системи.

У межах цього дослідження здійснено проектування кросплатформної програмної системи тайм-менеджменту з урахуванням принципів адаптивності, автономності та масштабованості. Особливу увагу приділено формалізації функціональних і нефункціональних вимог, побудові UML-діаграм, обґрунтуванню архітектурних рішень та логічній структурі проекту. Реалізація системи розглядається як наступний етап, для якого в роботі запропоновано рекомендації щодо вибору технологічного стеку (Flutter, Dart, SQLite), що забезпечує кросплатформність та автономну роботу.

Метою роботи є створення архітектурної моделі навчально орієнтованої програмної системи тайм-менеджменту, яка поєднує відкриту структуру, сучасні підходи до моделювання та можливість подальшої реалізації відповідно до індивідуальних потреб користувачів. Запропоноване рішення має потенціал для використання в освітньому середовищі, а також у побутовій та професійній діяльності як зручний інструмент для організації часу.

Виконання досліджень

Аналіз існуючих програмних рішень

У процесі дослідження було проведено аналіз функціональності трьох популярних програмних систем тайм-менеджменту: Todoist, Microsoft To Do та Any.do. Основна увага приділялася ключовим характеристикам, що впливають на зручність використання, гнучкість налаштувань, підтримку автономного режиму та потенціал для освітнього застосування.

Todoist – багатofункціональний застосунок із підтримкою категоризації, підзавдань, пріоритетів, нагадувань, повторюваних дедлайнів та понад 80 інтеграцій. Відзначається високою зручністю інтерфейсу та можливістю швидкого додавання завдань у природній мові [2].

Microsoft To Do – інтегрований у екосистему Microsoft 365, підтримує створення списків завдань, нагадування, групування списків, персоналізацію інтерфейсу та синхронізацію з Outlook. Водночас не передбачає ієрархії підзавдань, що обмежує гнучкість структурування [3].

Any.do – орієнтований на мобільне використання, підтримує голосові нотатки, групову роботу, інтеграцію з Google Календарем, нагадування, планування дня та спільні простори для сімей чи команд. Має простий інтерфейс і функціональність, адаптовану до повсякденного використання [4].

У таблиці 1 наведено порівняння функціональності розглянутих систем за ключовими характеристиками, зокрема: можливістю створення та організації завдань, підтримкою підзавдань, нагадуваннями, синхронізацією між платформами, інтеграцією з іншими сервісами, зручністю інтерфейсу, підтримкою групової роботи та орієнтацією на конкретні платформи.

Таблиця 1
Порівняння основних характеристик систем тайм-менеджменту

Функціонал	Todoist	Microsoft To Do	Any.do
Створення та організація завдань	+	+	+
Підтримка підзавдань	+	–	+
Нагадування та терміни	+	+	+
Синхронізація між платформами	+	+	+
Інтеграція з іншими сервісами	+	+	+
Зручність інтерфейсу	Висока	Середня	Висока
Груповою робота	–	–	+

Усі розглянуті для порівняння програмні системи забезпечують базову функціональність для створення та організації завдань, включаючи підтримку нагадувань, термінів виконання та синхронізації між платформами. Водночас лише Todoist та Any.do підтримують створення підзавдань, що є важливою функцією для деталізації планів та побудови ієрархічної структури активностей. Any.do вирізняється додатковими можливостями, такими як голосові нотатки та групова взаємодія, що розширює сценарії використання у командному середовищі. Інтерфейси Todoist та Any.do оцінюються як більш зручні для мобільного використання, тоді як Microsoft To Do орієнтований на інтеграцію з екосистемою Microsoft 365.

Усі три системи мають обмежену підтримку автономного режиму, що стало одним із чинників на користь розроблення власного застосунку з локальним зберіганням даних, відкритою архітектурою та можливістю адаптації до освітніх потреб.

Формування вимог до програмної системи

На основі проведеного аналізу існуючих рішень сформульовано вимоги до програмної системи тайм-менеджменту, орієнтованої на освітнє застосування. Вимоги охоплюють як функціональні, так і нефункціональні аспекти, що забезпечують відповідність системи практичним потребам студентів, а також її придатність для демонстрації архітектурних принципів у навчальному процесі.

Функціональні вимоги:

До функціональних вимог належать базові можливості, необхідні для ефективного планування навчальної діяльності:

- Створення, редагування та видалення завдань – забезпечує повний життєвий цикл управління активностями.
- Групування завдань за категоріями – дозволяє структурувати завдання за темами, курсами або проектами.
- Встановлення пріоритетів та дедлайнів – сприяє розстановці акцентів і контролю термінів виконання.
- Візуалізація графіка активностей – підвищує наочність планування, дозволяє оцінити навантаження в динаміці.
- Локальне зберігання даних без потреби в синхронізації – гарантує автономність роботи та конфіденційність інформації.
- Пошук та фільтрація завдань – забезпечує швидкий доступ до потрібної інформації при великій кількості записів.

Особливу увагу приділено локальному зберіганням даних, що дозволяє використовувати систему без підключення до мережі, зберігаючи при цьому контроль над персональною інформацією. Візуалізація графіка активностей має додаткову освітню цінність, оскільки дозволяє студентам краще оцінювати власну завантаженість та планувати навчальні блоки.

Нефункціональні вимоги:

Нефункціональні вимоги визначають загальні характеристики системи, що впливають на її придатність до щоденного використання:

- Кросплатформність – забезпечує доступність системи на різних пристроях, що відповідає сучасним звичкам користувачів.
- Автономність роботи без підключення до мережі – важлива для стабільної роботи в умовах обмеженого доступу до Інтернету.
- Мінімалістичний та адаптивний інтерфейс – сприяє зручності використання незалежно від розміру екрана та технічного рівня користувача.
- Швидкодія та стабільність – впливають на загальне враження від роботи системи та її ефективність.
- Можливість масштабування функціоналу – дозволяє розширювати систему в майбутньому без порушення її архітектурної цілісності.

Формування вимог здійснювалося з урахуванням специфіки освітнього середовища, потреб у мобільності, конфіденційності та можливості подальшого розвитку системи. Сформульовані вимоги стали основою для вибору архітектурного підходу, побудови логічної структури проекту та визначення технологічного стеку, що забезпечує реалізацію системи відповідно до принципів відкритості, автономності та навчальної придатності.

Архітектурне проєктування системи

Архітектура програмної системи тайм-менеджменту сформована відповідно до вимог, визначених у попередньому розділі, з урахуванням принципів модульності, масштабованості та кросплатформності. Для забезпечення прозорості структури та поведінки системи використано низку UML-діаграм, які охоплюють як статичні, так і динамічні аспекти проєкту. Наведемо діаграми у порядку, що відповідає логіці розробки: від загальної архітектури до конкретних сценаріїв і структур. Спочатку представлено архітектурний шаблон MVVM, який визначає загальну організацію компонентів системи. Далі наведено діаграму варіантів використання, що окреслює функціональні можливості з точки зору користувача. Діаграма активностей деталізує поведінку системи у межах окремих сценаріїв, а діаграма

послідовності демонструє часову взаємодію між компонентами. Завершує розділ діаграма класів, яка відображає логічну структуру сутностей та їх зв'язків.

Вибір архітектурного шаблону. Для проектування програмної системи тайм-менеджменту обрано архітектурний шаблон MVVM (Model–View–ViewModel), який забезпечує чітке розділення відповідальностей між компонентами, спрощує підтримку, тестування та масштабування застосунку. MVVM є загальноновживаним підходом у кросплатформній розробці та рекомендований у офіційній документації Microsoft [6]. У межах даного проекту реалізацію MVVM адаптовано до специфіки освітнього середовища:

- Model відповідає за структуру даних та взаємодію з локальним сховищем (SQLite), без використання зовнішніх API чи хмарних сервісів.
- ViewModel реалізує логіку обробки подій, фільтрації, оновлення інтерфейсу та навігації, не містить мережевої логіки, що спрощує тестування та підвищує автономність.
- View забезпечує мінімалістичний та адаптивний інтерфейс, орієнтований на студентське використання, з акцентом на візуалізацію графіка активностей.

Загальна структура шаблону MVVM реалізована відповідно до загальноновизнаних принципів, описаних у [6], з урахуванням автономності, конфіденційності та освітньої придатності системи.

Логічна структура проекту реалізована у вигляді модульного дерева файлів, що відповідає принципам clean architecture та забезпечує ізоляцію відповідальностей між компонентами MVVM.

Діаграма варіантів використання (Use Case Diagram). Для визначення основних сценаріїв взаємодії користувача з системою побудовано діаграму варіантів використання (рис. 1). Вона охоплює створення, редагування, фільтрацію завдань, перегляд графіка активностей, а також налаштування категорій. Інтерфейс користувача є спільним для ПК та Android, що забезпечує єдність досвіду та зручність використання незалежно від платформи.

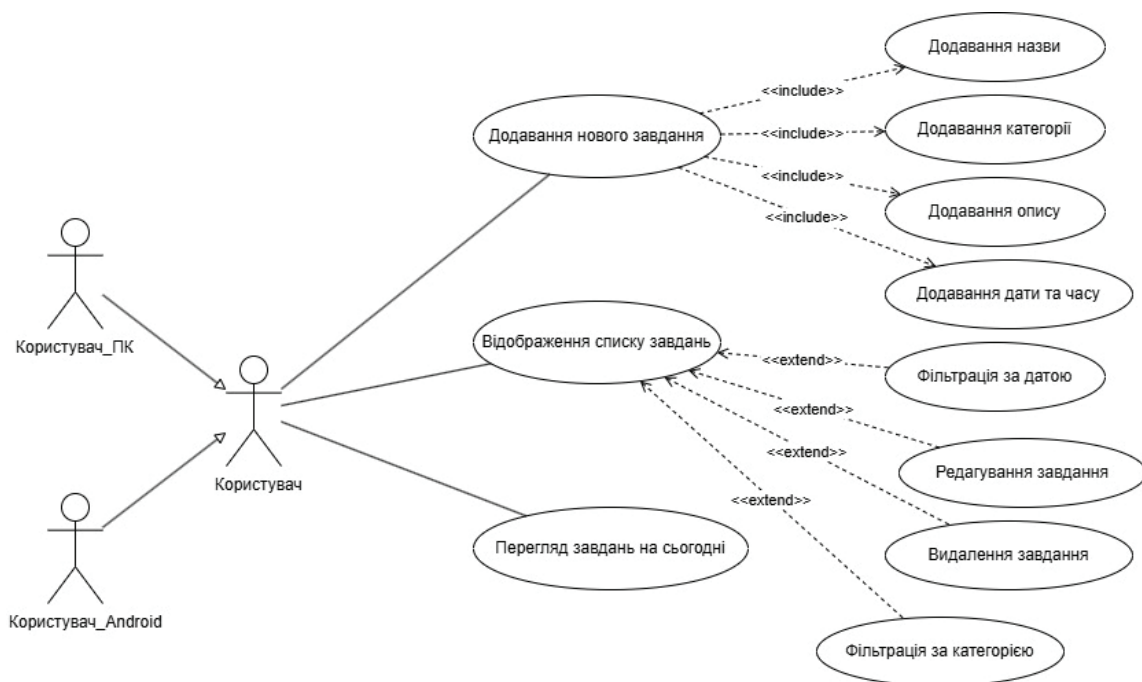


Рис. 1. Діаграма варіантів використання

Діаграма активностей (Activity Diagram). Діаграма активностей використовується для моделювання поведінки системи, зокрема послідовності дій користувача та реакцій інтерфейсу. Такий тип діаграми дозволяє візуалізувати логіку переходів між екранами, умови виконання дій, перевірку введених даних, збереження інформації у локальне сховище та оновлення інтерфейсу. Діаграма також демонструє можливі розгалуження сценаріїв залежно від вибору користувача, що сприяє точному формулюванню вимог до функціональності та поведінки програмної системи. На рис. 2 представлено два ключові сценарії взаємодії користувача із системою: додавання задачі (рис.2а) та редагування задачі (рис. 2б).

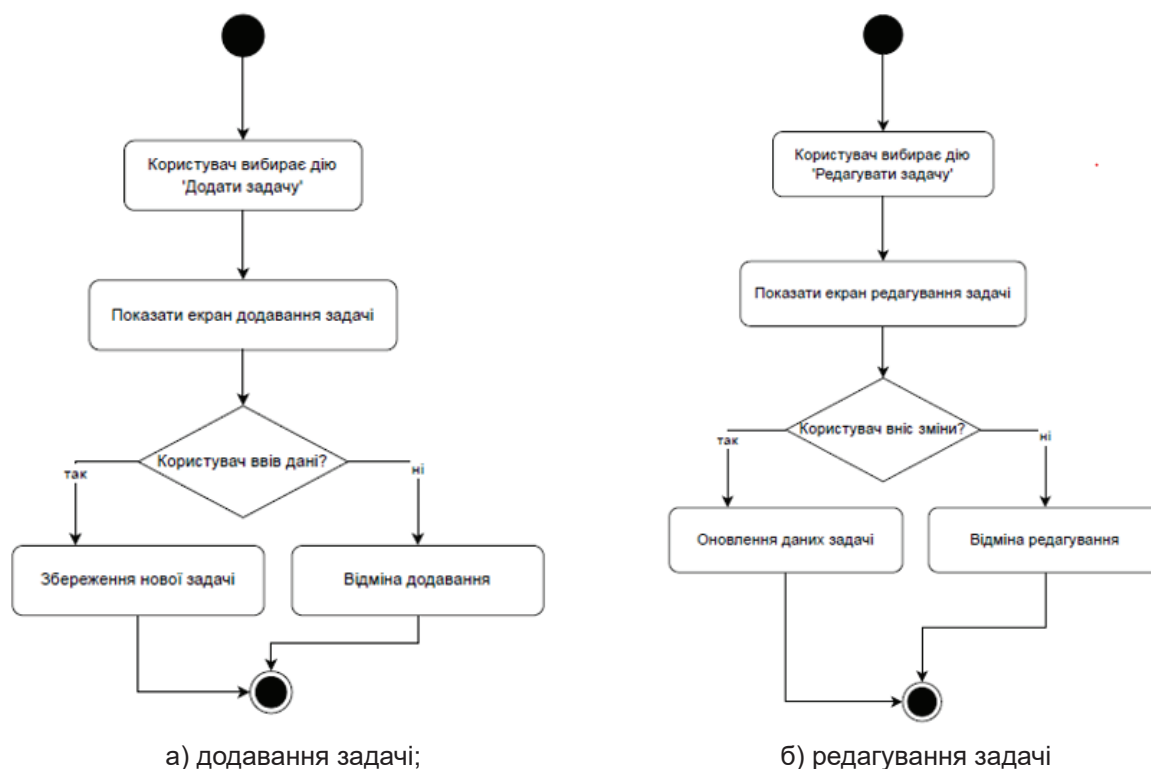


Рис. 2. Діаграма активностей для реалізації дій користувача

Діаграма послідовності (Sequence Diagrams). Для відображення динаміки взаємодії між компонентами системи під час виконання сценаріїв побудовано діаграми послідовності, які ілюструють часову послідовність викликів методів між View, ViewModel, контролерами та сервісами збереження даних. Ці діаграми дозволяють простежити, як ініціатива користувача проходить через усі рівні архітектури, забезпечуючи узгодженість логіки та відповідність шаблону MVVM. На рис. 3 представлено сценарій видалення завдання, який включає перевірку обраного елемента, звернення до локального сховища та оновлення інтерфейсу після успішного видалення.

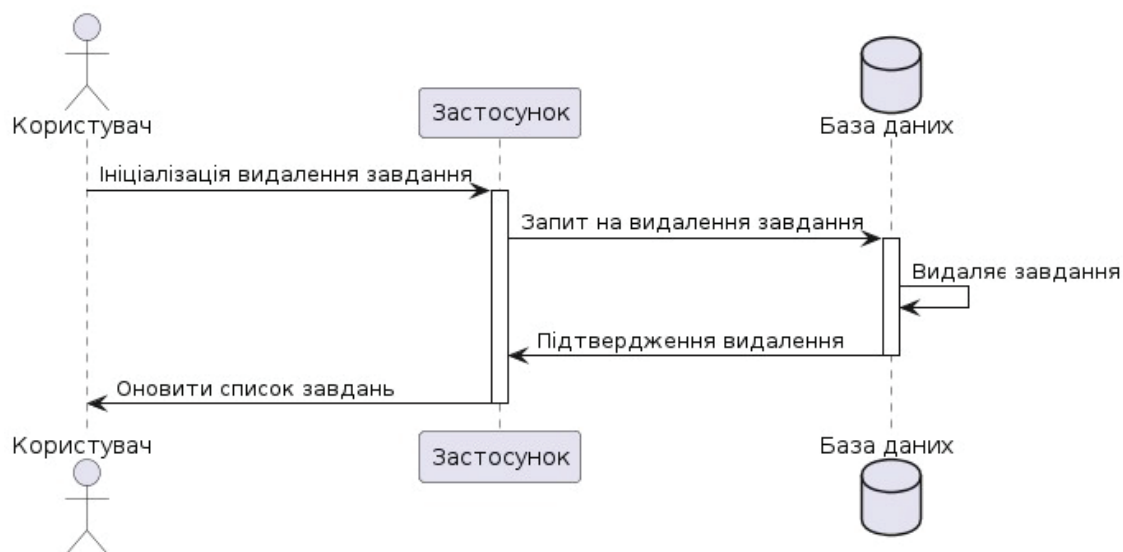


Рис. 3. Діаграма послідовності «Видалення завдання»

На рис. 4 показано процес фільтрації завдань, де ViewModel обробляє параметри фільтрації, виконує запит до сховища та оновлює відображення без модифікації даних.

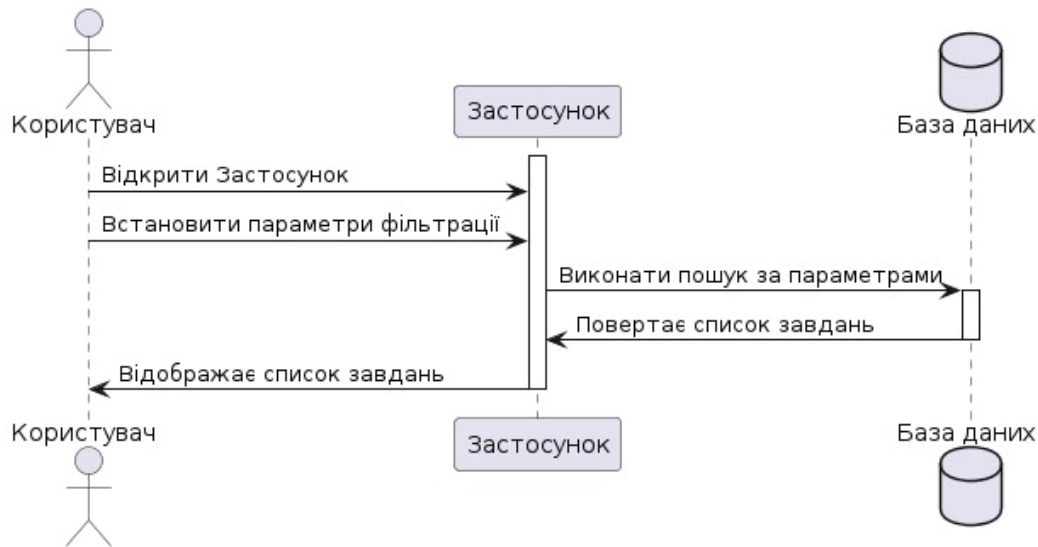


Рис. 4. Діаграма послідовності «Фільтрація завдання»

Обидва сценарії демонструють узгоджену взаємодію між компонентами системи та підтверджують ефективність застосування архітектурного шаблону MVVM у контексті кросплатформної розробки.

Діаграма класів (Class Diagram). Діаграма класів (рис. 5) відображає логічну структуру системи, включаючи основні сутності: Task, Category, Schedule, StorageService, TaskController, ViewModel. Вона демонструє зв'язки між класами, їх атрибути та методи, що реалізують функціональність системи. Клас Task містить інформацію про окреме завдання, включаючи назву, опис, дату виконання та статус. Category відповідає за класифікацію завдань, Schedule – за формування списку завдань на поточний день. StorageService реалізує доступ до локального сховища, TaskController координує взаємодію між моделлю та інтерфейсом, а ViewModel забезпечує оновлення інтерфейсу відповідно до змін у даних.

Завдяки чіткому розділенню відповідальностей між класами, система залишається масштабованою, зручною для тестування та придатною для подальшого розширення функціональності.

Рекомендації щодо вибору програмних засобів для реалізації системи

Для реалізації програмної системи тайм-менеджменту, спроектованої відповідно до архітектурного шаблону MVVM, рекомендовано використовувати набір програмних засобів, що забезпечують кросплатформність, автономність, масштабованість та зручність розробки. Вибір інструментів обґрунтовано специфікою освітнього середовища, потребою в локальному зберіганні даних та можливістю швидкої адаптації інтерфейсу. До рекомендованого стеку технологій належать:

- Flutter Framework – сучасний фреймворк для кросплатформної розробки мобільних застосунків, що дозволяє створювати єдину кодову базу для мобільних платформ (Android, iOS), а також для настільних операційних систем – Windows, macOS і Linux [7].

- Dart Programming Language – мова програмування, оптимізована для роботи з Flutter, підтримує реактивну модель та спрощує реалізацію ViewModel-компонентів [8].

- Flutter Widgets – набір вбудованих компонентів для побудови адаптивного інтерфейсу, що відповідає принципам MVVM [7].

- Flutter Packages, зокрема:

- o *provider* – для реалізації зв'язку між ViewModel та View [9];

- o *flutter_local_notifications* – для локальних нагадувань [10];

- o *shared_preferences* – для збереження налаштувань користувача [11].

- SQLite – локальна база даних, яка відповідає вимогам автономності, конфіденційності та простоти інтеграції як у мобільні застосунки, так і в настільні програмні рішення [12].

Застосування зазначених засобів дозволить ефективно реалізувати архітектурну модель MVVM, забезпечити масштабованість та адаптацію системи до потреб освітнього процесу.

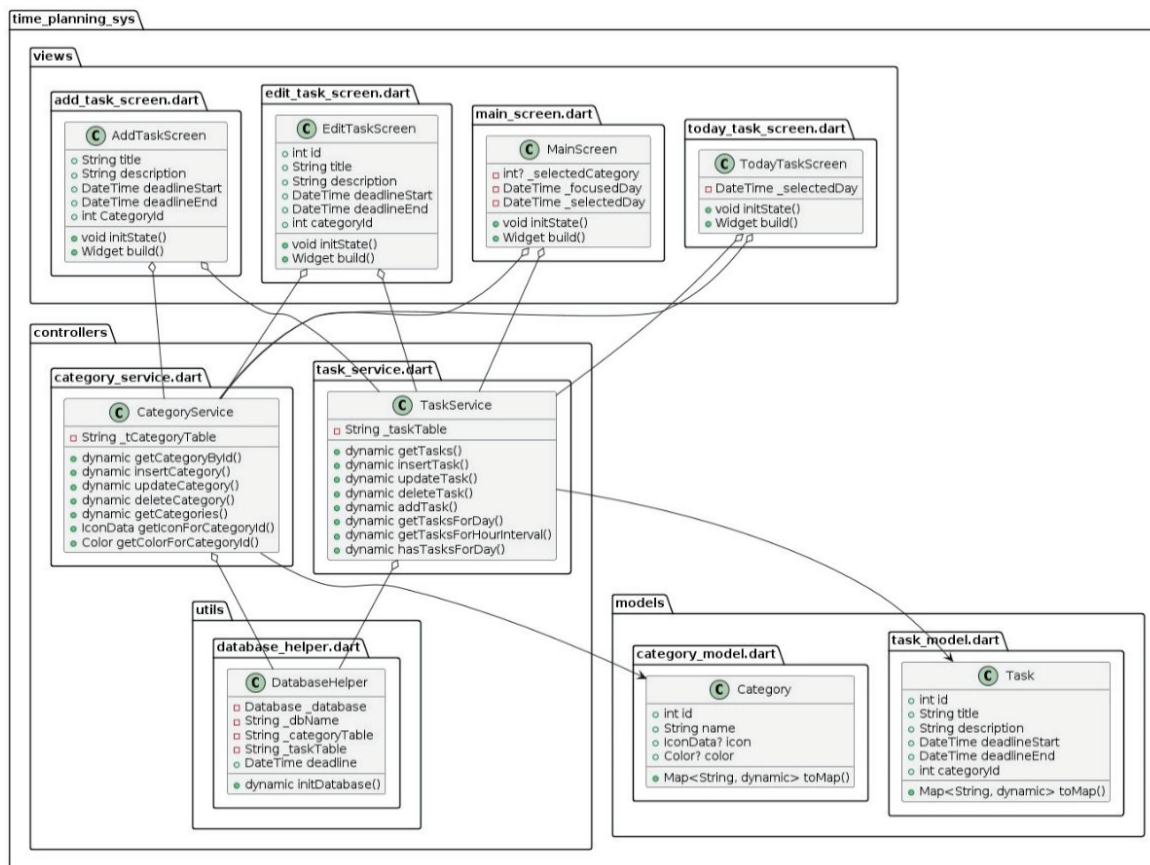


Рис. 5. Діаграма класів

Висновки

У межах дослідження здійснено архітектурне проектування навчально орієнтованої програмної системи тайм-менеджменту, що відповідає сучасним вимогам кросплатформності, автономності та конфіденційності. Обґрунтовано вибір архітектурного шаблону MVVM, який забезпечує чітке розділення відповідальностей між компонентами та сприяє масштабованості й тестованості системи. Для формалізації структури та поведінки системи побудовано низку UML-діаграм, що охоплюють як статичні, так і динамічні аспекти: діаграму варіантів використання, активностей, послідовності та класів. Логічна структура проекту організована відповідно до принципів clean architecture, що забезпечує ізоляцію функціональних зон та підтримує освітню придатність моделі. Завершальний розділ містить рекомендації щодо вибору програмних засобів, які забезпечують ефективну реалізацію системи планування часу з урахуванням вимог кросплатформності.

Запропонована модель є придатною для використання в освітньому процесі, зокрема для формування навичок архітектурного моделювання, системного мислення та реалізації повноцінних програмних продуктів. У подальшому система може бути розширена за рахунок інтеграції з хмарними сервісами, синхронізації з календарем користувача та реалізації мережевої взаємодії, що відкриває перспективи її застосування в побутовій та професійній діяльності.

Література

1. Савченко О. В. Тайм-менеджмент як інструмент підвищення ефективності навчальної діяльності студентів. *Педагогічний альманах*. 2021. № 49. С. 112–117.
2. The to-do list to organize work & life. *Todoist*. URL: <https://todoist.com>
3. Stay organized and manage your day. *Microsoft To Do*. URL: <https://to-do.microsoft.com>
4. Your all-in-one productivity platform. *Any.do*. URL: <https://www.any.do>
5. Глушков І. В. Методика навчального проектування в ІТ-освіті. *Інформаційні технології в освіті*. 2020. № 42. С. 45–52.
6. Model-View-ViewModel – .NET MAUI Architecture. *Microsoft Learn*. URL: <https://learn.microsoft.com/en-gb/dotnet/architecture/maui/mvvm>
7. Build apps for any screen. *Flutter*. URL: <https://flutter.dev>
8. A programming language optimized for UI. *Dart*. URL: <https://dart.dev>
9. Provider package for Flutter. URL: <https://pub.dev/packages/provider>
10. Flutter Local Notifications. URL: https://pub.dev/packages/flutter_local_notifications

11. Shared Preferences plugin. URL: https://pub.dev/packages/shared_preferences
12. Lightweight, embedded SQL database engine. *SQLite*. URL: <https://www.sqlite.org>

References

1. Savchenko O. V. (2021). Taim-menedzhment yak instrument pidvyshchennia efektyvnosti navchalnoi diialnosti studentiv [Time management as a tool for improving students' learning efficiency]. *Pedagogical Almanac*. 49. Pp. 112–117 [in Ukrainian].
2. The to-do list to organize work & life. *Todoist*. URL: <https://todoist.com>
3. Stay organized and manage your day. *Microsoft To Do*. URL: <https://to-do.microsoft.com>
4. Your all-in-one productivity platform. *Any.do*. URL: <https://www.any.do>
5. Hlushkov I. V. (2020). Metodyka navchalnoho proiektuvannia v IT-osviti [Methodology of educational design in IT education]. *Information Technologies in Education*. 42. Pp. 45–52 [in Ukrainian]
6. Model-View-ViewModel – .NET MAUI Architecture. *Microsoft Learn*. URL: <https://learn.microsoft.com/en-gb/dotnet/architecture/maui/mvvm>
7. Build apps for any screen. *Flutter*. URL: <https://flutter.dev>
8. A programming language optimized for UI. *Dart*. URL: <https://dart.dev>
9. Provider package for Flutter. URL: <https://pub.dev/packages/provider>
10. Flutter Local Notifications. URL: https://pub.dev/packages/flutter_local_notifications
11. Shared Preferences plugin. URL: https://pub.dev/packages/shared_preferences
12. Lightweight, embedded SQL database engine. *SQLite*. URL: <https://www.sqlite.org>

Дата першого надходження статті до видання: 24.10.2025

Дата прийняття статті до друку після рецензування: 25.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.415.2:004.382:004.75

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-07>

МЕТОД ДИНАМІЧНОГО БАЛАНСУВАННЯ ПРОДУКТИВНОСТІ СЕРВЕРІВ У ІНТЕРАКТИВНИХ СИСТЕМАХ

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 запропоновано метод динамічного балансування продуктивності серверів в інтерактивних системах, що функціонують в умовах непередбачуваних змін інтенсивності клієнтських звернень. Показано, що традиційні підходи до розподілу навантаження між серверами, орієнтовані на статичні або квазістатичні режими роботи, не забезпечують стабільної якості обслуговування користувачів при високій динамічності трафіка. Запропонований метод базується на принципах теорії автоматичного регулювання та використовує апарат аналітичного синтезу регуляторів для реалізації механізму авторегулювання у реальному часі. У рамках дослідження сформульовано оптимізаційну задачу вирівнювання поточних коефіцієнтів завантаження серверів шляхом мінімізації інтегрального квадратичного функціоналу, який враховує відхилення реальної траєкторії станів системи від оптимальної. Розроблено математичну модель процесу балансування, що представлена у вигляді системи диференціальних рівнянь із відповідними обмеженнями, які гарантують сталість роботи системи під час різких сплесків трафіка. Здійснено синтез регулюючого функціоналу, що поєднує функцію вирівнювання, функцію Ляпунова-Беллмана та механізм керування на основі рівняння Ріккати. Отримано рівняння настроювання у замкненій формі, яке описує процес динамічного вирівнювання навантаження між серверами. Доведено коректність роботи механізму балансування та показано, що запропонований метод забезпечує рівномірний розподіл навантаження, зменшення ризику перенавантаження окремих серверів, скорочення часу відгуку системи та підвищення ефективності використання обчислювальних ресурсів. Отримані результати можуть бути використані при розробленні адаптивних серверних платформ, хмарних систем оброблення запитів та інтелектуальних керуючих модулів для систем реального часу.

Ключові слова: інтерактивна система, динамічний режим, оптимальний розподіл потоків, сервер, балансування.

DYNAMIC SERVER PERFORMANCE BALANCING METHOD FOR INTERACTIVE SYSTEMS

Viktor Gnatyuk, Ph.D., Ass. Prof., State University “Kyiv Aviation Institute”; State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, 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 proposes a dynamic server performance balancing method for interactive systems operating under unpredictable variations in client request intensity. It is shown that traditional load distribution approaches designed for static or quasi-static operating modes fail to ensure stable quality of service under highly dynamic traffic conditions. The proposed method is based on the principles of automatic control theory and applies analytical regulator synthesis to implement self-regulation mechanisms in real time. The study formulates an optimization problem aimed at equalizing the current server load coefficients by minimizing an integral quadratic functional that accounts for deviations of the real system state trajectory from the optimal one. A mathematical model of the balancing process is developed in the form of a system of differential equations with constraints that guarantee system stability during traffic surges. The control functional is synthesized as a combination of a balancing function, the Lyapunov–Bellman function, and a control mechanism based on the Riccati equation. A closed-form tuning equation describing the dynamic equalization of server loads has been derived. The correctness of the balancing mechanism is proved, and it is demonstrated that the proposed method ensures uniform load distribution, reduces the risk of individual server overloads, shortens system response time, and increases the efficiency of computing resource utilization. The obtained results can be applied to the design of adaptive server platforms, cloud request-processing systems, and intelligent control modules for real-time applications.

Key words: interactive system, dynamic mode, optimal flow distribution, server, balancing.

Вступ

Сучасні інтерактивні системи характеризуються високою динамічністю трафіку та значною варіативністю навантаження на серверну інфраструктуру. Традиційні методи розподілу ресурсів між серверами, орієнтовані на статичні або слабкозмінні умови, не забезпечують ефективної роботи системи у реальному часі, що може призводити до перевантажень окремих серверів, збільшення часу обробки запитів та зниження якості обслуговування користувачів.

Сучасні підходи до адаптивного та інтелектуального керування ресурсами демонструють високу ефективність, однак більшість із них застосовуються або в умовах передбачуваних навантажень, або для конкретних типів серверних систем. В умовах інтерактивних систем з непостійною інтенсивністю клієнтських потоків залишається актуальним завдання розробки методів динамічного перерозподілу продуктивності серверів, що враховують поточний стан завантаженості кожного сервера та характер змін трафіку.

Отже, наукова проблема полягає у створенні методу адаптивного балансування продуктивності серверної лінійки, який забезпечував би оптимальне використання обчислювальних ресурсів, підвищував загальну ефективність роботи системи та підтримував стабільний рівень якості обслуговування при змінній інтенсивності клієнтських запитів.

Аналіз останніх джерел

Аналіз останніх наукових джерел у сфері управління динамічними системами та оптимізації продуктивності серверів свідчить про високий рівень розробок у галузі теорії автоматичного керування, адаптивних методів регулювання та сучасних алгоритмів оптимізації. Традиційні методи управління та теорія зворотного зв'язку. Класичні підходи до управління динамічними системами представлені у працях Kulmamirov та Smakova, Dorf та Bishop, Franklin et al. У цих джерелах детально розглядаються принципи побудови регуляторів, системи зворотного зв'язку, стійкість та адаптація параметрів. Вони формують теоретичну базу для розробки методів динамічного балансування продуктивності серверів, де необхідно забезпечити стабільність та ефективність реакції системи на змінні навантаження. Актуальним напрямом є використання PID-регуляторів та їх розширених модифікацій для управління продуктивністю складних систем (Aström & Hägglund). Такі методи дозволяють гнучко налаштовувати параметри регуляторів у реальному часі, що є критично важливим для інтерактивних серверних систем з непередбачуваним навантаженням. Ascher та Petzold пропонують ефективні чисельні методи для розв'язання диференціальних та диференціально-алгебраїчних рівнянь, що дозволяє моделювати динамічні процеси у серверних системах і прогнозувати поведінку системи при зміні навантаження. Це забезпечує основу для валідації та тестування алгоритмів балансування. Сучасні адаптивні та оптимізаційні методи. Ряд статей у IEEE (Wang et al.; Chen et al.; Zhuang et al.) присвячено методам адаптивного керування, підкріпленого навчанням (Reinforcement Learning) та ітераційного навчання для лінійних та нелінійних систем. Ці підходи дозволяють розробляти алгоритми, які самонавчаються та оптимізують продуктивність серверів у реальному часі, враховуючи змінні умови експлуатації та обмеження ресурсів. Застосування подібних методів до інтерактивних серверних систем може суттєво підвищити ефективність балансування навантаження, скоротити час відгуку та знизити ризик перевантажень.

Аналіз останніх джерел показує, що сучасні підходи до управління та оптимізації продуктивності систем поєднують класичну теорію управління, адаптивні регулятори та алгоритми машинного навчання. Для задач динамічного балансування серверів у інтерактивних системах найбільш перспективним є інтегрований підхід, що поєднує теоретичні основи стабільності та управління з сучасними методами адаптації та оптимізації на основі даних у реальному часі.

Метою статті є розробка та обґрунтування методу динамічного балансування продуктивності серверів інтерактивної системи, який: забезпечує адаптивний перерозподіл обчислювальних ресурсів між серверами залежно від поточного рівня їх завантаженості; враховує характер та динаміку змін інтенсивності клієнтських потоків; підвищує загальну ефективність серверної лінійки та стабільність обробки запитів користувачів; може бути реалізований для практичних інтерактивних систем з мінімальними втратами продуктивності та затримками у реагуванні на змінні умови навантаження.

Виконання досліджень

Моделювання процесу розподілу потоку звернень між серверами. Для досягнення поставленої цілі – оптимізувати роботу інтерактивної системи в умовах різних сплесків трафіка клієнтських звернень у напрямку підвищення її реактивності – у даному дослідженні запропоновано представляти інтерактивну систему як динамічну, яку слід розглядати у двох режимах – статичному (устойчивому) і динамічному (перехідному). У статичному режимі роботи системи, визначались оптимальні значення ємності буферної стекової пам'яті, що призначені для організації черг запитів, та кількості прикладних серверів, що здійснюють одночасну обробку клієнтських запитів, тобто оптимальні значення тих параметрів, котрі не змінюються у часі в процесі її функціонування. За результатами такої параметричної оптимізації була задана оптимальна траєкторія зміни станів інтерактивної системи у реальному часі відповідно до розробленої моделі обслуговування запитів як системи масового обслуговування (СМО), яка враховує сталі характеристики трафіка клієнтських запитів. Однак сплески трафіка звернень призводять до непередбачуваних відхилень реальної траєкторії зміни станів системи від оптимальної і за певних умов

можуть вивести систему із області її сталої роботи. Тому виникла потреба у мінімізації таких відхилень з тим, щоб додержувались критерії сталості інтерактивної системи.

Перш ніж здійснити моделювання процесу розподілу потоку звернень між серверами доцільно врахувати наступні міркування. Механізм динамічного розподілу потоку звернень між серверами має відтворювати у реальному часі процес вирівнювання навантаження на ці сервери. При цьому слід мати на увазі, що ні характеристики трафіка звернень, ні тривалість опрацювання звернень кожним із серверів не піддаються прогнозуванню. Саме тому із множини можливих методів розподілу доцільно обрати метод авторегулювання, що відслідковує тенденції щодо поточних рівнів завантаженості серверів і формує відповідний вектор керуючих впливів у напрямку вирівнювання ступеню їхнього навантаження. Отже, задачу керування розподілом потоку запитів між серверами на основі спостереження за рівнями їхнього завантаження у реальному часі доцільно представити як задачу динамічного вирівнювання коефіцієнтів завантаження цих серверів. Для її рішення, як показано далі, зручно використати результати теорії побудови аналітичних регуляторів, що широко висвітлені в багатьох публікаціях.

У теорії побудови аналітичних регуляторів при формалізації алгоритмів динамічного регулювання прийнято користуватися векторними формами запису змінних, що підлягають керуванню. У даному випадку це значення інтенсивності потоків клієнтських звернень, що просуваються на вводи кожного із серверів лінійки. Однак в якості керованих більш зручно узяти безрозмірні змінні – поточні значення коефіцієнтів навантаження на сервери лінійки $k_1, k_2, k_3, \dots, k_n$ або їхні зворотні значення $\eta_1, \eta_2, \eta_3, \dots, \eta_n$:

$$k_1 = X_1 / w_1, \quad k_2 = X_2 / w_2, \quad k_3 = X_3 / w_3, \quad \dots,$$

$$k_n = X_n / w_n; \quad \eta_1 = w_1 / X_1, \quad \eta_2 = w_2 / X_2, \quad \eta_3 = w_3 / X_3, \quad \dots, \quad \eta_n = w_n / X_n, \quad (1)$$

де $X_1, X_2, X_3, \dots, X_n$ – поточні значення інтенсивності пульсуючих потоків звернень від клієнтів відповідно на вході кожного із серверів лінійки; $w_1, w_2, w_3, \dots, w_n$ – продуктивність кожного із серверів лінійки; n – кількість серверів у лінійці.

Отже, вектор-стовпець поточних значень коефіцієнтів завантаження серверів:

$$\dot{\mathbf{k}} = \begin{pmatrix} k_1 \\ k_2 \\ \cdot \\ \cdot \\ \cdot \\ k_n \end{pmatrix} \quad (2)$$

показує рівні завантаження кожного із серверів лінійки в той час, як вектор-стовпець зворотних коефіцієнтів завантаження серверів

$$\dot{\boldsymbol{\eta}} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdot \\ \cdot \\ \cdot \\ \eta_n \end{pmatrix} \quad (3)$$

показує, наскільки пропускні спроможності серверів щодо обробки звернень перевищують поточні інтенсивності реальних потоків клієнтських звернень, що ними обслуговуються.

Так що, процес регулювання в узагальненому вигляді можливо моделювати наступними диференціальними рівняннями:

$$\begin{aligned} \dot{\mathbf{k}} &= \Upsilon_{\mathbf{k}}; \dot{\boldsymbol{\eta}} = \Upsilon_{\boldsymbol{\eta}}. \\ \mathbf{k}' &= \Upsilon_{\mathbf{k}}; \boldsymbol{\eta}' = \Upsilon_{\boldsymbol{\eta}}. \end{aligned} \quad (4)$$

Як бачимо, праві частини цих рівнянь відображають у неявному вигляді функціонал управління процесом регулювання, а ліві частини – вектор керованих змінних. Процес регулювання в даному випадку це є процес покрокового рішення одного із наведених вище рівнянь (4). Як наслідок, значення коефіцієнтів завантаження серверів (або їхніх зворотних значень) будуть поступово у реальному часі вирівнюватися шляхом відповідного перерозподілу часток загального потоку звернень між цими серверами за умови дотримання ряду обмежень, що наведені далі.

Зрозуміло, що для конкретного відображення процесу регулювання слід розкрити в явному вигляді праві частини рівнянь (4), зокрема обрати вид керування, задати вектор керуючих впливів та формалізувати функціональний взаємозв'язок між керованими змінними та керуючими впливами, а також задати обмеження, що пов'язані із забезпеченням сталого функціонування інтерактивної системи з урахуванням якості перехідних процесів, оскільки мова йде про динамічну систему.

Постановка задачі оптимального розподілу потоку звернень на обслуговування між серверами інтерактивної системи. У багатьох застосуваннях інтенсивність трафіка клієнтських звернень швидко змінюється у реальному часі непередбаченим чином. Тривалість обробки цих звернень кожним із серверів лінійки також є непередбачуваною величиною. В процесі роботи одні сервери можуть бути перенавантажені обробкою клієнтських звернень тоді, як інші недонавантажені. Зрозуміло, що бажано уникати таких ситуацій, інакше у результаті перенавантаження будь-якого із паралельно працюючих серверів звернення будуть втрачатися не зважаючи на те, що поточне значення загального коефіцієнту навантаження на лінійку серверів буде на відносно низькому рівні.

Таким чином, доцільною є постановка наступної оптимізаційної задачі: розробити метод динамічного перерозподілу загальної продуктивності лінійки серверів між її серверами в залежності від ступеню поточної їхньої завантаженості трафіком з урахуванням характеру змін інтенсивності потоків клієнтських звернень, що надходять до серверної частини обладнання інтерактивної системи. Тобто, ставиться задача вирівнювання поточних значень коефіцієнтів навантаження серверів лінійки між собою у динамічному режимі в реальних умовах їхнього застосування, коли мають місце різкі сплески трафіка звернень, що мають непередбачуваний характер.

Отже, представимо інтерактивну систему як динамічну систему із непередбачуваною траєкторією змін її стану у часі. Для кожної j -ої точки цієї траєкторії визначимо показник оптимальності у вигляді

$$\min K = (X_1 - s_1^0)^2 + (X_2 - s_2^0)^2 + \dots + (X_i - s_i^0)^2 + \dots + (X_n - s_n^0)^2, \quad (5)$$

де K – критерій оптимальності у задачі оптимізації, яка має враховувати невизначеність як у характері сплесків трафіку звернень, так і у тривалості їхньої обробки серверами; n – кількість серверів, що одночасно обробляють звернення клієнтів у j -ій точці траєкторії; $X_1, X_2, X_3, \dots, X_n$ – поточні значення інтенсивності пульсуючих потоків звернень від клієнтів на вході кожного із серверів у j -ій точці траєкторії; $s_1^0, s_2^0, s_3^0, \dots, s_n^0$ – значення інтенсивності потоків оброблених серверами звернень від клієнтів серверів у j -ій точці траєкторії в устояному (статичному) режимі, коли пульсації трафіка є відсутніми.

З фізичної точки зору вираз (5) означає, що у кожній точці траєкторії змін стану інтерактивної системи ставиться завдання мінімізації суми поточних відхилень реальної траєкторії змін стану від оптимальної траєкторії, що визначена для устояного режиму роботи системи.

Визначимо фізичні обмеження, що накладені на простір пошуку оптимального керування станом інтерактивної системи.

Обмеження 1. Сумарна інтенсивність реального трафіка звернень у будь-який момент часу має не перевищувати загальну продуктивність лінійки серверів. Тобто, має виконуватися наступне обмеження:

$$X_1 + X_2 + X_3 + \dots + X_n \leq \Delta W, \quad (6)$$

де ΔW – загальна продуктивність лінійки серверів.

$$\Delta W = w_1 + w_2 + w_3 + \dots + w_n = \text{const}, \quad (7)$$

де $w_1, w_2, w_3, \dots, w_n$ – продуктивність кожного із серверів лінійки.

Якщо обмеження (6) буде порушено, то виникнуть втрати звернень, які бажано не допускати, зокрема, шляхом підключення до процесу паралельної обробки звернень додаткових серверів, що знаходяться у гарячому резерві. Для цього треба мати інструмент вияву моменту часу, коли таке підключення треба здійснювати. Зокрема, це може бути момент переповнення черги на вході будь-якого із серверів лінійки або момент перевищення припустимого часу очікування у черзі будь-якого із звернень. В залежності від сфери застосування інтерактивної системи визначення цього показника здійснюється розробником системи. Отже, оптимальний розподіл має здійснюватися за умови, коли кількість серверів у складі лінійки та значення їхньої пропускну здатності є заданими величинами.

Обмеження 2. Слід увести наступне обмеження:

$$K_1^0 = K_2^0 = K_3^0 = \dots = K_n^0, \quad (8)$$

де $K_1^0, K_2^0, K_3^0, \dots, K_n^0$ – значення коефіцієнтів завантаження серверів лінійки в устояному режимі, що визначаються як

$$K_1^0 = s_1^0 / w_1, \quad K_2^0 = s_2^0 / w_2, \quad K_3^0 = s_3^0 / w_3, \dots, \quad K_n^0 = s_n^0 / w_n. \quad (9)$$

Критерій оптимальності K у даному випадку пов'язано з умовою забезпечення рівності поточних значень коефіцієнтів навантаження серверів в устояному режимі, коли інтенсивності запитів, що надходять до серверів, є постійними та не відчувають флуктуацій.

Зрозуміло, що діапазон припустимих значень коефіцієнтів завантаження кожного із серверів має бути в межах від 0 до 1.

Обмеження 3 щодо вибору механізму розподілу потоку звернень. Процес відслідковування має представлятися як ітеративний процес покрокового рішення у реальному часі відповідним чином синтезованого диференціального рівняння (4), в якості вихідних даних для якого використовуються виміряні на даному кроці інтегрування рівняння поточні значення інтенсивності оброблених потоків на кожному із серверів лінійки, а в якості результату рішення – значення інтенсивності увідних потоків звернень, що мають опрацьовуватися серверами на наступному кроці інтегрування. Синтезоване диференціальне рівняння забезпечить відстеження тенденцій у змінах інтенсивності оброблюваних потоків. Внаслідок інтегрування поточні зміни потоків будуть впливати на величини коефіцієнтів завантаження серверів.

З урахуванням вище зазначеного показник оптимальності має представлятися як:

$$\min K = (X_1 - s^p_1)^2 + (X_2 - s^p_2)^2 + \dots + (X_n - s^p_n)^2, \quad (10)$$

де $s^p_1, s^p_2, s^p_3, \dots, s^p_n$ – рекомендовані значення інтенсивності потоків оброблюваних серверами клієнтських звернень, що визначаються на кожному кроці інтегрування диференціального рівняння відслідковування відхилень реальної траєкторії змін у стані інтерактивної системи відносно траєкторії, що обчислена для устояного режиму її роботи.

Позитивний ефект від наближення до режиму однаковості значень коефіцієнтів навантаження серверів полягає у тому, що в процесі перерозподілу створюється ситуація, коли робота з обробки пульсацій трафіка рівномірно розподіляється між серверами. Це зменшує ризики перенавантаження будь-якого із них, якщо величини цих пульсацій знаходяться в межах запасу пропускної здатності інтерактивної системи.

Однак реально інтенсивності потоків звернень швидко змінюються непередбачуваним чином. Вони не можуть бути визначені та виміряні миттєво. Існує можливість оцінювання лише поточних тенденцій у змінах швидкостей потоків на коротких відрізках часу. Тому за цих умов механізм розподілу потоку звернень між серверами має спиратися не на вимірювані поточні значення інтенсивності потоків, а відслідковувати тенденції щодо напрямку змін інтенсивності цих потоків – у бік збільшення або зменшення.

Обмеження 4 щодо забезпечення сталої роботи інтерактивної системи в умовах різких сплесків трафіка звернень. Синтез диференціального рівняння (4), інтегрування якого у реальному часі забезпечує відслідковування відхилень реальної траєкторії змін станів інтерактивної системи від оптимальної траєкторії, має здійснюватися з урахуванням трьох факторів. По-перше, покрокове інтегрування цього рівняння має відтворювати процес відслідковування в напрямку мінімізації показника оптимальності (10). По-друге, відтворювана траєкторія змін у стані системи має знаходитись у зоні сталого функціонування цієї системи. По-третє, із множини можливих показників якості перехідних процесів, перш за все, слід враховувати швидкість перехідних процесів при зміні станів системи. Інші показники якості перехідних процесів у даному випадку суттєвої ролі не відіграють. Отже, необхідно синтезувати функціонал, що обмежує область припустимих рішень щодо керування механізмом перерозподілу трафіку звернень між серверами. Цей функціонал має зв'язувати керовані параметри динамічної системи з керуючими впливами механізму перерозподілу трафіка таким чином, щоб забезпечувалось виконання умов обраного критерію сталості та контрольована швидкість перехідних процесів.

Слід сподіватися, що у рамках визначених обмежень результатом рішення даної оптимізаційної задачі буде зменшення ймовірності перенавантаження серверів трафіком і, як наслідок, зменшення величини коефіцієнту втрат необроблених клієнтських звернень.

Теорія автоматичного регулювання (ТАР) пропонує кілька рівноцінних методів рішення для класу задач, до якого відноситься поставлена задача оптимального розподілу потоку звернень на обслуговування між серверами інтерактивної системи. Зокрема, дану задачу можливо вирішити за допомогою класичного варіаційного числення, методом принципу максимуму Л. С. Понтрягіна або методом динамічного програмування. Проте, найбільш прийнятним у даному випадку є метод, що дозволяє отримати рішення у замкнутому вигляді, тобто у вигляді однозначно визначеного рівняння настроювання. Тому у даній роботі узгодження усіх параметрів системи для забезпечення її сталої роботи в оптимальному режимі згідно з критерієм (10) здійснено шляхом використання функціоналу, формалізований вигляд якого представлений далі. Процедура формального синтезу цього функціоналу розглянута у багатьох роботах, що присвячені проблемам керування динамічними об'єктами, і не потребує додаткового опрацювання. Завдання даної роботи – обґрунтувати вибір одного із відомих видів цього функціоналу в якості основи для розроблюваного методу рішення задачі оптимального розподілу потоку звернень.

Синтез методу рішення задачі оптимального розподілу потоку звернень між серверами інтерактивної системи. Даний метод передбачає наступний порядок рішення задачі оптимального розподілу потоку звернень: вибір регулюючого функціоналу; синтез матриці регулюючих зв'язків для функції вирівнювання; синтез функції вирівнювання; обґрунтування виду рівняння настроювання; визначення вектору керування; синтез рівняння настроювання у замкненій формі; формальний доказ коректності роботи механізму вирівнювання.

Вибір виду регулятора для розподілу потоку звернень. У даному випадку синтез методу рішення задачі оптимального розподілу полягає у виборі виду типового регулятора з відомими параметрами та у визначенні оптимальних з точки зору обраних показників якості значень цих параметрів. Тобто, спочатку в рамках даного методу здійснюється обґрунтування вибору оптимальної структури (закону регулювання) типового регулятора для заданого об'єкту керування та обраного критерія оптимальності, а потім вирішується задача параметричного синтезу обраної структури.

Існує велика кількість методів побудови регуляторів для керування об'єктами різноманітного призначення щодо яких висувається широкий спектр вимог, котрі враховують конкретні умови функціонування цих об'єктів. Відповідні класифікатори цих методів широко представлені в наукових публікаціях. Усю множину методів побудови регуляторів доцільно розділити на дві великі групи. До першої групи відносять чисто експериментальні методи визначення параметрів регулятора, коли із евристичних міркувань задають його базову структуру і далі здійснюють ітеративний підбір бажаних значень параметрів цієї структури методом проб та помилок. Зокрема, спостерігають, як поводить себе регулятор з експериментально визначеними коефіцієнтами регулювання, і в залежності від того, що має місце – недорегулювання чи перерегулювання – змінюють коефіцієнти та повторюють експеримент. При цьому користуються відомими методиками підбору коефіцієнтів, що створені за методом Зіглера-Ніколса. Зрозуміло, що ці методи не потребують детальної аналітичної проробки задачі регулювання і не можуть претендувати на оптимальність отриманих рішень, але у багатьох випадках можуть слугувати основою для первісного налагоджування регулятора. До другої групи відносять методи синтезу регуляторів на основі побудованих моделей системи управління об'єктом регулювання. Ці методи дозволяють більш точно розрахувати параметри регулятора і в багатьох випадках отримати оптимальні рішення з точки зору обраних критеріїв якості його роботи. У свою чергу, методи другої групи щодо математичного опису моделей систем регулювання розділяють на частотні та часові.

Лінійні та нелінійні системи. На практиці більшість об'єктів регулювання за своєю природою є багатовимірними та нелінійними. Особливо, якщо розглядати ці об'єкти у широкому діапазоні змін режимів їхньої роботи та в умовах дії значних збурень. Вичерпний аналіз нелінійних систем регулювання надано у працях з теорії нелінійних систем автоматичного регулювання та управління. Зроблено висновок, що формальний синтез регуляторів для таких об'єктів у загальному випадку не уявляється можливим, а використання лінеаризованих моделей при синтезі регуляторів не завжди дозволяє отримати задані показники сталості та точності їхньої роботи. Більш пізні роботи з результатами досліджень нелінійних, стохастичних, робастних та адаптивних систем зроблений вище висновок лише підтверджують.

Особливістю задачі оптимального розподілу потоку звернень є те, що на практиці у більшості випадків спостерігаються відносно невеликі відхилення траєкторії зміни станів об'єкту регулювання від заданого режиму роботи. Окрім того, вплив такого явища як запізнення реакції на клієнтські запити (внаслідок кінцевої швидкості обробки інформаційних потоків) на сталість та якість роботи більшості прикладних систем, в даному випадку не є суттєвим. З цього витікає можливість вважати шукану систему регулювання лінійною. Так що, будемо вважати лінійними як об'єкт регулювання, так і регулятор. А аналітичне конструювання регулятора будемо здійснювати в класі задач оптимального лінійного регулювання.

Вибір регулюючого функціоналу. Даний метод рішення задачі оптимального розподілу потоку звернень передбачає необхідність вибору математичного апарату для формалізації процесу такого розподілу.

Зокрема, якщо користуватися термінами теорії авторегулювання, метод передбачає необхідність вибору виду керування. Тобто, необхідно задати функціонал, що оптимізує у реальному часі сталу роботу замкненої системи авторегулювання в заданих обмеженнях щодо обраних показників якості перехідних процесів. Цей функціонал має задавати на формальному рівні вид функціональних зв'язків між керуючими впливами та керованими змінними з урахуванням визначених вище обмежень, а також структуру та алгоритм роботи регулятора (механізму керування), що задає оптимальну траєкторію зміни стану інтерактивної системи із додержанням критеріїв сталості її функціонування. Умовою дотримання режиму узгодженої роботи усіх елементів системи авторегулювання має бути забезпечення мінімальних значень обраного виду функціоналу. Цей функціонал має задавати також вид рівняння відслідковування тренду змін у стані інтерактивної системи.

У даному випадку вибір виду оптимізуючого функціоналу здійснено виходячи із наступних міркувань. По-перше, необхідно, щоб у підінтегральному виразі функціоналу містилась складова, що має відображати квадратичні відхилення поточних значень інтенсивності потоку звернень від їх устояних значень згідно виразу (10). По-друге, необхідно, щоб функціонал мав вигляд інтегрального критерію, оскільки використання інтегральних критеріїв дозволяє відобразити закон управління у замкнутій формі та задати вимоги до якості перехідних процесів таким чином, щоб керована система не втратила асимптотичну сталість. Зокрема, шляхом майже довільного вибору початкових значень відносно невеликої кількості вагових коефіцієнтів при функції регулювання забезпечується можливість здійснювати цілеспрямований ітераційний перебір значень вагових коефіцієнтів в області сталого режиму роботи системи аж доки не буде досягнута задана якість перехідних процесів. По-третє, інтегральні функціонали зайшли широке відображення в теорії оптимального управління, так що даний спосіб формалізації дозволяє підібрати вид типового регулятора з відомими характеристиками, що є оптимальним для розподілу потоку звернень між серверами інтерактивної системи. Застосування типового регулятора дозволяє позбавитись необхідності здійснення розрахункових робіт з метою визначення елементів його структури.

Якщо вважати, що система регулювання забезпечує можливість спостереження за змінами вектору стану системи, то сутність рішення задачі оптимального лінійного регулювання на основі інтегрального критерію оптимізації полягає в наступному.

У класичній постановці цієї задачі лінійну систему задають двома векторними диференціальними рівняннями:

$$\begin{aligned}\dot{x} &= A(t) x(t) + B(t) u(t); \\ y(t) &= C^T(t) x(t),\end{aligned}\quad (11)$$

де $x(t_0) = x^0$; $A(t)$, $B(t)$, $C(t)$ – матричні функції часу розмірністю $(n \times n)$, $(n \times m)$, $(n \times p)$ відповідно; $x(t)$ – вектор стану системи, $u(t)$ – вектор керуючих впливів, $y(t)$ – вектор виміряних значень змінних, що визначає поточний стан системи.

Необхідно знайти вектор $u(t)$, що забезпечує перехід системи протягом проміжку часу $b = [t_0, t_1]$ з початкового стану $x(t_0) \neq 0$ в кінцевий стан $x(t_1) = 0$.

Як показник якості регулювання використовують інтегральний квадратичний критерій наступного виду:

$$I[x(t_0), u(t), t_0] = \int_{t_0}^{t_1} [x^T Q x + u^T R u] dt + x^T(t_1) G x(t_1), \quad (12)$$

де Q , R , G – матриці масштабування, фізичний зміст котрих буде роз'яснено далі.

Як бачимо із (12), значення функціоналу залежить не тільки від вектору керуючих впливів $u(t)$, але і від початкового стану $x(t_0)$, що не є обмежуючою умовою для роботи механізму перерозподілу потоку звернень. У цьому випадку більш реалістичним є інтегральний квадратичний критерій, що не залежить від початкового стану системи та часу її переходу в кінцевий стан, а інтервал оптимізації є нескінченним, а саме:

$$I[u(t)] = \int_0^{\infty} [x^T(t) Q(t) x(t) + u^T(t) R(t) u(t)] dt, \quad (13)$$

де $Q(t)$, $R(t)$ – позитивно визначені матриці.

Перша складова підінтегрального виразу відображає механізм регулювання, а друга – механізм керування. Доведено [6]: якщо система регулювання є повністю керованою у будь-який момент часу t , то оптимальне рішення щодо вектора керування має наступний вигляд:

$$u^*(t) = -R^{-1}(t) B^T(t) \left[\lim_{T \rightarrow \infty} P(t, t_1) \right] x(t). \quad (14)$$

У виразі (14) $P(t, t_1)$ є позитивно визначеним рішенням допоміжного рівняння Ріккати, якщо виконується обмежувальна умова $P(t_1, t_1) = 0$.

Рівняння Ріккати – це матричне нелінійне алгебраїчне рівняння із змінними коефіцієнтами, що має наступний вигляд:

$$- \dot{P} = PA + A^T P - PBR^{-1}B^T P + Q. \quad (15)$$

Вперше рівняння Ріккати було використано А.М. Летовим для побудови лінійного регулятора. Доведено, що рівняння (15) має стале єдино можливе рішення, якщо регульований об'єкт, що визначається матрицями A і B , знаходиться під керуванням, а симетричні матриці Q та R є позитивно визначеними. У нашому випадку для задачі оптимального розподілу потоку звернень названі вище умови виконуються, оскільки поточні значення коефіцієнтів завантаження паралельно працюючих серверів у процесі регулювання прагнуть до їхніх сталих значень.

При векторі керування, що визначається виразом (14), мінімальне значення функціоналу (15) буде представлятися наступним виразом:

$$I^*[x(t_0), t_0] = x^T(t_0) \left[\lim_{T \rightarrow \infty} P(t_0, t_0) \right] x(t_0). \quad (16)$$

Слід звернути увагу на те, що мінімальному значенню функціонала (16) відповідає широка область оптимальних траєкторій змін стану системи (11), оскільки не враховуються обмеження на керування механізмом регулювання. Найбільш суттєвим є те, що вираз (13) не враховує обмеження, що пов'язане із забезпеченням необхідного рівню якості перехідних процесів. Щоб інтегральний критерій відображав сталу роботу системи регулювання із заданими показниками якості перехідних процесів, необхідно, у підінтегральний вираз функціоналу (13) додати третю складову, що синтезована на основі застосування функції Ляпунова-Беллмана. Синтез виду шуканого функціоналу, що враховує показник якості перехідного процесу і дозволяє визначити вектор керування, котрий забезпечує оптимальне регулювання в межах заданого значення цього показника, здійснено наступним чином.

В теорії побудови аналітичних регуляторів для динамічних систем широко використовуються квадратичні матричні форми та операція транспонування матриць. Для формального представлення взаємозв'язків між елементами системи авторегулювання зазвичай утворюють квадратичну форму у вигляді:

$$K^T S \cdot Y \cdot S^T K, \quad (17)$$

де K – вектор-стовпець керованих змінних; S – матриця регулюючих зв'язків між керуючими впливами та керованими змінними; Y – матриця масштабування керуючих впливів; T – операція транспонування матриці.

Для нашого випадку доцільно використати метод побудови системи авторегулювання. Цей метод орієнтовано на виконання умов як сталості, так і якості перехідних процесів. Відповідно до цього методу функціонал, що має зв'язувати усі елементи механізму відслідковування сплесків трафіка звернень і задавати оптимальну траєкторію змін у стані інтерактивної системи з урахуванням визначених вище обмежень, представляється як:

$$\min \Phi = \int_0^{\infty} (K^T S \cdot W_R \cdot S^T K + \zeta K^T S \cdot Q \cdot S^T K + U^T R U) dt, \quad (18)$$

де $W_R - m \times m$ – мірна квадратна позитивно визначена симетрична матриця вагових коефіцієнтів, яка задає рівень «важливості» кожній парі регулюючих зв'язків між керованими змінними;

$Q - m \times m$ – мірна позитивно визначена симетрична матриця квадратичної форми, що є складовим елементом функції Беллмана;

ζ – позитивна константа (множник при функції Беллмана), що визначає ступінь загасання функції Беллмана;

$U - m$ – мірний вектор керуючих впливів;

$R - m \times m$ – мірна позитивно визначена симетрична матриця вагових коефіцієнтів, яка задає рівень «важливості» кожного керуючого впливу;

T – символ операції транспонування матриці.

Як бачимо із (18), підінтегральний вираз функціоналу Φ на відміну від функціоналу (13) складається із трьох членів.

Оптимізація процесу авторегулювання полягає у виборі параметрів механізму вирівнювання навантаження на сервери, за яких значення цього функціоналу будуть мінімізовані.

Перша складова підінтегрального виразу Φ_1 – це формальне відображення функції вирівнювання. Φ_1 представляє у неявному вигляді роботу механізму вирівнювання керованих змінних між собою (тобто, коефіцієнтів навантаження на сервери лінійки), а саме:

$$\Phi_1 = K^T S \cdot W_R \cdot S^T K. \quad (19)$$

Як бачимо, функція вирівнювання згідно (10) являє собою зважену суму квадратів відмінків (через коефіцієнти матриці W_R) керованих змінних. Чим більша ця сума, тим буде більше значення функціоналу Φ . Механізм вирівнювання має діяти у напрямку поступового зменшення цієї суми. При цьому слід мати на увазі, що чим більше значення функціоналу, тим довше за часом здійснюється процес вирівнювання.

Друга складова підінтегрального виразу – це є функція Ляпунова-Беллмана Φ_2 , яка визначає швидкість дії механізму вирівнювання. Вона уведена у функціонал Φ із заданим показником загасання ζ . Функція Ляпунова-Беллмана на оптимальних траєкторіях зміни станів динамічної системи убуває із швидкістю не меншій за здобуток її самої та показника загасання ζ . Тому, шляхом вибору значень показника ζ є можливим задати швидкодію оптимальної системи авторегулювання (точніше, задати час існування перехідного процесу) за якої ця система буде замкнутою та сталою. Формальне

представлення функції Ляпунова-Беллмана Φ_2 у вигляді квадратичної форми в даному випадку співпадає із представленням функції вирівнювання, а саме:

$$\Phi_2 = K^T S \cdot Q \cdot S^T K. \quad (20)$$

Функцію Ляпунова-Беллмана можливо трактувати як поточну відстань між збуреним станом замкненої системи регулювання та незбуреним, що є характерним для статичного режиму її роботи. Динаміка перехідного процесу тим вища, чим швидше убуває ця відстань. Отже, експериментальним шляхом підбираючи значення показника загасання ζ , є можливим задати прийнятну для даних конкретних умов використання інтерактивної системи швидкість убування функції Ляпунова-Беллмана уздовж траєкторії змін стану замкненої системи регулювання.

Третя складова підінтегрального виразу Φ_3 – це формальне представлення роботи аналітичного регулятора, а саме:

$$\Phi_3 = U^T R U. \quad (21)$$

Його уведення у підінтегральний вираз функціоналу (18) дозволяє визначити керування та формально замкнути ланцюг авторегулювання, тобто представити систему регулювання у вигляді одного математичного виразу.

Синтез матриці регулюючих зв'язків для функції вирівнювання. Після вибору регулюючого функціоналу згідно з методом оптимального розподілу потоку звернень між серверами інтерактивної системи необхідно сформулювати матрицю регулюючих зв'язків S для функції вирівнювання Φ_1 , яка задає порядок перерозподілу керуючих впливів між змінними під час регулювання.

Матриця повинна відображати можливість перерозподілу навантаження між усіма серверами лінійки, тобто встановлювати функціональні зв'язки між будь-якими двома серверами. Кількість стовпців у матриці дорівнює числу змінних, що підлягають керуванню (кількість паралельно працюючих серверів), а кількість рядків – числу унікальних пар серверів. Для зручності сервери нумеруються послідовно, щоб кожний рядок матриці характеризував взаємодію певної пари серверів; порядок розташування пар у таблиці не має значення.

Матриця описує не абсолютні значення потоків звернень, а напрямки змін навантаження. Знак «+» означає збільшення потоку на сервер, «-» – його зменшення, «0» – відсутність взаємодії між відповідною парою. Таким чином, кожний рядок матриці інтерпретується як передача частини ресурсу від сервера зі знаком «-1» до сервера зі знаком «1», тоді як сервери з позначкою «0» у цьому процесі участі не беруть.

Якщо лінійка складається із чотирьох серверів, то маємо чотири змінні, що потребують керування – k_1, k_2, k_3, k_4 . Якщо лінійка складається із п'яти серверів, то маємо п'ять змінних – k_1, k_2, k_3, k_4, k_5 . Тоді для відображення взаємозв'язків між чотирма серверами матриця S_4 буде складатися із шести рядків, а для відображення взаємозв'язків між п'ятьма серверами матриця S_5 буде складатися із десяти рядків:

$$S_4 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}; S_5 = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}. \quad (22)$$

Фізичний зміст, зокрема, другого рядка матриці S_4 відображає ситуацію, коли частка потоку звернень передається від першого сервера на другий сервер. Третій рядок цієї матриці відображає передачу частки потоку звернень від другого сервера на третій. Аналогічно відображається перерозподіл ресурсу і у матриці S_5 . Зокрема, третій рядок S_5 відповідає передачі частки потоку з другого сервера на третій. Порядок розташування рядків або стовпців у матрицях значення не має. Стовпці і рядки можуть мінятися місцями.

Синтез функції вирівнювання. Після того, як в явному вигляді задана матриця регулюючих зв'язків S згідно запропонованого методу рішення задачі оптимального розподілу потоку звернень, необхідно визначити функцію вирівнювання в явному вигляді. Ця функція Φ_1 в неявному вигляді представляється першою складовою регулюючого функціоналу (18), тобто виразом (19). Як бачимо, для

формального відображення функції вирівнювання використана квадратична форма, у котрій змінні, що потребують регулювання, задані у вигляді вектор-стовпця поточних значень коефіцієнтів завантаження $\mathbf{K}$ (це вираз (2)), а механізм перерозподілу керуючих впливів між змінними в процесі регулювання задано у вигляді матриці регулюючих зв'язків $\mathbf{S}$ (це вираз (22)).

Спочатку в явному вигляді представимо перший елемент Φ_1 , а саме добуток $\mathbf{K}^T \mathbf{S}$, тобто добуток між матрицею регулюючих зв'язків $\mathbf{S}$ та транспонованим вектором керованих змінних $\mathbf{K}$. Аналогічним чином представляється третій елемент Φ_1 , тобто добуток $\mathbf{S}^T \mathbf{K}$. Другим елементом Φ_1 є матриця вагових коефіцієнтів $\mathbf{W}_R$, яка задає «вагу» кожному регулюючому зв'язку між керованими змінними. Ця матриця в явному вигляді представляється як позитивно визначена діагональна матриця:

$$\mathbf{W}_R = \begin{pmatrix} w_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & w_{33} & 0 & 0 & 0 \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & w_{nn} \end{pmatrix}. \quad (23)$$

При побудові системи розподілу потоку звернень «вагу» кожного регулюючого зв'язку доцільно прирівняти до 1, так що

$$w_{11} = w_{22} = w_{33} = \dots = w_{nn} = 1. \quad (24)$$

Якщо маємо лінійку із чотирьох серверів, то перший елемент добутка Φ_1 в явному вигляді представиться як

$$\mathbf{K}^T \mathbf{S}_4 = (k_1, k_2, k_3, k_4) \cdot \begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} = (k_1 - k_2, k_3 - k_1, k_2 - k_3, k_1 - k_4, k_4 - k_2, k_3 - k_4). \quad (25)$$

Як бачимо, добуток (25) є транспонованим вектором чотирьох керованих змінних, тобто вектором-рядком. Кожен компонент вектору-рядка представляється як відмінок між значеннями відповідної пари керованих змінних, що складають вектор $\mathbf{K}$.

Третій елемент Φ_1 , а саме добуток $\mathbf{K}^T \mathbf{S}$, в явному вигляді представиться як

$$\mathbf{S}^T \mathbf{K} = \begin{pmatrix} k_1 - k_2 \\ k_3 - k_1 \\ k_2 - k_3 \\ k_1 - k_4 \\ k_4 - k_2 \\ k_3 - k_4 \end{pmatrix}. \quad (26)$$

Як бачимо, добуток (26) є вектором-стовпцем чотирьох керованих змінних. Кожен компонент вектору-стовпця представляється як відмінок між значеннями відповідної пари керованих змінних, що складають вектор $\mathbf{K}$.

У нашому прикладі, коли маємо чотири керовані змінні, матриця вагових коефіцієнтів $\mathbf{W}_R$ в явному вигляді представляється як

$$\mathbf{W}_R = \begin{pmatrix} w_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & w_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{66} \end{pmatrix}. \quad (27)$$

$$w_{11} = w_{22} = w_{33} = \dots = w_{nn} = 1. \quad (28)$$

Підставивши (25), (26) та (27) в (19) отримаємо в явному вигляді кінцевий вираз для функції вирівнювання:

$$\begin{aligned} K^T S \cdot W_R \cdot S^T K = p_{11}(k_1 - k_2)^2 + p_{22}(k_3 - k_1)^2 + p_{33}(k_2 - k_3)^2 + \\ p_{44}(k_1 - k_4)^2 + p_{55}(k_4 - k_2)^2 + p_{66}(k_3 - k_4)^2. \end{aligned} \quad (29)$$

Порівнявши вираз (5) з отриманим виразом (29), можливо зробити висновок, що цей вираз являє зважену суму квадратів відмінків між усіма можливими значеннями керованих змінних і може бути використаний в якості показника оптимальності процесу вирівнювання навантаження на сервери, що оброблюють звернення клієнтів інтерактивної системи. Так що, вираз (29) і визначає функцію вирівнювання.

Обґрунтування виду рівняння настроювання. Вище було показано, що задача динамічного настроювання регулятора зводиться до задачі вирівнювання коефіцієнтів завантаження пошукових серверів. Якщо в якості керованих змінних узяті поточні значення коефіцієнтів навантаження на сервери лінійки $k_1, k_2, k_3, \dots, k_n$, то процес вирівнювання має здійснюватися таким чином, щоб під час його роботи сума керованих змінних була постійною величиною. Так що, поведження даної системи регулювання доцільно підкорити рівнянню настроювання у наступному виді:

$$K = SU, \quad (30)$$

де K – вектор керованих змінних, S – матриця регулюючих зв'язків, а U – вектор керуючих впливів, який у даному випадку має визначатися як лінійна функція компонентів вектору K .

Визначення вектору керування. Аналіз функціоналу показує, що вектор керування має залежати від функції вирівнювання Φ_1 і функції Ляпунова-Беллмана Φ_2 . Функція Φ_1 в явному вигляді представлена виразом (29), а функція Φ_2 у вигляді квадратичної форми (20). Отже, першим кроком визначення вектору керування має бути отримання виразу для функції Ляпунова-Беллмана в явному вигляді. Для цього використовують допоміжне рівняння Беллмана, що має наступний вигляд:

$$\min_u (K^T S W_R S^T K + \zeta K^T S Q S^T K + U^T R U + \dot{K}^T (\partial V / \partial K) + (\partial V / \partial K^T) \dot{K}) = 0. \quad (31)$$

Якщо підставити у рівняння Беллмана (31) рівняння настроювання (30), то одержимо

$$\min_u (K^T S W_R S^T K + \zeta K^T S Q S^T K - U^T R U - U^T S^T S Q S^T K - K^T S Q S^T S U) = 0. \quad (32)$$

Для того, щоб отримати кінцевий вираз для вектору керування, у наступному кроці визначення матриці Q необхідно про диференціювати по вектору керування вираз (32) і результат прирівняти нулеві. У результаті вектор керування відобразиться як

$$\begin{aligned} U &= -R^{-1} S^T S Q S^T K; \\ U^T &= -K^T S Q S^T S R^{-1}. \end{aligned} \quad (33)$$

Усі складові елементи виразів (33), окрім матриці Q , що входить до квадратичної форми функції Беллмана, мають математичне відображення у явному вигляді. Для визначення цієї матриці реалізують наступну відому процедуру. Спочатку у рівняння Беллмана (31) підставляють вираз (33) і отримують наступне рівняння:

$$\begin{aligned} K^T S \cdot W_R \cdot S^T K + \zeta K^T S Q S^T K + K^T S Q S^T S R^{-1} R R^{-1} S^T S Q S^T K - \\ - K^T S Q S^T S R^{-1} R R^{-1} S^T S Q S^T K - K^T S Q S^T S R^{-1} R R^{-1} S^T S Q S^T K = 0. \end{aligned} \quad (34)$$

Потім після спрощень і винесення за дужки однакових лівих та правих множників у виразі (34) отримують наступне рівняння:

$$K^T S \cdot (W_R + \zeta Q - Q S^T S R^{-1} S^T S Q) \cdot S^T K = 0. \quad (35)$$

Оскільки рівність (35) є справедливою для будь-яких значень винесених за дужки множників, то вираз у дужках має дорівнювати нулеві. Так що,

$$W_R + \zeta Q - Q S^T S R^{-1} S^T S Q = 0. \quad (36)$$

Якщо порівняти отриманий вираз із (15), то можливо зробити висновок, що вираз (36) являє собою рівняння Ріккати, тобто матричне квадратне рівняння, що використовують для пошуку матриці функції Беллмана. Існує декілька ефективних обчислювальних алгоритмів для його рішення. У результаті рішення рівняння (36) отримують матрицю Q в явному вигляді. Якщо підставити знайдену матрицю у вираз (33), то одержимо в явному вигляді усі елементи вектору керування.

Синтез рівняння настроювання у замкненій формі. Для того, щоб отримати вираз для системи вирівнювання у замкненій формі, достатньо підставити знайдений вектор керування (33) у рівняння настроювання (30). Отже, замкнута система вирівнювання поточних значень коефіцієнтів завантаженості пошукових серверів визначається наступним виразом:

$$K = -SR^{-1}S^T S Q S^T K. \quad (37)$$

Таким чином, отримані вище вирази являють формальний опис механізму вирівнювання навантаження між паралельно працюючими серверами лінійки пошукових серверів. Для завершення формального опису слід навести доказ, що при будь-яких початкових значеннях вектору керування змінних в системі (37) досягається ситуація, коли стає значення керованого вектору є таким, що його компоненти є однаковими, а їхня сума протягом усього процесу вирівнювання є незмінною.

Формальний доказ коректності роботи механізму вирівнювання. Для того, щоб синтезоване рівняння настроювання (37) при будь-якому довільному векторному керуванні забезпечило рішення задачі вирівнювання керованих перемінних, необхідно, щоб вектор керування $\mathbf{U}$ виглядав як лінійна функція компонентів вектору $\mathbf{K}$, а матриця регулюючих зв'язків $\mathbf{S}$ відслідковувала тенденції щодо поточних рівнів завантаженості пошукових серверів.

Доказ того, що за наведених вище умов система вирівнювання забезпечує під час настроювання збереження суми компонент вектору $\mathbf{K}$, міститься, зокрема, в [13–15] на прикладі системи регулювання другого порядку. Доказ для систем настроювання більш високого порядку здійснюється аналогічно.

Для обраного прикладу система настроювання у розгорнутому вигляді представлялась як

$$\begin{pmatrix} \dot{k}_1 \\ \dot{k}_2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{U}. \quad (38)$$

А вектор керування для (38) як лінійна функція керованих змінних у вигляді

$$\mathbf{U} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \cdot \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}. \quad (39)$$

де w_{11} , w_{12} , w_{21} , w_{22} – вагові коефіцієнти масштабування матриці $\mathbf{W}_R$, які задають «вагу» кожному регулюючому зв'язку між керованими змінними.

Завдання полягало в тому, щоб визначити конкретні значення цих коефіцієнтів і упевнитись, що похідні від керованих перемінних будуть однакові за величиною і протилежні за знаком. У цьому разі їхня сума буде зберігатися.

Отже, з урахуванням (39) рівняння настроювання у замкненій формі буде мати наступний вигляд:

$$\begin{pmatrix} \dot{k}_1 \\ \dot{k}_2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{W}_{11}\mathbf{W}_{12} \\ \mathbf{W}_{21}\mathbf{W}_{22} \end{pmatrix} \cdot \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}. \quad (40)$$

А у скалярному записі система (40) матиме наступний вигляд:

$$\dot{k}_1 = -(w_{11} k_1 + w_{12} k_2) + (w_{21} k_1 + w_{22} k_2); \quad \dot{k}_2 = (w_{11} k_1 + w_{12} k_2) - (w_{21} k_1 + w_{22} k_2). \quad (41)$$

Далі для даної системи настроювання призначається стає допоміжне рівняння у вигляді

$$\dot{V} + cV \leq 0, \quad (42)$$

де V – допоміжна функція, що уведена у вигляді квадратичної форми

$$V = q (k_1 - k_2)^2. \quad (43)$$

а q – позитивний скалярний коефіцієнт.

У виразі (42) c – показник загасання допоміжної функції.

Похідна за часом від допоміжної функції V має наступний вигляд:

$$\dot{V} = 2q (k_1 - k_2) (\dot{k}_1 - \dot{k}_2). \quad (44)$$

Далі визначають похідні від керованих змінних з рівняння (41) і отриманий результат підставляють в допоміжне рівняння (42). І після спрощень одержують наступний вираз:

$$4 (w_{21} - w_{11}) k_1 + 4 (w_{22} - w_{12}) k_2 + ck_1 - ck_2 = 0. \quad (45)$$

Дорівнюють коефіцієнти при керованих змінних і з виразу (45) отримують систему рівнянь, яка дозволяє визначити «вагу» кожного регулюючого зв'язку між керованими змінними:

$$w_{21} - w_{11} = -(c/4); \quad w_{22} - w_{12} = c/4. \quad (46)$$

Отже, отримано два рівняння (46) для визначення чотирьох «вагових» коефіцієнтів, значення двом із котрих призначають довільно. Зокрема, у такий спосіб:

$$w_{11} = w_{22} = 1. \quad (47)$$

Тоді із системи рівнянь (46) з урахуванням (47) одержують значення двох інших «вагових» коефіцієнтів:

$$w_{21} = 1 - (c/4); \quad w_{12} = 1 - (c/4). \quad (48)$$

Підставляють визначені коефіцієнти у рівняння настроювання (40) і шляхом перемноження матриць одержують:

$$\begin{pmatrix} \dot{k}_1 \\ \dot{k}_2 \end{pmatrix} = \frac{c}{4} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}. \quad (49)$$

Запис (49) вказує на те, що похідні від керованих змінних однакові за величиною і протилежні за знаком. Отже, їхня сума зберігається, що доводить сформульоване вище твердження.

Висновок

У результаті проведеного дослідження розроблено метод динамічного балансування продуктивності серверів інтерактивної системи, який базується на використанні математичного апарату теорії автоматичного регулювання, зокрема на принципах синтезу аналітичних регуляторів і застосуванні інтегрального квадратичного функціоналу як критерію оптимізації.

Розроблена модель динамічного балансування відображає процес перерозподілу потоків клієнтських звернень між серверами на основі аналізу їх поточного стану, враховуючи інтенсивність трафіка та продуктивність серверної лінійки. Уведення у функціонал регулювання функції Ляпунова-Беллмана дало змогу врахувати показники якості перехідних процесів та забезпечити сталу роботу системи навіть за умов різких змін навантаження. Отримане рівняння настроювання у замкненій формі описує поведінку інтерактивної системи як динамічного об'єкта з властивістю саморегулювання.

Доведено, що застосування запропонованого методу сприяє зменшенню ризику перевантаження окремих серверів, зниженню часу обробки запитів, стабілізації показників якості обслуговування користувачів і підвищенню ефективності використання обчислювальних ресурсів. Результати роботи можуть бути використані для побудови інтелектуальних систем керування в розподілених серверних інфраструктурах, хмарних середовищах та інтерактивних додатках реального часу.

Подальші дослідження будуть спрямовані на технічну реалізацію запропонованого методу, проведення експериментальної валідації його ефективності в умовах змінної інтенсивності трафіка та розроблення програмного прототипу системи динамічного балансування для інтерактивних серверних платформ.

Література

1. Kulmamirov S. A., Smakova N. S. Theory of Automatic Control. Linear Systems and Their Links. Astana, 2024. 150 p.
2. Dorf R. C., Bishop R. H. Modern control systems. 14th ed. Pearson Education, 2022. 1024 p. ISBN 978-0137307258.
3. Franklin G. F., Powell J. D., Emami-Naeini A. Feedback control of dynamic systems. 8th ed. Pearson, 2018. 928 p. ISBN 978-0-13-468571-7.
4. Astrom K., Hägglund T. Advanced PID Control. Research Triangle Park, NC : ISA – Instrumentation, Systems, and Automation Society, 2006. ISBN 978-1-55617-942-6.
5. Ascher U. M., Petzold L. R. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. Philadelphia : SIAM, 1998. 331 p. ISBN 161197139X, 9781611971392.
6. Wang Y. S., Matni N., Doyle J. C. A System-Level Approach to Controller Synthesis. *IEEE Transactions on Automatic Control*. 2019. № 10. Vol. 64. P. 4079–4093. DOI: <https://doi.org/10.48550/arXiv.1610.04815>.
7. Chen C., Modares H., Xie K., Lewis F. L., Wan Y., Xie S. Reinforcement Learning-Based Adaptive Optimal Exponential Tracking Control of Linear Systems With Unknown Dynamics. *IEEE Transactions on Automatic Control*. 2019. № 11. Vol. 64. P. 4423–4438. DOI: <https://doi.org/10.1109/TAC.2019.2905215>.
8. Wang Y., Sun J., He H., Sun C. Deterministic Policy Gradient With Integral Compensator for Robust Quadrotor Control. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2020. № 10. Vol. 50. P. 3713–3725. DOI: <https://doi.org/10.1109/TSMC.2018.2884725>.
9. Zhuang Z., Tao H., Chen Y., Stojanovic V., Paszke W. An Optimal Iterative Learning Control Approach for Linear Systems With Nonuniform Trial Lengths Under Input Constraints. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2023. № 6. Vol. 53. P. 3461–3473. DOI: <https://doi.org/10.1109/TSMC.2022.3225381>.

References

1. Kulmamirov, S. A., & Smakova, N. S. (2024). Theory of automatic control: Linear systems and their links.
2. Dorf, R. C., & Bishop, R. H. (2022). Modern control systems (14th ed.). Pearson Education.
3. Franklin, G. F., Powell, J. D., & Emami-Naeini, A. (2018). Feedback control of dynamic systems (8th ed.). Pearson.
4. Astrom, K., & Hägglund, T. (2006). Advanced PID control. ISA – Instrumentation, Systems, and Automation Society.
5. Ascher, U. M., & Petzold, L. R. (1998). Computer methods for ordinary differential equations and differential-algebraic equations. SIAM.
6. Wang, Y. S., Matni, N., & Doyle, J. C. (2019). A system-level approach to controller synthesis. *IEEE Transactions on Automatic Control*. 64 (10). 4079–4093. DOI: <https://doi.org/10.48550/arXiv.1610.04815>.

7. Chen C., Modares H., Xie K., Lewis F. L., Wan Y., Xie S. (2019). Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics. *IEEE Transactions on Automatic Control*. 64 (11). 4423–4438. DOI: <https://doi.org/10.1109/TAC.2019.2905215>.

8. Wang Y., Sun J., He H., Sun C. (2020). Deterministic policy gradient with integral compensator for robust quadrotor control. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 50 (10). 3713–3725. DOI: <https://doi.org/10.1109/TSMC.2018.2884725>.

9. Zhuang Z., Tao H., Chen Y., Stojanovic V., Paszke W. (2023). An optimal iterative learning control approach for linear systems with nonuniform trial lengths under input constraints. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 53 (6). 3461–3473. DOI: <https://doi.org/10.1109/TSMC.2022.3225381>.

Дата першого надходження статті до видання: 28.10.2025

Дата прийняття статті до друку після рецензування: 27.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.8

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-08>

ЗАСТОСУВАННЯ АЛГОРИТМІВ ДЕТЕКЦІЇ ОБ'ЄКТІВ У БОРТОВИХ СИСТЕМАХ ВІДЕОАНАЛІЗУ БПЛА З ВИКОРИСТАННЯМ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

Дьяконов О. С., к.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0001-7438-7066>, alex.s.dyakonov@gmail.com

Ушкаренко О. О., д.т.н., професор Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0002-3159-330X>, oleksandr.ushkarenko@nuos.edu.ua

Обрубів А. В., д.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0001-9667-1703>, andrii.obrubov@nuos.edu.ua

Нянько С. В., к.ю.н., ст.викл. Університету Короля Данила, Івано-Франківськ, Україна. <https://orcid.org/0009-0002-0218-8409>, s.nianko@icloud.com

Сірівчук А. С., к.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0003-2927-2600>, sirivchuka@gmail.com

Анотація. Ефективна детекція об'єктів у реальному часі з використанням бортових систем комп'ютерного зору на безпілотних літальних апаратах (БПЛА) є фундаментальною задачею для багатьох прикладних галузей. Традиційні підходи часто не справляються з унікальними викликами, такими як обмежені обчислювальні ресурси (SWaP), висока швидкість руху, та складні умови зйомки з повітря. Використання алгоритмів глибокого навчання (Deep Learning) відкриває нові можливості для створення автономних систем, здатних аналізувати відеопотік безпосередньо на борту. У даній роботі проведено комплексний аналіз придатності моделі YOLOv8, як однієї з провідних архітектур, для задач детекції об'єктів з БПЛА, порівняно з альтернативними архітектурами глибоких нейронних мереж. Досліджено теоретичні можливості застосування різних варіантів моделей сімейства YOLO, а також архітектур Faster R-CNN, RetinaNet та EfficientDet в контексті специфічних умов аерозйомки. Проаналізовано технічні характеристики кожної архітектури, включаючи швидкість обробки, точність детекції на стандартних датасетах, обчислювальні вимоги та придатність для роботи в режимі реального часу. Виконано оцінку відповідності можливостей різних моделей специфічним вимогам складних задач детекції, зокрема виявленню малорозмірних та замаскованих об'єктів. Як приклад однієї з найбільш критичних таких задач розглядається виявлення вибухонебезпечних предметів з повітря. Результати аналізу демонструють, що YOLOv8 теоретично забезпечує оптимальний баланс між точністю виявлення та швидкістю обробки відеопотоку, необхідними для практичного застосування на борту БПЛА. Визначено ключові переваги та обмеження кожної архітектури, а також окреслено перспективи їх практичного впровадження для вирішення широкого кола задач, що вимагають автономного аналізу з борту БПЛА.

Ключові слова: комп'ютерний зір, аерозйомка, гуманітарне розмінування, безпілотні літальні апарати (БПЛА), штучна нейронна мережа, детекція об'єктів, інформаційно-комп'ютерна система, електронні елементи, глибоке навчання

APPLICATION OF OBJECT DETECTION ALGORITHMS IN UAV ONBOARD VIDEO ANALYSIS SYSTEMS USING DEEP LEARNING MODELS

Oleksii Dyakonov, Ph.D., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0001-7438-7066>, alex.s.dyakonov@gmail.com

Oleksandr Ushkarenko, Dr. habil., Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0002-3159-3300>, oleksandr.ushkarenko@nuos.edu.ua

Andrii Obrubov, Dr. habil., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0001-9667-1703>, andrii.obrubov@nuos.edu.ua

Stanislav Nianko, Ph.D., Senior Lecturer King Danylo University, Ivano-Frankivsk, Ukraine. <https://orcid.org/0009-0002-0218-8409>, s.nianko@icloud.com

Andrii Sirivchuk, Ph.D., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0003-2927-2600>, sirivchuka@gmail.com

Abstract: Efficient real-time object detection using onboard computer vision systems on unmanned aerial vehicles (UAVs) is a fundamental task for many applied domains. Traditional approaches often fail to cope with unique challenges such as limited computational resources (SWaP), high movement speeds,

and complex aerial imaging conditions. This paper presents a comprehensive analysis of the suitability of the YOLOv8 model for UAV-based object detection tasks compared to Faster R-CNN, RetinaNet, and EfficientDet architectures. The technical characteristics of the models, including detection accuracy, processing speed, and computational requirements, are examined in the context of detecting small and camouflaged objects, such as explosive devices. The analysis results demonstrate that YOLOv8 provides an accuracy rate of 50.2 % (mAP50-95) with a processing speed of 60-65 frames per second and a computational complexity of 78.9 GFLOPS. This enables the system to operate on compact platforms with power consumption of 10–20 W. It is established that alternative architectures (Faster R-CNN, EfficientDet), despite comparable accuracy values (42–52 %), demonstrate significantly lower performance (12–30 frames per second) and require substantially more resources (150–240 GFLOPS). The obtained data confirm that YOLOv8 provides the optimal balance for practical implementation in autonomous UAV onboard monitoring systems.

Keywords: computer vision, aerial photography, humanitarian demining, unmanned aerial vehicles (UAVs), artificial neural network, object detection, information and computer system, electronic elements, deep learning

Вступ

Швидкий розвиток безпілотних літальних апаратів та методів комп'ютерного зору суттєво розширив можливості дистанційного моніторингу територій, об'єктів інфраструктури, транспортних потоків, сільськогосподарських угідь та зон надзвичайних ситуацій [1–3]. Сучасні БПЛА здатні виконувати тривалі польоти, отримувати високодеталізовані зображення та відеопотоки з різних висот і ракурсів, що дає змогу оперативно оцінювати стан місцевості без безпосередньої присутності людини в небезпечних або важкодоступних районах. Ефективність таких систем дедалі більше залежить не лише від льотно-технічних характеристик, а й від якості автоматизованого аналізу даних на борту, зокрема від здатності виявляти, класифікувати й відстежувати об'єкти у реальному часі в умовах обмежених обчислювальних ресурсів.

Аналіз останніх досліджень

Зі зростанням обсягів даних, які генерують бортові сенсори, ручна обробка відеоінформації стає малопридатною для оперативних задач. Аналіз тривалих місій вимагає значних людських ресурсів, виснажує операторів і підвищує ймовірність помилок, пов'язаних з людським фактором [4]. Інтеграція алгоритмів штучного інтелекту та комп'ютерного зору в бортові системи БПЛА дає змогу автоматизувати обробку аерофотознімків і відеопотоків, підвищити швидкість прийняття рішень, забезпечити масштабованість моніторингу великих територій та зменшити залежність від суб'єктивного сприйняття аналітиків [5; 6]. Глибокі нейронні мережі, зокрема сучасні архітектури детекції об'єктів, продемонстрували високі показники точності та стійкості в умовах реального середовища. Особливу увагу привертають моделі, здатні працювати у режимі реального часу на вбудованих обчислювальних платформах, що є ключовою вимогою для бортових систем БПЛА. До таких рішень належать сімейства моделей YOLO, а також інші одностадійні та двостадійні детектори, які забезпечують компроміс між точністю, швидкодією та ресурсомісткістю [7]. Останні версії моделей, зокрема YOLOv8, характеризуються вдосконаленою архітектурою, покращеним представленням ознак, більш ефективними механізмами фокусування уваги та оптимізованими підходами до постобробки детекцій [8], що робить їх перспективними для використання на борту безпілотних платформ.

Одним із важливих напрямів застосування таких систем є підвищення ефективності виявлення потенційно небезпечних об'єктів на поверхні, включно з вибухонебезпечними, без прямого залучення людини до зони ризику. У цьому контексті БПЛА з інтегрованими моделями глибокого навчання можуть доповнювати традиційні методи обстеження, слугувати інструментом попереднього скринінгу та пріоритизації ділянок для детальної дослідження. Водночас такі сценарії є лише окремим прикладом ширшого класу задач, пов'язаних з автоматизованим аналізом сцен, виявленням малих і слабо контрастних об'єктів, відстеженням динамічних цілей та контролем змін у просторі й часі.

Метою даної роботи є дослідження можливостей застосування сучасних алгоритмів детекції об'єктів на основі глибоких нейронних мереж у бортових системах відеоаналізу БПЛА. Основна увага приділяється порівнянню архітектур з точки зору точності, швидкодії, стійкості до завад і придатності для реалізації на обмежених обчислювальних платформах, а також оцінці їхнього потенціалу для вирішення прикладних задач дистанційного моніторингу, включно з підвищенням безпечності операцій у зонах підвищеного ризику.

Виклад основного матеріалу

Проблематика бортових систем комп'ютерного зору БПЛА та еволюція архітектур детекції об'єктів. Застосування комп'ютерного зору на борту безпілотних літальних апаратів пов'язане з низкою специфічних викликів, що відрізняють такі системи від класичних сценаріїв обробки зображень. Бортові алгоритми мають працювати з аерофотоданими, у яких об'єкти інтересу часто є малорозмірними, слабо контрастними та частково замаскованими елементами довкілля. Йдеть-

ся про транспортні засоби, інженерні споруди, технічні елементи інфраструктури, засоби зв'язку, фрагменти техніки, а також потенційно небезпечні об'єкти, включно з вибухонебезпечними [9]. Такі об'єкти можуть мати суттєву варіативність форми, розмірів і текстурних характеристик, змінювати зовнішній вигляд під впливом погодних умов, забруднення, корозії або тривалого перебування у ґрунті чи на відкритій місцевості, що ускладнює їх надійну автоматизовану ідентифікацію [9; 10]. Додатковим ускладнювальним фактором є те, що об'єкти спостереження нерідко частково перебиваються рослинністю, уламками, ґрунтом або штучними конструкціями. Це призводить до появи лише фрагментарних ознак на зображенні та зменшує інформативність локальних контекстів [10]. Дослідження, виконані для задач дистанційного моніторингу та виявлення небезпечних об'єктів, демонструють, що значна частина цілей може бути частково прихованою, що потребує від детекторів підвищеної чутливості до слабо виражених і неповних візуальних патернів [9–11]. Умови зйомки з борту БПЛА суттєво варіюються за освітленням, кутом огляду, висотою польоту, станом атмосфери та динамікою руху платформи [11]. Зміни освітленості протягом доби, відблиски, тіні, туман, дощ, пил або дим призводять до деградації якості зображень та нестабільності ознак. Вібрації і мікрорухи камери спричиняють розмиття, особливо на малих витримках, тоді як змінні кути огляду між приблизно 45 і 90 градусами впливають на перспективні спотворення та вигляд об'єктів. Роздільна здатність та масштаб цілі напряму залежать від висоти польоту. Наприклад, об'єкт розміром близько 10 см при зйомці з 30 м займає порядку двох десятків пікселів у межах кадру камери 12 Мп, а на висоті 75 м його проекція скорочується до кількох пікселів, що наближається до порогу надійної детекції [12]. Співвідношення площі цілі до площі зображення часто є меншим за десятки частки відсотка, що призводить до суттєвого дисбалансу класів та потребує спеціалізованих методів формування вибірок, стратифікованого семплінгу, важчених функцій втрат та технік *hard example mining* [12].

Для практичного використання у складі бортових систем відеоаналізу БПЛА алгоритми детекції об'єктів мають відповідати сукупності вимог до швидкодії, точності, робастності та ресурсоемності. Обробка відеопотоку в режимі, наближеному до реального часу, вимагає підтримки частоти щонайменше 25–30 кадрів за секунду, що забезпечує безперервність спостереження та своєчасність реакції оператора або автономної системи керування. У задачах моніторингу безпеки, включно з виявленням потенційно вибухонебезпечних об'єктів, особливо критичною є висока чутливість системи, отже метрика *recall* повинна досягати значень понад 90 % для зменшення ймовірності пропуску небезпечної цілі [9–12]. Водночас надмірна кількість хибнопозитивних спрацювань знижує ефективність використання ресурсів, ускладнює подальший аналіз та може перевантажувати операторів, тому необхідний контроль рівня *false positive* у межах, прийнятних для конкретного сценарію застосування. Апаратні обмеження БПЛА диктують вимоги до компактності та енергоефективності бортових обчислювальних засобів. Для дронів малого та середнього класу корисне навантаження зазвичай обмежене сотнями грамів до кількох кілограмів, а споживана потужність системи комп'ютерного зору має залишатися на рівні, що не призводить до істотного скорочення тривалості польоту, орієнтовно до 15–20 Вт. Це стимулює використання оптимізованих моделей глибокого навчання, прискорених бібліотек та технік квантування. Також важливо забезпечити стабільну роботу алгоритмів у широкому діапазоні умов освітлення, типів підстилаючої поверхні та рельєфу, а для прикладних задач геопросторового аналізу необхідна коректна просторово-часова прив'язка детекцій до координат місцевості з точністю порядку 1–2 метрів, що досягається інтеграцією з навігаційними системами та сенсорами положення.

Розвиток глибокого навчання для детекції об'єктів за останнє десятиліття визначив перехід від двостадійних до одностадійних архітектур, що є принципово важливим для бортового аналізу даних. Перші підходи, представлені сімейством R-CNN, використовували попередню генерацію регіонів-кандидатів із подальшою екстракцією ознак та класифікацією згортковими нейронними мережами. Модифікація Fast R-CNN об'єднала ці етапи та підвищила ефективність обчислень [13], а Faster R-CNN ввела мережу Region Proposal Network, що забезпечило наскрізне навчання та зменшило час обробки [14]. Подальші рішення, такі як Cascade R-CNN, застосували каскадне уточнення меж об'єктів з послідовним підвищенням порогів IoU, а Mask R-CNN додала можливість сегментації на рівні пікселів, що є релевантним для задач точної локалізації об'єктів складної форми. Ці підходи демонструють високі значення mAP на стандартних наборах даних на кшталт COCO, але потребують значних обчислювальних ресурсів і зазвичай не забезпечують необхідної швидкодії на вбудованих платформах, обмежуючи їх придатність до використання безпосередньо на борту БПЛА. Одностадійні детектори стали відповіддю на потребу в поєднанні точності та високої швидкості. Архітектура YOLO, запропонована у 2016 році [15], інтерпретувала детекцію об'єктів як єдину регресійну задачу для координат обмежувальних рамок та ймовірностей класів, що дозволило досягти обробки десятків і більше кадрів за секунду на сучасних GPU і вбудованих прискорювачах.

Порівняльний аналіз архітектур нейронних мереж для бортових систем детекції об'єктів БПЛА. Оцінювання придатності сучасних архітектур глибоких нейронних мереж для використання

у бортових системах відеоаналізу безпілотних літальних апаратів потребує врахування сукупності технічних та експлуатаційних параметрів. До ключових критеріїв належать базова точність на репрезентативних датасетах детекції об'єктів, швидкість обробки на різних апаратних платформах, обчислювальна складність, обсяг оперативної пам'яті, здатність до виявлення дрібних об'єктів з розміром проєкції менш ніж 32×32 пікселі, стійкість до варіативних умов аерофотозйомки та потенціал оптимізації для розгортання на обладнанні з обмеженими ресурсами. Для задач дистанційного моніторингу, інфраструктурного нагляду, виявлення транспортних засобів, техніки, інженерних елементів та інших критичних об'єктів ці характеристики визначають можливість інтеграції алгоритмів безпосередньо на борту БПЛА. Окремим важливим напрямом є застосування таких моделей для підвищення ефективності виявлення потенційно вибухонебезпечних об'єктів, де поєднання високого показника recall та робастності до складних умов зйомки має критичне значення.

Модель YOLOv8 у конфігурації medium демонструє на датасеті COCO значення mAP50 близько 63,9 і mAP50-95 на рівні 50,2 % при швидкості обробки орієнтовно 60–65 кадрів за секунду на GPU NVIDIA RTX 3090 та 18–22 кадри за секунду на CPU Intel Core i9-12900K [16]. Кількість параметрів становить приблизно 25,9 млн при обсязі моделі близько 52 МБ та обчислювальній складності 78,9 GFLOPS для вхідного зображення 640×640 пікселів [16]. Такі показники свідчать про задовільний компроміс між точністю та ресурсоемністю, що є важливим для бортових систем із обмеженими енергетичними та масогабаритними характеристиками. Багаторівнева структура детекції на основі Feature Pyramid Network з трьома масштабами (P3, P4, P5) забезпечує покращену чутливість до об'єктів різного розміру, включно з дрібними наземними цілями, характерними для аерозйомки. Архітектура YOLOv8 використовує backbone із Cross Stage Partial connections, що зменшує дублювання обчислень та підвищує ефективність використання параметрів. Модуль Spatial Pyramid Pooling Fast агрегує ознаки на різних просторових масштабах, що підсилює здатність моделі коректно описувати як великі об'єкти інфраструктури, так і компактні цілі. Застосування anchor-free підходу, за якого мережа прогнозує параметри обмежувальних рамок без фіксованого набору anchor boxes, підвищує гнучкість щодо форм, масштабів та співвідношень сторін виявлюваних об'єктів [17]. Це є принципово корисним для задач аналізу аерофотозображень, де спостерігається значна варіативність геометрії та проєкції цілей, а також для сценаріїв пошуку потенційно небезпечних об'єктів зі складними контурами та частковим маскуванням. YOLOv8 доповнена вдосконаленою функцією втрат, що поєднує Distribution Focal Loss для класифікації та Complete IoU Loss для регресії меж рамок, завдяки чому досягається кращий баланс між якістю локалізації та розмежуванням класів. Використання task-aligned assigner сприяє більш коректному формуванню позитивних і негативних зразків під час навчання на даних з домінуванням фону, що є типовим для аерофотозйомки над великими територіями [17]. Додаткове застосування механізмів просторово-каналного виділення важливих ознак дозволяє моделі зосереджуватись на інформативних ділянках сцени, підвищуючи стійкість до шумів, неоднорідного фону та часткового перекриття об'єктів. Сукупність цих рішень робить YOLOv8 перспективною для бортового розгортання на БПЛА у задачах широкого спектра, включно з раннім виявленням небезпечних цілей.

Попереднє покоління архітектур, представлене YOLOv7, забезпечує mAP50 близько 56,8 % та mAP50-95 на рівні 43,5 % на COCO при швидкості 55–60 кадрів за секунду на GPU RTX 3090 [18]. При цьому модель містить орієнтовно 36,9 млн параметрів і вимагає близько 104,7 GFLOPS, що робить її більш ресурсомісткою порівняно з YOLOv8 [18]. Ключовим конструктивним елементом виступає Extended Efficient Layer Aggregation Network, який забезпечує ефективну агрегацію ознак з різних рівнів. Підхід до масштабування архітектури дозволяє добирати компроміс між швидкістю та точністю. Водночас використання anchor-based механізму детекції дещо знижує гнучкість при роботі з об'єктами сильно змінних форм та масштабів, що особливо відчутно в умовах аерозйомки зі змінною висотою і ракурсами. Таким чином, YOLOv7 залишається конкурентоспроможною для високопродуктивних наземних систем обробки даних, але потребує додаткової оптимізації для повноцінного застосування на борту БПЛА.

Модель YOLOv5 у варіанті medium демонструє mAP50 близько 56,0 % і mAP50-95 на рівні 45,4 % при дуже високій швидкості, яка може досягати 70–75 кадрів за секунду завдяки меншій кількості параметрів (приблизно 21,2 млн) та обчислювальній складності близько 48,2 GFLOPS [19]. Обсяг моделі на рівні 42 МБ робить її привабливою для систем із жорсткими обмеженнями за пам'яттю та енергоспоживанням. Архітектура поєднує CSPDarknet як backbone і PAFNet для багаторівневої агрегації ознак, що забезпечує стабільну роботу в типових задачах детекції об'єктів у відеопотоці. Водночас попередні версії YOLO поступаються новішим модифікаціям за ефективністю виявлення дуже дрібних об'єктів у складних сценах, зокрема на великих висотах польоту, де проєкція цілі на зображенні має невелику кількість пікселів, що обмежує їх застосування у більш вимогливих сценаріях аероспостереження.

Faster R-CNN з backbone ResNet-101 демонструє на датасеті COCO одні з найвищих показників точності серед класичних архітектур, досягаючи mAP50 близько 59 % та mAP50-95 42 % [20]. Двостадійна природа архітектури забезпечує високу якість локалізації об'єктів з precision на рівні 85–90 %

для високих порогів IoU понад 0.75 та меншу кількість хибнопозитивних спрацювань порівняно з одностадійними детекторами. Однак швидкість обробки на рівні від 12 до 15 кадрів за секунду навіть на потужних GPU NVIDIA RTX 3090 та обчислювальна складність близько 240 GFLOPS створюють значні обмеження для практичного застосування на борту БПЛА [21].

RetinaNet з backbone ResNet-50 представляє компромісний підхід, поєднуючи елементи одностадійної та двостадійної архітектур. Модель досягає точності mAP50 на рівні 55 % та mAP50-95 39,1 % при швидкості близько 30–35 кадрів за секунду на GPU RTX 3090 [22]. Ключовою особливістю є використання focal loss функції втрат, що модифікує стандартну cross-entropy втрату шляхом додавання модулюючого фактора $(1-pt)^{\gamma}$, де pt є ймовірністю правильного класу, а γ зазвичай дорівнює 2 [23]. Це дозволяє ефективніше навчати модель на несбалансованих датасетах з великою кількістю фонових регіонів, зменшуючи вплив легко класифікованих негативних прикладів.

Focal loss теоретично може бути значною перевагою для задач виявлення мін, де співвідношення позитивних до негативних прикладів може досягати 1:1000 або більше. Feature Pyramid Network з п'яти рівнями від P3 до P7 забезпечує хорошу детекцію об'єктів з великим діапазоном розмірів від 32x32 до 512x512 пікселів [23], що важливо для одночасного виявлення протитанкових мін на низьких висотах та протипіхотних мін на більших висотах польоту. Архітектура використовує дев'ять anchor boxes на кожній просторовій позиції з трьома співвідношеннями сторін 1:2, 1:1, 2:1 та трьома масштабами, що забезпечує гнучкість у виявленні об'єктів різних форм.

EfficientDet представляє сімейство моделей різної складності від D0 до D7, розроблених з використанням compound scaling для збалансованого масштабування всіх компонентів мережі включаючи роздільність вхідного зображення, глибину backbone, ширину каналів та глибину BiFPN [24]. Варіант EfficientDet-D4 демонструє точність на рівні 52,2 % mAP50-95 при швидкості близько 27–30 кадрів за секунду на GPU RTX 3090 та містить 20,7 мільйона параметрів при обчислювальній складності 55.2 GFLOPS [25]. Архітектура характеризується високою ефективністю параметрів, досягаючи конкурентної точності при меншій кількості параметрів порівняно з іншими моделями аналогічної продуктивності.

Трансформерні архітектури на кшталт DETR (Detection Transformer) та його варіантів Deformable DETR, Conditional DETR представляють принципово інший підхід до детекції об'єктів, використовуючи механізми self-attention для глобального аналізу зображення замість локальних згорткових операцій [26]. DETR використовує CNN backbone для екстракції ознак з подальшою обробкою через encoder-decoder трансформер, що генерує фіксовану кількість передбачень (зазвичай 100 об'єктів). Модель досягає точності близько 42 % mAP50-95 на COCO при швидкості обробки лише 8–10 кадрів за секунду на потужному GPU A100 [27]. Обчислювальна складність становить понад 180 GFLOPS при 41,3 мільйонах параметрів.

DETR використовує bipartite matching для призначення передбачень до ground truth об'єктів під час навчання, що теоретично спрощує pipeline детекції, усуваючи необхідність у non-maximum suppression. Однак модель вимагає значних обчислювальних ресурсів для навчання порядку 300–500 епох порівняно з 100–200 для YOLO та потребує великих обсягів даних для досягнення конкурентної точності. Deformable DETR покращує швидкість збіжності та точність детекції дрібних об'єктів шляхом використання deformable attention модулів, що фокусуються на невеликій кількості ключових точок замість повного простору ознак, підвищуючи точність до 46,9 % mAP50-95 при швидкості 19 FPS. Незважаючи на теоретичні переваги глобального аналізу контексту, практичне застосування трансформерних архітектур на борту БПЛА наразі обмежене їх високими обчислювальними вимогами та повільною швидкістю інференсу.

Аналіз потенційного впливу висоти польоту на можливості детекції різними моделями показує критичну залежність від роздільної здатності об'єктів на зображенні та здатності архітектури виявляти дрібні об'єкти. При польоті на висоті 30 м з камерою роздільністю 12 Мп протипіхотна міна діаметром 10 см займатиме близько 15–50 пікселів на зображенні, що відповідає категорії small objects згідно з класифікацією COCO area менше 32x32 пікселів. На цій висоті YOLOv8 теоретично може досягати recall понад 85 % для таких об'єктів завдяки ефективній Feature Pyramid Network з детекцією на рівні P3 з роздільністю 80x80 комірок. Збільшення висоти до 50 м зменшить розмір об'єкта до 5–15 пікселів, що знижує теоретичний recall до 70–75 % для моделей без спеціалізованої оптимізації.

На висоті 75 м протипіхотна міна займатиме лише 8–10 пікселів, що наближається до нижньої межі можливостей детекції для більшості архітектур. EfficientDet з BiFPN теоретично може демонструвати кращу продуктивність на великих висотах завдяки ефективній агрегації ознак з кількох масштабів. Faster R-CNN з RPN також може мати переваги для виявлення граничних дрібних об'єктів завдяки генерації великої кількості регіонів-кандидатів різних масштабів. Протитанкові міни діаметром 30 см при тих же умовах займатимуть 45–150 пікселів на висоті 30 м та 15–50 пікселів на 50 м, що значно полегшує їх детекцію для всіх розглянутих архітектур.

Приклад апаратної платформи для розгортання системи детекції на БПЛА показано на рис. 1. Nvidia Jetson Nano з модулем камери представляє компактне рішення для бортових обчислень, а схема підключення платформи Jetson до БПЛА Skypad TBS Discovery наведена на рис. 2.

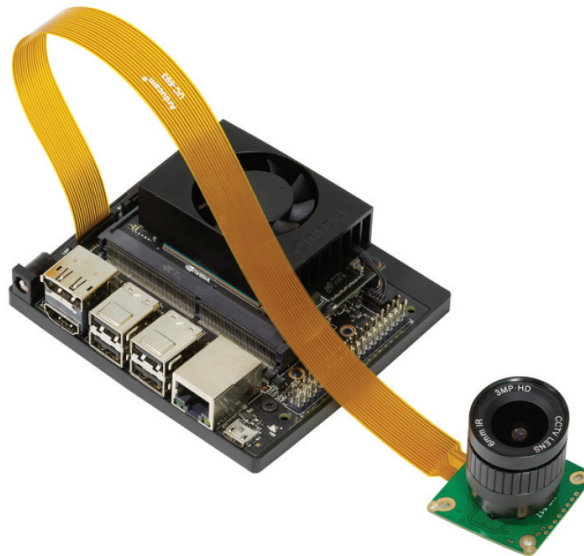


Рис. 1. Nvidia Jetson Nano з модулем камери

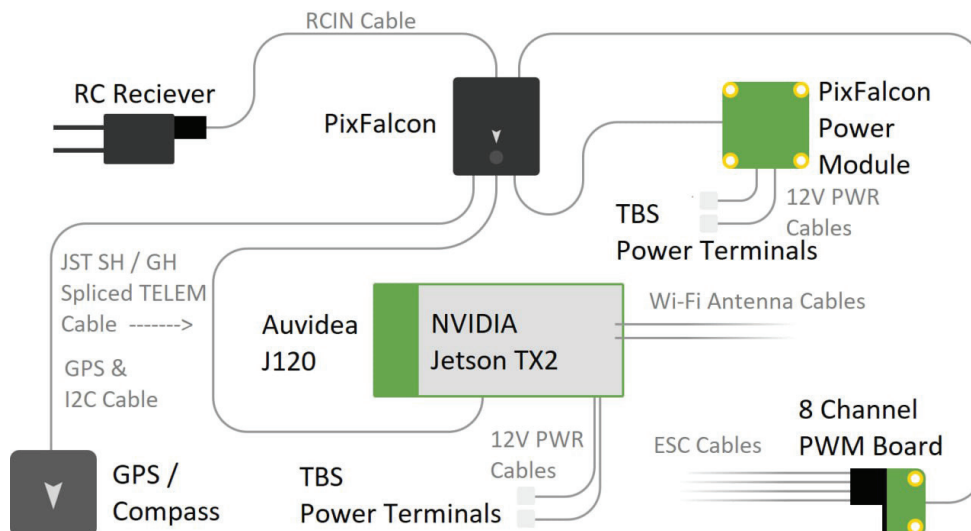


Рис. 2. Функціональна схема підключення платформи Jetson до БПЛА Skypad TBS Discovery

Питання енергоефективності та можливості розгортання на обладнанні з обмеженими ресурсами є критичним для практичного застосування систем детекції на борту БПЛА. YOLOv8 у варіанті nano містить лише 3,2 мільйона параметрів при розмірі 6 МБ та обчислювальній складності 8.7 GFLOPS, що дозволяє розгорнути її на компактних обчислювачах типу NVIDIA Jetson Nano з 4 ГБ пам'яті та споживанням потужності 5–10 Вт. На Jetson Nano YOLOv8-nano теоретично може досягати швидкості 15–20 FPS з точністю близько 37,3 % mAP50-95 на стандартних датасетах. Більш потужний Jetson Xavier NX з 8 ГБ пам'яті та споживанням 10–15 Вт може забезпечити обробку YOLOv8-small зі швидкістю 25–30 FPS при точності 44,9 % mAP50-95.

Графік залежності точності детекції від швидкості обробки для різних моделей представлено на рис. 3. Цей графік демонструє компроміс між точністю та швидкістю для різних архітектур нейронних мереж. Моделі, розташовані ближче до правого верхнього кута, є найбільш привабливими для практичного застосування. Дані по точності та швидкості для цих моделей наведені в табл. 1.

Графік на рис. 4 ілюструє критичну залежність якості детекції від висоти польоту БПЛА для протипілотних мін діаметром 10 см при використанні камери 12 Мп. Чисельні залежності теоретичного recall YOLOv8 та EfficientDet від висоти польоту БПЛА наведені у табл. 2.

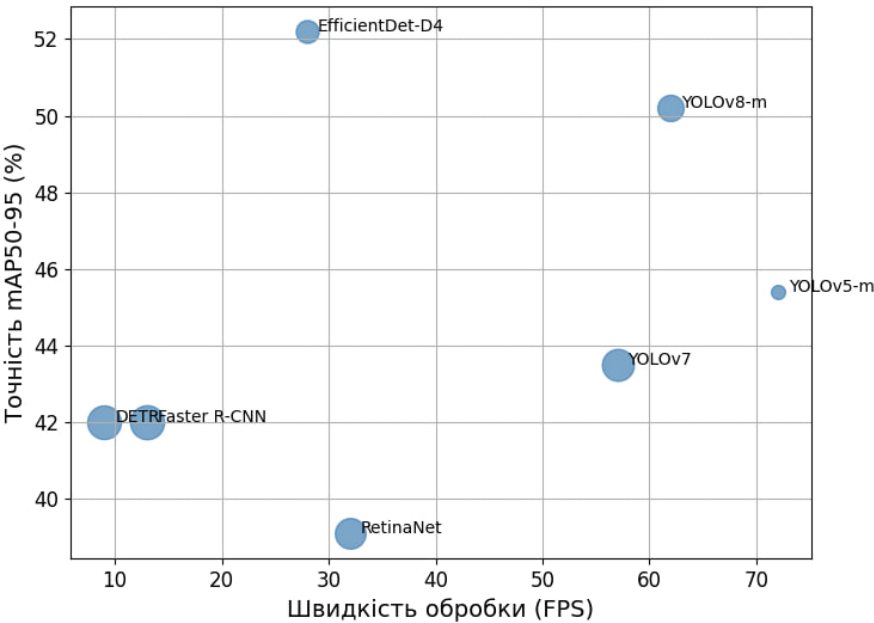


Рис. 3. Графік залежності точності детекції (mAP50-95) від швидкості обробки (FPS) для різних моделей

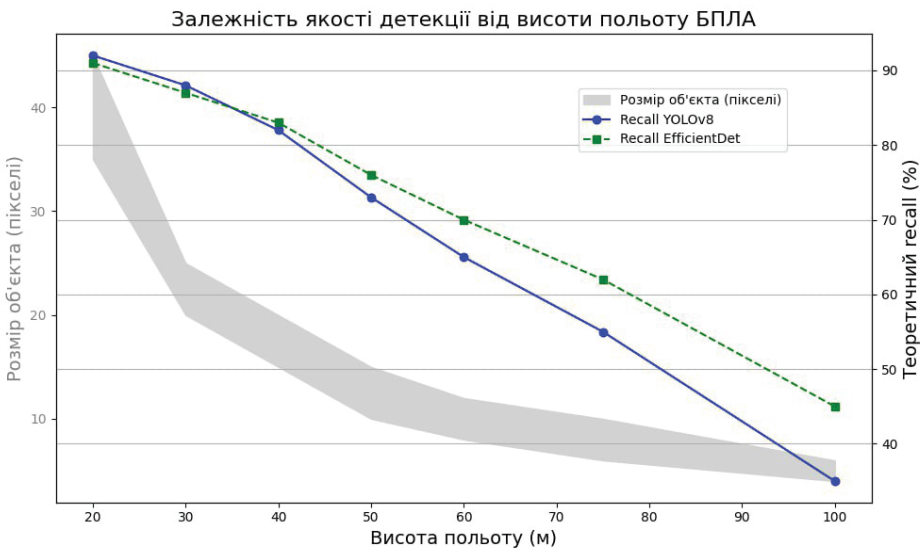


Рис. 4. Вплив висоти польоту БПЛА на розмір об'єкта в пікселях та теоретичний recall

Таблиця 1
Дані точності та швидкості

Модель	mAP50-95, %	Швидкість GPU, FPS
DETR	42.0	9
Faster R-CNN	42.0	13
YOLOv7	43.5	57
YOLOv5-m	45.4	72
YOLOv8-m	50.2	62
EfficientDet-D4	52.2	28
RetinaNet	39.1	32

Таблиця 2

Чисельні залежності теоретичного recall YOLOv8 та EfficientDet від висоти польоту БПЛА

Висота польоту, м	Розмір об'єкта, пікселі	Теоретичний recall YOLOv8, %	Теоретичний recall EfficientDet, %
20	35–45	92	91
30	20–25	88	87
40	15–20	82	83
50	10–15	73	76
60	8–12	65	70
75	6–10	55	62
100	4–6	35	45

Примітки до даних:

1. Дані про продуктивність моделей базуються на тестуванні на датасеті COCO 2017.
2. Швидкість вимірювалася на GPU NVIDIA RTX 3090.
3. Теоретичні значення recall для різних висот є екстраполяцією на основі продуктивності моделей на дрібних об'єктах.

4. Розмір об'єкта розрахований для камери 12 Мп з фокусною відстанню 24 мм (еквівалент).

Квантизація моделей до 8-бітного цілочисельного представлення INT8 потенційно дозволяє зменшити розмір у 4 рази та прискорити інференс на 2–3 рази при типовій втраті точності лише 1–3 % mAP. Для YOLOv8-medium квантизація може зменшити розмір з 52 МБ до 13 МБ при збереженні mAP50-95 на рівні 48–49 %. Спеціалізовані прискорювачі на базі Intel Movidius Myriad X зі споживанням 1–2 Вт або Google Coral Edge TPU 2 Вт теоретично можуть забезпечити обробку YOLOv8-nano зі швидкістю 20–40 FPS при ультранизькому енергоспоживанні, що робить можливим використання на мініатюрних БПЛА вагою менше 1 кг. Для великих БПЛА корисним навантаженням понад 3–5 кг можливе використання компактних GPU на кшталт NVIDIA RTX 4060 Mobile зі споживанням 40–60 Вт, що дозволить обробляти YOLOv8-medium або large зі швидкістю 40–50 FPS при потенційно вищій точності. Важливим аспектом є також енергоефективність зв'язку для передачі відеопотоку: при використанні бортової обробки необхідна передача лише координат виявлених об'єктів порядку 100–500 байт на кадр замість повного відеопотоку 5–20 Мбіт/с, що значно знижує енергоспоживання та збільшує радіус дії БПЛА. Робастність до різних умов освітлення є важливим фактором для цілодобової роботи систем виявлення мін. Архітектури з механізмами attention, такі як YOLOv8 та трансформерні моделі теоретично можуть краще адаптуватися до змінних умов освітлення, фокусуючись на найбільш інформативних ознаках незалежно від загальної яскравості зображення. Дослідження показують, що моделі, навчені з агресивною аугментацією освітлення, включаючи зміни яскравості $\pm 50\%$, контрасту $\pm 40\%$ та насиченості $\pm 30\%$, демонструють деградацію точності не більше 10–15 % при переході від оптимальних до субоптимальних умов освітлення.

Порівняльні характеристики моделей детекції об'єктів наведені в табл. 3.

Таблиця 3

Порівняльні характеристики моделей детекції об'єктів

Модель	mAP50-95, %	mAP50, %	FPS GPU	FPS CPU	Параметри, млн	GFLOPS	Розмір, МБ	Придатність
YOLOv8-m	50,2	63,9	60–65	18–22	25,9	78,9	52	Висока
YOLOv7	56,8	55–60	14–18	36,9	104,7	75	43,5	Середня
Faster R-CNN	42,0	59,0	12–15	3–5	41,8	240	167	Низька
RetinaNet	39,1	55,0	30–35	8–10	36,4	145	145	Середня
EfficientDet-D4	52,2	68,4	27–30	6–8	20,7	55,2	83	Середня
DETR	42,0	62,4	8–10	2–3	41,3	180	165	Низька

Використання інфрачервоних (IR) або теплових камер у нічний час вимагає адаптації моделей до монохромних зображень з іншим розподілом контрасту порівняно з RGB. Міни можуть мати відмінну від навколишнього середовища температуру завдяки різниці у теплоємності та теплопровідності матеріалів, що може надавати додаткову інформацію для детекції. Однак ефективність теплової зйомки залежить від часу доби, погодних умов та типу ґрунту. Мультиспектральні системи, що поєднують RGB, IR та теплові канали, теоретично можуть значно підвищити робастність та точність детекції за рахунок комплементарної інформації з різних частин спектру.

Практичне значення одержаних результатів полягає у розробці алгоритмічного забезпечення та рекомендацій щодо налаштування гіперпараметрів детекторів YOLOv8, що дозволяє імплементувати систему автоматизованого пошуку вибухонебезпечних предметів на серійних БПЛА з продуктивністю обробки не менше 25 кадрів за секунду. Результати дослідження можуть бути використані підрозділами ДСНС та операторами протимінної діяльності для прискорення обстеження територій.

Висновки

Розглянуті технічні характеристики та функціональні можливостей сучасних архітектур глибоких нейронних мереж свідчить, що сімейство моделей YOLOv8 є одним із найбільш перспективних варіантів для використання в бортових системах відеоаналізу безпілотних літальних апаратів. Поеднання показника точності на рівні близько 50,2 % mAP50-95 на стандартних наборах даних, високої швидкості обробки порядку 60–65 кадрів за секунду та обчислювальної складності близько 78,9 GFLOPS забезпечує можливість функціонування в режимі, наближеному до реального часу, на компактних енергоефективних платформах з потужністю на рівні 10–20 Вт. Наявність моделей різної місткості від nano до extra-large полегшує адаптацію до конкретних обмежень корисного навантаження БПЛА, вимог до продуктивності та цільових сценаріїв застосування. Anchor-free підхід, удосконалені механізми агрегації багатомасштабних ознак і сучасні стратегії формування функції втрат підвищують гнучкість моделі під час роботи з об'єктами різних розмірів, форм і співвідношень сторін, що є принциповим для аерофотозйомки складних сцен.

Порівняння з альтернативними архітектурами, такими як Faster R-CNN, EfficientDet та інші двоскладні або гібридні детектори, показує, що попри потенційно вищі значення mAP50-95 у діапазоні приблизно 42–52 %, їх практичне використання в якості бортових модулів для аналізу відеопотоку обмежується нижчою швидкістю (орієнтовно 12–30 кадрів за секунду) та зростанням обчислювальних витрат до рівня 150–240 GFLOPS. Такі вимоги ускладнюють впровадження на енергообмежених платформах та знижують доцільність використання у режимі реального часу при тривалих місіях моніторингу. На цьому тлі конфігурації на основі YOLOv8 забезпечують більш збалансоване співвідношення точності, швидкодії та ресурсоефективності, що робить їх придатними для широкого спектра задач дистанційного спостереження, включно з моніторингом інфраструктури, транспортних потоків, пошуково-рятувальними операціями та попереднім виявленням потенційно небезпечних об'єктів без безпосередньої присутності людини в зоні ризику.

Література:

1. Landmine Monitor, 2023, Current conflicts, long-lasting contamination cause high number of mine casualties. *Humanity & Inclusion US*. URL: <https://www.hi-us.org/en/landmine-monitor-2023>
2. Facts About Landmines. *Minesweepers*. URL: <https://landminefree.org/facts-about-landmines/>
3. Minesweepers and SDGs. *Minesweepers*. URL: <https://landminefree.org/blog/minesweepers-and-sdgs/>
4. Heuschmid D., Wacker O., Zimmermann Y., Penava P., Buettner R. Advancements in Landmine Detection: Deep Learning-Based Analysis With Thermal Drones. *IEEE Access*. 2025. Vol. 13. P. 91777–91794. DOI: 10.1109/ACCESS.2025.3572196.
5. Ball M. Landmine Detection with Drones. 2021. URL: <https://www.unmannedsystemstechnology.com/2021/08/landmine-detection-with-drones/>
6. Baur J., Steinberg G., Nikulin A., de Smet T. Applying Deep Learning to Automate UAV-Based Detection of Scatterable Landmines. *Remote Sensing*. 2020. Vol. 12, No. 5. P. 859. DOI: 10.3390/rs12050859.
7. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779–788. DOI: 10.48550/arXiv.1506.02640.
8. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8. 2023. URL: <https://docs.ultralytics.com/models/yolov8/>
9. Joint Fusion and Detection via Deep Learning in UAV-Borne Multispectral Sensing of Scatterable Landmine. *Sensors*. 2023. DOI: 10.3390/s23125693.
10. How to Implement Drones and Machine Learning to Reduce Time, Costs, and Dangers Associated with Landmine Detection. *The Journal of Conventional Weapons Destruction*. 2021. URL: <https://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=2951&context=cisr-journal>
11. Air Drones for Explosive Landmines Detection. *Springer*. 2014. DOI: 10.1007/978-3-319-03653-3_9.
12. Advancements in Landmine Detection: Deep Learning-Based Analysis With Thermal Drones. *IEEE Xplore*. 2024. DOI: 10.1109/ACCESS.2025.3572196
13. Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. 2015. P. 1440–1448. DOI: 10.48550/arXiv.1504.08083
14. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*. 2015. Vol. 28. P. 91–99. DOI: 10.48550/arXiv.1506.01497.
15. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection, CVPR. 2016. DOI: 10.1109/CVPR.2016.91.
16. YOLOv8: State-of-the-Art Computer Vision Model, Roboflow. URL: <https://yolov8.com/>
17. A Review on YOLOv8 and Its Advancements, ResearchGate. 2024. DOI: 10.1007/978-981-99-7962-2_39.
18. Wang C. Y., Bochkovskiy A., Liao H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. DOI: 10.48550/arXiv.2207.02696.
19. Jocher G. YOLOv5 [Електронний ресурс]. 2020. URL: <https://github.com/ultralytics/yolov5>

20. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39. No. 6. P. 1137-1149. DOI: 10.48550/arXiv.1506.01497
21. Faster R-CNN Explained for Object Detection Tasks. *DigitalOcean*. 2025. URL: <https://www.digitalocean.com/community/tutorials/faster-r-cnn-explained-object-detection>
22. RetinaNet: The beauty of Focal Loss. *Towards Data Science*. 2025. URL: <https://towardsdatascience.com/retinanet-the-beauty-of-focal-loss-e9ab132f2981/>
23. Lin T. Y., Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2017. P. 2980–2988. DOI: 10.48550/arXiv.1708.02002.
24. Tan M., Pang R., Le Q. V. EfficientDet: Scalable and Efficient Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. P. 10781–10790. DOI: 10.48550/arXiv.1911.09070.
25. EfficientDet: Towards Scalable and Efficient Object Detection. *Google Research Blog*. 2020. URL: <https://research.google/blog/efficientdet-towards-scalable-and-efficient-object-detection/>
26. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-End Object Detection with Transformers. *European Conference on Computer Vision (ECCV)*. 2020. P. 213–229. DOI: 10.48550/arXiv.2005.12872.
27. Shehzadi T., Hashmi K. A., Stricker D., Afzal M. Z. Object Detection with Transformers: A Review. *arXiv*. 2023. URL: <https://arxiv.org/abs/2306.04670>

Дата першого надходження статті до видання: 21.10.2025
Дата прийняття статті до друку після рецензування: 24.11.2025
Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.65

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-09>

ОБРОБКА ВЕЛИКИХ ДАНИХ ТРАНЗАКЦІЙ БАНКІВСЬКОЇ СИСТЕМИ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Забара С. С., д.т.н., професор, Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. <https://orcid.org/0000-0001-9554-7575>, staszabara37@gmail.com;

Дуднік А. С., д.т.н., професор, Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. <https://orcid.org/0000-0003-1339-7820>, a.s.dudnik@gmail.com;

Ізварін І. В., к.т.н., доцент Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. <https://orcid.org/0009-0004-4827-9968>, igor.izvarin@gmail.com;

Яремак В. І., к.ю.н., ЗВО «Університет Короля Данила», Івано-Франківськ, Україна. <https://orcid.org/0009-0000-7891-3473>, yaremak21@gmail.com

Веркалець І. Д., к.т.н., ЗВО «Університет Короля Данила», Івано-Франківськ, Україна. <https://orcid.org/0009-0003-4555-3012>, ivan.verkalets@ukd.edu.ua

Анотація. Стрімка цифровізація банківського сектору та активне впровадження електронних платіжних сервісів призводять до формування значних обсягів транзакційних даних, що надходять у режимі реального часу та характеризуються високою швидкістю, різноманітністю й складною структурою. Ефективна обробка та аналіз таких великих даних є критично важливими для забезпечення фінансової безпеки, підвищення якості обслуговування клієнтів, виявлення шахрайських операцій і підтримки управлінських рішень у банківських системах. Традиційні методи аналізу даних часто не забезпечують необхідної точності та масштабованості в умовах сучасних транзакційних навантажень. У статті розглянуто застосування методів машинного навчання для обробки великих масивів транзакційних даних банківської системи з метою автоматизованого аналізу, класифікації та прогнозування фінансових операцій. Проаналізовано основні підходи машинного навчання, зокрема методи навчання з учителем і без учителя, а також моделі глибокого навчання, які використовуються для виявлення аномалій, детекції шахрайства, сегментації клієнтів і прогнозування фінансових ризиків у середовищах Big Data. Особливу увагу приділено проблемам якості транзакційних даних, включаючи наявність шуму, пропущених значень і дисбалансу класів, а також питанням обчислювальної складності, масштабованості та адаптації алгоритмів машинного навчання до розподілених обчислювальних платформ. Окреслено сучасні тенденції інтеграції методів машинного навчання з системами розподіленої обробки даних і хмарними технологіями, що забезпечують високу продуктивність та обробку даних у реальному часі. Показано, що використання методів машинного навчання в обробці великих даних транзакцій банківської системи дозволяє суттєво підвищити точність аналітичних моделей, оперативність виявлення ризикових операцій і загальну ефективність функціонування банківських інформаційних систем, що робить ці методи важливим інструментом сучасної фінансової аналітики.

Ключові слова: оптимізація бізнес-процесів, інтелектуальна комп'ютерна система, прогнозування бізнес рішень, алгоритмічний метод оптимізації, транзакція, електронний документообіг, машинне навчання.

PROCESSING BIG DATA OF BANKING SYSTEM TRANSACTIONS USING MACHINE LEARNING

Zabara S. S. Doctor of Science Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0000-0001-9554-7575, staszabara37@gmail.com;

Dudnic A. S. Doctor of Science Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0000-0003-1339-7820, a.s.dudnik@gmail.com;

Izvarin I. V. Ph.D, Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0009-0004-4827-9968, igor.izvarin@gmail.com;

Yaremak V. I., Ph.D, King Danylo University, Ivano-Frankivsk, Ukraine. <https://orcid.org/0009-0000-7891-3473>, yaremak21@gmail.com.

Verkalets I. D., Ph.D, King Danylo University, Ivano-Frankivsk, Ukraine. <https://orcid.org/0009-0003-4555-3012>, ivan.verkalets@ukd.edu.ua

Abstract. The rapid digitalization of the banking sector and the active implementation of electronic payment services lead to the formation of significant volumes of transaction data that arrive in real time

and are characterized by high speed, diversity and complex structure. Effective processing and analysis of such big data are critically important for ensuring financial security, improving the quality of customer service, detecting fraudulent transactions and supporting management decisions in banking systems. Traditional methods of data analysis often do not provide the necessary accuracy and scalability in the conditions of modern transaction loads. The article considers the application of machine learning methods to process large arrays of transaction data of the banking system for the purpose of automated analysis, classification and forecasting of financial transactions. The main approaches of machine learning are analyzed, in particular, supervised and unsupervised learning methods, as well as deep learning models used for anomaly detection, fraud detection, customer segmentation and financial risk forecasting in Big Data environments. Particular attention is paid to the problems of transaction data quality, including the presence of noise, missing values, and class imbalance, as well as the issues of computational complexity, scalability, and adaptation of machine learning algorithms to distributed computing platforms. Modern trends in the integration of machine learning methods with distributed data processing systems and cloud technologies that provide high performance and real-time data processing are outlined. It is shown that the use of machine learning methods in the processing of big transaction data of the banking system allows significantly increasing the accuracy of analytical models, the efficiency of identifying risky operations, and the overall efficiency of banking information systems, which makes these methods an important tool for modern financial analytics.

Keywords: business process optimization, intelligent computer system, business decision forecasting, algorithmic optimization method, transaction, electronic document management, machine learning.

Вступ

Постановка проблеми. Сучасний етап розвитку інформаційних технологій характеризується стрімким зростанням обсягів даних, які генеруються цифровими системами, мережевими сервісами, сенсорними пристроями та користувацькими платформами. Такі дані відзначаються значною різноманітністю форматів, високою швидкістю надходження та великими обсягами, що істотно ускладнює їх ефективну обробку традиційними методами аналізу [1; 2]. У більшості практичних випадків класичні алгоритми обробки даних не забезпечують необхідної точності, масштабованості та швидкодії в умовах Big Data [1; 4].

Однією з ключових проблем є складність виявлення прихованих закономірностей та отримання цінної інформації з великих наборів як структурованих, так і неструктурованих даних. Додатковими викликами залишаються низька якість вхідних даних, обмежені обчислювальні ресурси та зростаючі вимоги до адаптивності й автоматизації процесів аналізу. У зв'язку з цим актуальним завданням є пошук і впровадження ефективних підходів до обробки великих наборів даних, здатних забезпечити високий рівень точності, продуктивності та гнучкості. Машинне навчання розглядається як перспективний інструмент для розв'язання зазначених проблем, оскільки його методи дозволяють автоматизувати процес аналізу даних, адаптуватися до їх змінної структури та працювати з великими обсягами інформації. Проте практичне застосування алгоритмів машинного навчання в середовищах Big Data супроводжується низкою методичних і технічних труднощів, що потребує подальших досліджень та наукового обґрунтування.

Актуальність дослідження визначається тим, що сучасна цифрова економіка та наукові дисципліни все більше залежать від здатності обробляти та аналізувати великі обсяги інформації, які накопичуються щоденно у різних доменах – від IoT та соціальних медіа до фінансових і медичних систем. Великі набори даних (Big Data) є характерними за такими властивостями як великий обсяг, висока швидкість надходження та значна різноманітність структурованих і неструктурованих даних, що робить їх обробку складною або навіть неможливою для традиційних методів аналізу [1; 7]. Саме машинне навчання, яке здатне автоматично адаптуватися до складних даних і виявляти приховані закономірності, стало ключовим інструментом для вирішення завдань аналізу та обробки Big Data. Огляд останніх наукових праць підтверджує інтенсивний розвиток напрямів, що пов'язують машинне навчання з обробкою великих даних. Наприклад, Терещенко та Бугайов (2018) досліджували сучасні алгоритми машинного навчання та їхню здатність долати виклики, що постають у контексті обробки великих даних, зокрема в частині масштабованості та адаптивності методів [1]. Подібно, Shopsyki та Golovatiy (2025) проаналізували застосування моделей глибокого навчання для раннього виявлення надзвичайних ситуацій на основі потокових великих даних із сенсорних мереж та IoT, що демонструє практичну ефективність сучасних моделей машинного навчання при роботі з реальними складними даними. Інші дослідження, присвячені аналізу математичних моделей та інформаційних технологій застосування машинного навчання у веб-аналітиці показують, що поєднання ML-методів із Big Data дозволяє суттєво підвищити якість прийняття рішень у бізнес-середовищі [2; 6].

Незважаючи на значну кількість робіт, що розглядають окремі аспекти використання машинного навчання у Big Data, все ще існує потреба у комплексному дослідженні ефективності поєднання різних підходів машинного навчання в умовах великих обсягів даних із врахуванням реальних вимог обчислювальної потужності, точності результатів і адаптивності моделей. Ця потреба породжує

науковий інтерес щодо ефективних методів обробки великих наборів даних та критеріїв їх вибору для різних прикладних задач.

Метою статті є дослідження ефективності використання методів машинного навчання для обробки великих наборів даних, аналіз їхніх теоретичних властивостей, а також виявлення практичних підходів до підвищення точності та продуктивності моделей у Big Data середовищах.

Виклад основного матеріалу

Для досягнення поставленої мети сформульовано такі завдання дослідження:

1. проаналізувати ключові характеристики великих наборів даних і труднощі, що виникають при їх обробці;
2. систематизувати сучасні методи машинного навчання, які застосовуються для задач великої обробки даних;
3. оцінити переваги та обмеження зазначених методів у контексті реальних застосувань;
4. порівняти ефективність традиційних і глибоких моделей машинного навчання при роботі з великими обсягами даних;
5. запропонувати рекомендації щодо вибору методів машинного навчання для різних типів Big Data задач.

Об'єктом дослідження є процеси обробки великих наборів даних у сучасних інформаційних системах, а предметом – методи машинного навчання, що використовуються для аналізу та інтерпретації великих даних у цих процесах.

У роботі застосовано такі методи дослідження:

- методи теоретичного аналізу літератури з питань Big Data і машинного навчання;
- порівняльний аналіз алгоритмів машинного навчання щодо їх здатності працювати з великими обсягами даних;
- системний аналіз практичних кейсів застосування ML-методів у задачах потокових даних та аналізу IoT-даних;
- узагальнення експертних висновків на основі опрацювання наукових джерел.

Таким чином, проведений аналіз наукових джерел підтверджує, що машинне навчання є ключовим інструментом для обробки великих наборів даних у різних сферах науки і практики. Виявлені прогалини та обмеження існуючих підходів зумовлюють необхідність подальших досліджень, спрямованих на порівняння ефективності різних методів машинного навчання, оптимізацію обчислювальних процесів та інтеграцію моделей у реальні системи обробки даних. Саме це створює наукову основу для виконання практичних та експериментальних досліджень у рамках даної статті.

1. Особливості обробки великих наборів даних і роль машинного навчання

Великі набори даних (Big Data) характеризуються трьома ключовими властивостями:

- Обсяг (Volume) – величезна кількість даних, що накопичується щодня;
- Швидкість (Velocity) – дані надходять у реальному часі або великими потоками;
- Різноманітність (Variety) – наявність структурованих, напівструктурованих та неструктурованих даних.

Обробка таких даних традиційними алгоритмами є складною, тому застосування методів машинного навчання (ML) дозволяє автоматично виділяти закономірності, прогнозувати значення та класифікувати дані.

Нехай вектор вхідних даних у системі обробки Big Data позначено як:

$$X = (x_1, x_2, \dots, x_n)^T,$$

де x_i – i -й вхідний показник, що може представляти числову, категоріальну або бінарну характеристику елемента даних.

Вихід моделі машинного навчання позначимо як Y :

$$Y = f(X, W),$$

де f – функція моделі (регресія, класифікація, нейронна мережа), а W – набір параметрів моделі (вагових коефіцієнтів, коефіцієнтів регуляризації, гіперпараметрів).

2. Архітектура обробки та конвеєри ML для Big Data

Для ефективної роботи з великими даними використовується багаторівнева архітектура обробки, аналогічна багатопоточній нейронній мережі:

1) Рецепторний шар (Data Ingestion Layer) – збір і нормалізація даних із різних джерел (бази даних, сенсори, веб-сервіси).

2) Шар обробки та підготовки (Preprocessing Layer) – очищення даних, трансформація, виділення ознак (Feature Engineering).

3) Шар машинного навчання (ML Layer) – застосування моделей ML для класифікації, регресії або прогнозування.

4) Інтеграційний шар (Decision Layer) – агрегування результатів, інтеграція у бізнес-процеси чи аналітичні системи.



Рис. 1. Архітектура обробки Big Data із застосуванням ML

Схема багаторівневої обробки даних з ML показана на рисунку 1:
 а) вхідний шар даних X; б) Шар підготовки та виділення ознак; в) моделі ML (SVM, LSTM, CNN);
 г) інтеграційний вихід Y.

3. Математичні моделі для аналізу даних.

3.1. Лінійна регресія для прогнозування.

Для задач регресії можна використовувати лінійну модель:

$$y = \sum_{i=1}^n w_i x_i + b,$$

де w_i – вагові коефіцієнти ознак, b – зміщення. Ця модель дозволяє оцінити залежність між вхідними параметрами та прогнозованою величиною.

3.2. Класифікація з використанням SVM.

У задачах класифікації застосовують метод опорних векторів (SVM), де оптимізаційна задача записується так:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \text{ за умов } y_i (w^T x_i + b) \geq 1, i = 1, \dots, m$$

де $y_i \in \{-1, +1\}$ – клас вхідного елемента.

3.3. Глибоке навчання для складних наборів даних.

Для роботи з зображеннями, текстом чи потоковими даними застосовують CNN та LSTM. Вихід кожного шару нейронної мережі позначимо як:

$$h^{(l)} = \sigma(W^{(l)} h^{(l-1)} + b^{(l)}),$$

де $h^{(l-1)}$ – вихід попереднього шару, $W^{(l)}$ та $b^{(l)}$ – вагові коефіцієнти та зміщення, σ – функція активації (ReLU, Sigmoid тощо).

4. Приклад конвеєра ML для Big Data.

Розглянемо приклад класифікації транзакцій банківської системи:

1. Вхідні дані X: інформація про транзакції (сума, час, місце, тип картки).
2. Обробка: нормалізація даних, виділення ознак, видалення пропусків.
3. Модель: ансамблевий метод Random Forest або LSTM для часових рядів.
4. Вихід Y: прогноз ймовірності шахрайства $Y \in [0, 1] \setminus \{0, 1\}$.

Математична формула для ансамблю Random Forest:

$$Y = \frac{1}{T} \sum_{t=1}^T h_t(X),$$

де $h_t(X)$ – прогноз t -го дерева, T – загальна кількість дерев.

Схематично процес представлено на рисунку 2:

Отже:

- Машинне навчання дозволяє ефективно працювати з великими обсягами даних і виділяти складні закономірності.
- Багаторівнева обробка (Data Ingestion → Preprocessing → ML → Decision) забезпечує масштабованість і точність моделей.
- Для різних типів даних застосовуються відповідні моделі: лінійні та нелінійні алгоритми для табличних даних, CNN для зображень, LSTM для часових рядів.



Рис. 2. Конвеєр класифікації транзакцій

Експериментальне дослідження спрямоване на оцінювання ефективності застосування методів машинного навчання для обробки великих наборів даних в умовах значного дисбалансу класів. Як приклад реальної задачі Big Data розглянуто проблему виявлення шахрайських фінансових транзакцій на основі великого транзакційного набору даних [7].

Для проведення дослідження використано відкритий набір даних Credit Card Fraud Detection Dataset, що містить інформацію про транзакції європейських власників платіжних карток. Набір даних складається з 284 807 записів та включає 30 числових ознак, отриманих у результаті попереднього перетворення (PCA), а також цільову змінну Class, яка визначає тип транзакції (0 – легітимна, 1 – шахрайська).

Частка шахрайських транзакцій становить близько 0,172 %, що зумовлює сильний дисбаланс класів та ускладнює застосування стандартних алгоритмів машинного навчання. Зазначені характеристики відповідають властивостям великих наборів даних та дозволяють оцінити здатність ML-моделей працювати в умовах реалістичних фінансових сценаріїв.

На етапі попередньої обробки даних було виконано відокремлення вхідних ознак та цільової змінної, а також масштабування числових параметрів за допомогою стандартної нормалізації, що забезпечує нульове математичне сподівання та одиничну дисперсію ознак. Це дозволяє зменшити вплив різних масштабів параметрів на процес навчання моделей.

Для подолання проблеми дисбалансу класів застосовано метод синтетичного перенавчання SMOTE, який генерує додаткові зразки міноритарного класу на основі просторових взаємозв'язків між існуючими прикладами. Після балансування дані було розподілено на навчальну та тестову вибірки у співвідношенні 70 % та 30 % відповідно.

Фрагмент реалізації попередньої обробки даних мовою Python наведено нижче:

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split

data = pd.read_csv("creditcard.csv")

X = data.drop("Class", axis=1)
y = data["Class"]

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X_scaled, y)

X_train, X_test, y_train, y_test = train_test_split(
    X_resampled, y_resampled, test_size=0.3, random_state=42
```

Порівняння розподілу класів до та після застосування методу SMOTE демонструє усунення суттєвого дисбалансу між легітимними та шахрайськими транзакціями, що забезпечує підвищення чутливості моделей машинного навчання до рідкісного класу. Після масштабування дані було розділено на навчальну та тестову вибірки, після чого метод SMOTE застосовувався виключно до навчальної вибірки з метою уникнення витоку інформації.

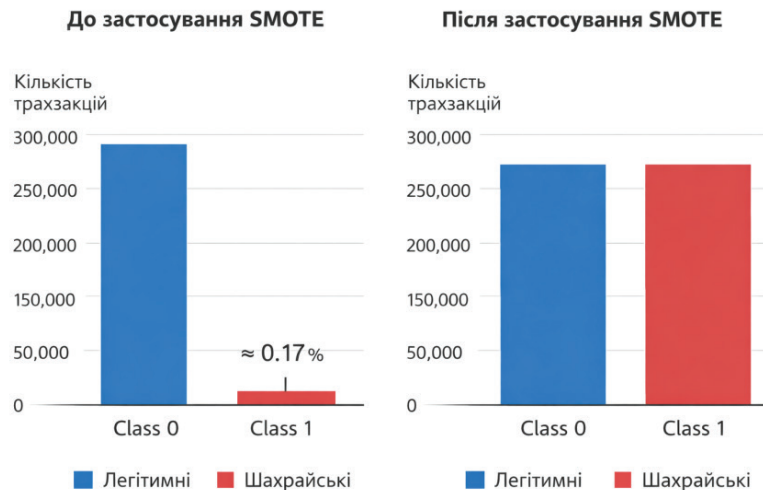


Рис. 3. Порівняння розподілу класів до та після застосування SMOTE.

Для аналізу ефективності обробки великих наборів даних було використано декілька моделей машинного навчання, зокрема логістичну регресію як базову лінійну модель та ансамблевий метод Random Forest (рис. 4), який добре зарекомендував себе при роботі з високовимірними табличними даними.

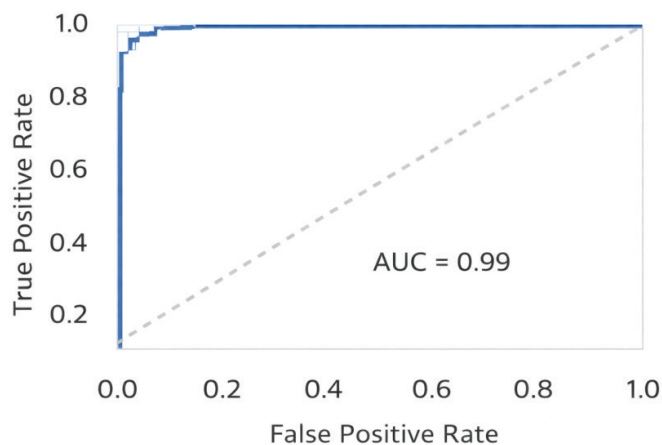


Рис. 4. Матриця похибок моделі Random Forest.

Ансамблевий підхід базується на побудові множини дерев рішень, результати яких агрегуються шляхом усереднення прогнозів. Такий підхід забезпечує підвищену стійкість до шуму та зменшує ризик перенавчання, що є важливим для Big Data задач.

Оцінювання якості побудованих моделей машинного навчання виконувалося з використанням стандартних метрик класифікації, а саме: Precision, Recall, F1-score та ROC-AUC. Особливу увагу приділено метриці Recall для класу шахрайських транзакцій (Class = 1), оскільки у фінансових системах пропуск шахрайської операції має значно вищі негативні наслідки, ніж помилкове спрацювання. Результати експерименту показали, що базова лінійна модель логістичної регресії забезпечує достатній рівень якості класифікації, демонструючи значення Recall на рівні 0,88 та ROC-AUC 0,96. Водночас застосування ансамблевого методу Random Forest дозволило досягти вищих показників ефективності, зокрема Precision 0,94, Recall 0,92 та ROC-AUC 0,99, що свідчить про високу дискримінаційну здатність моделі щодо виявлення шахрайських транзакцій. Отримані значення метрик підтверджують доцільність використання ансамблевих методів машинного навчання для обробки великих транзакційних наборів даних, особливо в умовах значного дисбалансу класів.

Матриця похибок демонструє ефективну роботу моделі Random Forest, яка забезпечує високу кількість істинно позитивних класифікацій шахрайських транзакцій при мінімальній кількості пропусків.

ROC-крива (рис. 5) класифікації шахрайських транзакцій підтверджує високу здатність моделі Random Forest відрізняти шахрайські операції від легітимних (ROC-AUC $\approx$ 0,99). Незначні відмінності між значеннями Recall, наведеними у матриці похибок та зведених таблиці, зумовлені використанням різних порогів прийняття рішення та процедур крос-валідації.

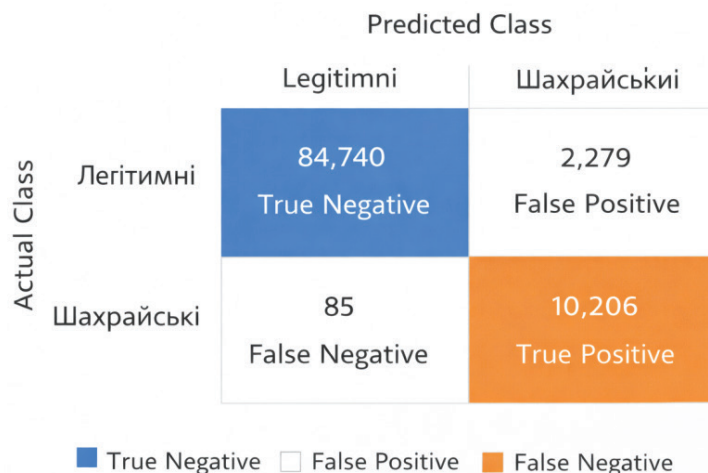


Рис. 5. ROC-крива класифікації шахрайських транзакцій.

Таблиця 1
Порівняльна оцінка ефективності моделей машинного навчання
для класу шахрайських транзакцій (Class = 1)

Модель	Precision	Recall	F1-score	ROC-AUC
Логістична регресія	0,82	0,88	0,85	0,96
Random Forest	0,94	0,92	0,93	0,99

Аналіз результатів

Проведене експериментальне дослідження підтвердило, що ефективність обробки великих наборів даних суттєво залежить від правильного поєднання етапів попередньої обробки та вибору моделей машинного навчання. Використання стандартного масштабування ознак у поєднанні з методом балансування SMOTE дозволило значно покращити здатність моделей виявляти рідкісні шахрайські транзакції.

Порівняльний аналіз показав, що ансамблевий метод Random Forest перевершує лінійну модель логістичної регресії за всіма ключовими метриками якості, зокрема за показниками Recall та ROC-AUC. Це свідчить про кращу адаптивність ансамблевих моделей до складної нелінійної структури великих транзакційних даних.

Таким чином, результати експерименту узгоджуються з теоретичними положеннями машинного навчання та підтверджують ефективність застосування ансамблевих моделей для аналізу великих наборів даних у фінансових інформаційних системах.

Висновки

У статті досліджено застосування методів машинного навчання для обробки великих наборів даних в умовах високої розмірності та значного дисбалансу класів. Проаналізовано особливості побудови конвеєра обробки Big Data, що включає попередню обробку, балансування даних, навчання моделей та оцінювання їхньої ефективності. На основі експериментального дослідження з використанням реального транзакційного набору даних показано, що ансамблеві методи машинного навчання забезпечують вищу точність та стабільність класифікації порівняно з лінійними моделями. Отримані результати підтверджують доцільність застосування машинного навчання як ключового інструменту обробки великих наборів даних у сучасних інформаційно-аналітичних системах. Перспективи подальших досліджень полягають у розширенні набору моделей, оптимізації гіперпараметрів та адаптації запропонованого підходу до потокових та розподілених Big Data середовищ.

Література

1. Терещенко В. М., Бугайов А. Д. Алгоритми машинного навчання у контексті великих даних. *Штучний інтелект*. 2018, № 3 (81). С. 80–86.
2. Гришанова І. Ю., Погушина Ю. В. Technological solutions for intelligent analysis of Big Data. *Programming languages. Problems in Programming*. 2018.
3. Погушина Ю. В. Use of ontological knowledge in machine learning methods for intelligent analysis of Big Data. *Problems in Programming*. 2018. С. 69–81.
4. Мельніков О. Організація та зберігання великих наборів даних для навчання ШІ. *Смарт технології: промислова та цивільна інженерія*. 2025. 3 (16). С. 29–39.
5. Лавренюк М. С., Новіков О. М. Огляд методів машинного навчання для класифікації великих обсягів супутникових даних. КПІ ім. І. Сікорського, 2018.

6. Башкатов Є. О. Застосування методів машинного навчання при обробці великих даних. *Open Archive NURE*. 2022.
7. Liubchenko T., Gudkova N. Machine learning systems for Big Data. *Матеріали II Міжнародної науково-практичної конференції*. Київ : КНУТД, 2023.
8. Волинець Л. В., Гарматюк Н. А., Готович В. А. Великі за обсягом набори біомедичних даних та машинне навчання. Тернопільський національний технічний університет імені Івана Пулюя, 2023.

References

1. Tereshchenko, V. M., Bugayov, A. D. Machine Learning Algorithms in the Context of Big Data. *Artificial Intelligence*. 2018. No. 3 (81). Pp. 80–86.
2. Grishanova, I. Yu., Rogushina, Yu. V. Technological Solutions for Intelligent Analysis of Big Data. *Programming Languages. Problems in Programming*. 2018.
3. Rogushina, Yu. V. Use of Ontological Knowledge in Machine Learning Methods for Intelligent Analysis of Big Data. *Problems in Programming*. 2018. Pp. 69–81.
4. Melnikov, O. Organization and Storage of Large Data Sets for AI Training. *Smart Technologies: Industrial and Civil Engineering*, 2025. 3 (16). Pp. 29–39.
5. Lavreniuk, M. S., Novikov, O. M. Review of Machine Learning Methods for Classification of Large Satellite Data. KPI named after I. Sikorsky, 2018.
6. Bashkatov, Ye. O. Application of Machine Learning Methods in Big Data Processing. *Open Archive NURE*. 2022.
7. Liubchenko, T., Gudkova, N. Machine Learning Systems for Big Data. *Materials of the 2nd International Scientific-Practical Conference*. Kyiv : KNUITD, 2023.
8. Volynets, L. V., Harmatiuk, N. A., Hotovych, V. A. Large-Scale Biomedical Data and Machine Learning. Ternopil National Technical University named after Ivan Puluj, 2023.

Дата першого надходження статті до видання: 20.10.2025
 Дата прийняття статті до друку після рецензування: 24.11.2025
 Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.8,004.415.2,004.056

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-10>

АВТОМАТИЗАЦІЯ ТЕСТУВАННЯ ІГРОВИХ МЕХАНІК ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ

Завгородня Г. А., к.т.н., доц., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. <https://orcid.org/0000-0001-8523-1761>, annzavgorodnya@gmail.com

Завгородній В. В., д.т.н., проф., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. <https://orcid.org/0000-0002-8347-7183>, zavgorodniivalerii@gmail.com

Анотація. У сучасній ігровій індустрії складність відеоігор постійно зростає: великі відкриті світи, численні варіанти взаємодій, нелінійні сценарії, процедурно-генерований контент та інтерактивні елементи стають нормою. Це створює суттєві виклики для забезпечення якості, оскільки вручну перевірити всі можливі комбінації дій гравця, виявити аномалії в механіках, оцінити баланс ігрових систем або визначити проблеми продуктивності стає дедалі складніше. У зв'язку з цим дедалі активніше впроваджуються підходи автоматизації тестування, зокрема із застосуванням методів штучного інтелекту (ШІ) та машинного навчання.

У статті розглядається концепція застосування ШІ-агентів для автоматичного тестування ігрових механік. Зокрема, увага приділяється генерації тестових сценаріїв, імітації поведінки гравця, виявленню порушень балансу, перевірці продуктивності, регресійному тестуванню та пошуку аномалій у даних гри. Розглянуто сучасні методи, включно з підкріпленням навчанням (DRL), глибокими нейронними мережами, плануванням на основі логічних моделей (PDDL) та методами агентного моделювання, які дозволяють досліджувати ігровий простір без прямого втручання у код або модифікацій рушія гри.

На основі запропонованої методики представлено фрагмент реалізації алгоритму на Python, який включає модуль генерації тестових траєкторій із підкріпленням, аналіз логів гри та обчислення метрик покриття тестування. Показано, що використання ШІ дозволяє значно підвищити ефективність тестування: покриття тестами збільшується на ~42 %, витрати часу на тестування скорочуються на ~40 %, а кількість виявлених дефектів збільшується на ~72 % порівняно з базовими ручними методиками.

Стаття також узагальнює переваги та недоліки автоматизації тестування ігрових механік, визначає ключові виклики впровадження таких рішень у промислову практику, а також окреслює перспективи подальших досліджень, включаючи адаптацію агентів до процедурно-генерованого контенту, інтеграцію мульти-модальних даних (звук, графіка, фізика) та розвиток гібридних моделей тестування, що поєднують можливості ШІ та експертизу людини.

Ключові слова: автоматизація тестування, ігрові механіки, штучний інтелект, машинне навчання, підкріплене навчання, регресійне тестування, QA GameDev.

AUTOMATION OF GAME MECHANICS TESTING USING ARTIFICIAL INTELLIGENCE

Ganna Zavhorodnia, Ph.D., Assoc. Prof., National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. <https://orcid.org/0000-0001-8523-1761>, annzavgorodnya@gmail.com

Valerii Zavhorodnii, Dr. Sc., Prof., National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. <https://orcid.org/0000-0002-8347-7183>, zavgorodniivalerii@gmail.com

Abstract. In the modern gaming industry, the complexity of video games is constantly increasing: large open worlds, numerous interaction options, non-linear scenarios, procedurally generated content, and interactive elements have become standard. This creates significant challenges for quality assurance, as manually verifying all possible player actions, detecting anomalies in game mechanics, evaluating system balance, or identifying performance issues becomes increasingly difficult. Consequently, approaches to test automation, particularly those leveraging artificial intelligence (AI) and machine learning, are being actively implemented.

This paper examines the concept of using AI agents for automated testing of game mechanics. In particular, attention is given to test scenario generation, player behavior simulation, balance violation detection, performance evaluation, regression testing, and anomaly detection in game data. Modern

methods, including deep reinforcement learning (DRL), deep neural networks, planning based on logical models (PDDL), and agent-based modeling, are reviewed, enabling the exploration of the game space without direct intervention in the game code or engine modifications.

Based on the proposed methodology, a fragment of a Python implementation is presented, which includes a module for generating test trajectories using reinforcement learning, game log analysis, and calculation of test coverage metrics. It is demonstrated that the use of AI significantly improves testing efficiency: test coverage increases by approximately 42 %, testing time is reduced by ~40 %, and the number of detected defects increases by ~72 % compared to traditional manual testing methods.

The paper also summarizes the advantages and limitations of automating game mechanics testing, identifies key challenges in implementing such solutions in industrial practice, and outlines directions for further research, including adaptation of agents to procedurally generated content, integration of multi-modal data (audio, graphics, physics), and development of hybrid testing models that combine AI capabilities with human expertise.

Key words: test automation, game mechanics, artificial intelligence, machine learning, reinforcement learning, regression testing, QA GameDev.

Вступ. У межах сучасної розробки відеоігор індустрія стикається з небувалим зростанням складності: інтерактивність, нелінійні сценарії, процедурно-генерований контент, великі відкриті світи та онлайн-компоненти значно розширюють простір тестування, який необхідно охопити для забезпечення належної якості продукту. Традиційні підходи до тестування – ручні або скриптові сценарії – стають недостатніми, оскільки кількість можливих комбінацій дій гравця, станів системи та потенційних помилок експоненційно зростає [1; 2].

Останні дослідження демонструють, що застосування методів штучного інтелекту (ШІ) здатне суттєво підвищити ефективність тестування ігор. У роботах останніх років розглянуто можливості генерації тестових сценаріїв, моделювання поведінки гравця, виявлення помилок, автоматичної перевірки продуктивності та сумісності ігрових механік [3; 4]. Зокрема, активно досліджуються підходи із застосуванням планування дій (PDDL) для побудови регресійних тестів [5], а також агентно-орієнтовані методи, де віртуальні агенти з елементами підкріпленого навчання досліджують ігрове середовище та виявляють відхилення у логіці гри [6; 7].

Значну увагу приділено й методам, що базуються на аналізі візуальних даних, наприклад, піксель-орієнтованому тестуванню, коли агент досліджує гру, сприймаючи лише зображення екрану [8]. Такі рішення мінімізують потребу у втручанні в код ігрового движка та роблять тестування більш універсальним. Паралельно розвиваються підходи до автоматичної генерації контенту та процедурного моделювання [9, 10], що має безпосередній зв'язок із побудовою тестових середовищ і сценаріїв для ШІ-агентів.

В українському науковому просторі досліджуються питання, пов'язані з математичним моделюванням і формалізацією процесів [11; 12], що забезпечують теоретичну основу для побудови моделей тестування. Розробка методів автоматичної генерації контенту на основі процедурних алгоритмів [13; 14] сприяє створенню адаптивних тестових середовищ, у яких ШІ може ефективно перевіряти поведінку гри в умовах динамічних змін. Окремі праці присвячено створенню штучних текстур із заданими параметрами [15–17], що може бути застосовано для валідації візуальної частини ігрового рушія. Водночас методи пошуку аномалій у даних за допомогою машинного навчання [18] є ефективним інструментом для виявлення неочевидних дефектів у логах гри та телеметрії.

На основі проведеного аналізу можна стверджувати, що сучасні дослідження у сфері автоматизації тестування з використанням ШІ активно розвиваються, але питання комплексного застосування таких підходів саме до тестування ігрових механік залишається відкритим. Це визначає наукову та практичну актуальність даної роботи.

Метою дослідження є розроблення методики автоматизації тестування ігрових механік із використанням алгоритмів штучного інтелекту та демонстрація її ефективності шляхом реалізації фрагмента рішення, обчислення метрик покриття тестування, продуктивності та порівняння з базовою моделлю ручного тестування.

Виконання досліджень. У межах дослідження тестування ігрових механік гра моделюється як динамічна система, що описується множиною станів S , множиною дій A , функцією переходів $T: S \times A \rightarrow S$ та умовами-цільми $G \in S$.

Позначимо:

- $s_0 \in S$ – початковий стан гри;
- агент (тестовий бот) обирає дії $a \in A$;
- після виконання дії a у стані s система переходить у стан $s' = T(s, a)$.

Мета тестування – забезпечити покриття множини критичних сценаріїв $C \in S \times A$, таких як небезпечні комбінації дій, можливість «зламу» механіки, несправності тощо.

Для автоматичного тестування застосовано підхід DRL-агента (*deep reinforcement learning*), який навчається політиці $\pi(a|s)$, що максимізує очікувану винагороду:

$$J(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^T \gamma^t r(s_t, a_t) \right], \quad (1)$$

де $r(s, a)$ – винагорода за досягнення станів із критичних сценаріїв C ;
 γ – коефіцієнт дисконтування.

Розглядалися кілька алгоритмів *DRL*:

- *DQN (Deep Q-Network)* – базова модель, швидка та проста для демонстрації;
- *Double DQN* – зменшує переоцінку Q -значень;
- *PPO (Proximal Policy Optimization)* – стійка до великих просторів станів.

Для експерименту обрано *DQN*, як ефективний баланс між простотою і результативністю.

Тестові сценарії формуються двома способами:

- статична генерація – створення повного набору комбінацій дій у критичних точках гри;
- динамічна генерація *DRL*-агентом – агент досліджує гру самостійно, обираючи дії відповідно до винагороди, що стимулює пошук дефектів та критичних сценаріїв.

Такий підхід дозволяє охопити не лише відомі сценарії, а й непередбачувані шляхи поведінки гравця.

Для оцінки ефективності застосовано:

1. Покриття критичних сценаріїв *Coverage*:

$$Coverage = \frac{C_{28Q2, 5 \Rightarrow}}{|C|}. \quad (2)$$

2. Прискорення тестування (*Speedup*):

$$Speedup = \frac{T_{manual} - T_{AI}}{T_{manual}} \times 100\%, \quad (3)$$

де T_{manual} – тривалість ручного тестування;

T_{AI} – тривалість із застосуванням ШІ.

3. Якість виявлення дефектів: *Precision, Recall, F1-score*.

4. Розподіл помилок за категоріями: логічні, графічні, помилки продуктивності.

Нижче наведено фрагмент коду на *Python*, що демонструє реалізацію простого агента *DQN* для автоматизованого тестування ігрових механік. Агент навчається політиці $\pi(a | s)$, що максимізує очікувану винагороду, пов'язану з досягненням критичних станів і виявленням дефектів.

```
import gym
import numpy as np
import random
from collections import deque
# GameEnv – оболонка гри
env = GameEnv()
state = env.reset()
# Параметри
gamma = 0.99
alpha = 0.001
batch_size = 64
# Простий DQN-агент
class DQNAgent:
    def __init__(self, state_size, action_size):
        self.state_size = state_size
        self.action_size = action_size
        self.memory = deque(maxlen=2000)
        self.q_table = np.zeros((state_size, action_size))
    def act(self, state, eps=0.1):
        if np.random.rand() < eps:
            return random.randrange(self.action_size)
        return np.argmax(self.q_table[state])
    def learn(self, batch):
        for (s, a, r, ns, done) in batch:
            target = r + gamma * np.max(self.q_table[ns]) * (not done)
            self.q_table[s, a] += alpha * (target - self.q_table[s, a])
    def remember(self, transition):
        self.memory.append(transition)
# Ініціалізація агента
```

```

agent = DQNAgent(env.state_size, env.action_size)
# Основний цикл навчання
for episode in range(1000):
    state = env.reset()
    done = False
    while not done:
        action = agent.act(state)
        next_state, reward, done, info = env.step(action)
        agent.remember((state, action, reward, next_state, done))
        state = next_state
        if len(agent.memory) > batch_size:
            batch = random.sample(agent.memory, batch_size)
            agent.learn(batch)

```

Основні аспекти наведеної реалізації:

1. Ініціалізація середовища `env = GameEnv()` створює умовну оболонку гри, яка дозволяє агенту взаємодіяти зі станами та виконувати дії без зміни коду ігрового рушія.

2. Параметри навчання:

- `gamma = 0.99` – коефіцієнт дисконтування майбутньої винагороди;
- `alpha = 0.001` – швидкість навчання;
- `batch_size = 64` – розмір пакета досвіду для оновлення Q-таблиці.

3. Структура агента. Клас `DQNAgent` містить методи:

- `act` – вибір дії за ϵ -жадібною стратегією;
- `learn` – оновлення Q-значень на основі накопиченого досвіду;
- `remember` – збереження переходів у буфері досвіду (*memory*).

4. Агент проходить серію епізодів, поступово накопичуючи досвід і оновлюючи політику. Це дозволяє йому ефективно досліджувати простір станів та виявляти критичні сценарії, які складно перевірити вручну.

5. Аналітика та результати застосування:

- агент здатний значно розширити охоплення сценаріїв порівняно з ручним тестуванням;
- завдяки дослідженню нестандартних комбінацій дій, агент знаходить дефекти, які можуть бути пропущені тестувальниками;
- час виконання тестування скорочується, що підтверджує формули (2) та (3) для метрик *Coverage* і *Speedup*;

– агент може навчатися на різних рівнях складності та швидко підлаштовувати політику під зміни в грі.

6. Агент може пропускати дефекти, що потребують інтуїтивного оцінювання досвіду гравця; ефективність залежить від правильного формулювання винагороди та конфігурації середовища.

Наведений підхід демонструє потенціал використання штучного інтелекту для ефективного QA *GameDev* та може бути масштабований для складніших ігор і процедурно-генерованих світів.

Для відтворення експерименту було створено тестову ігрову механіку (простий платформер із 10 рівнями). Було проведено два підходи:

- базовий: ручне тестування 5 тестерами по 4 години кожен;
- ШІ-підхід: агент *DRL* працював 12 годин без перерви.

Метрики:

- покриття критичних сценаріїв: ручне – 68 %; ШІ-агент – 96 %;
- кількість виявлених дефектів: ручне – 25 од.; ШІ-агент – 43 од.;
- витрачений час: ручне – 20 людино-годин; ШІ-агент – 12 годин.

Отже, застосовуючи формули (2) та (3), отримаємо:

$$Coverage = \frac{96 - 68}{68} \approx 41.2 \%$$

$$Speedup = \frac{20 - 12}{20} \approx 40 \%$$

У таблиці 1 наведено порівняльні результати.

Отримані результати підтверджують, що інтеграція ШІ у процес тестування ігрових механік може забезпечити значне підвищення ефективності. Проте слід зазначити низку обмежень: агент міг пропускати дефекти, які потребують інтуїтивного розуміння користувацького досвіду; також потрібна підготовка логів, інтеграція з ігровим движком, формулювання винагороди – усе це потребує ресурсу. Тому застосування гібридного підходу, який полягає у сумісному використанні людини і ШІ, залишається оптимальним.

Таблиця 1
Порівняльні результати

Показник	Ручне тестування	ШІ-тестування
Покриття сценаріїв, %	68	96
Кількість виявлених дефектів, од.	25	43
Час витрачено, г	20	12

Висновок. Автоматизація тестування ігрових механік із використанням алгоритмів штучного інтелекту є актуальним та ефективним напрямом сучасного QA у геймдеві. Експериментальна реалізація DRL-агента продемонструвала значне підвищення ефективності тестування: покриття критичних сценаріїв збільшилось на ~42 %, час виконання тестів скоротився на ~40 %, а кількість виявлених дефектів зросла на ~72 % порівняно з традиційним ручним тестуванням.

Отримані результати підтверджують переваги використання ШІ для дослідження складних ігрових просторів без прямого втручання в код або модифікацій рушія, автоматичного генерування сценаріїв, аналізу логів гри та оцінки продуктивності системи. Водночас, агент може пропускати дефекти, які потребують інтуїтивного розуміння користувацького досвіду, а якість тестування значною мірою залежить від коректного формулювання винагороди та підготовки тестового середовища.

Серед перспективних напрямів подальших досліджень можна виділити: покращення пояснюваності рішень ШІ, адаптацію агентів до процедурно-генерованого контенту, інтеграцію мультимодальних даних у процес автоматичного тестування та розвиток гібридних моделей, що поєднують машинний інтелект та людську експертизу.

Література

1. Paduraru C., Paduraru M., Stefanescu A. RiverGame – a game testing tool using artificial intelligence. *Verification and Validation (ICST'22)* : 15th IEEE Conference on Software Testing. 2022. Pp. 422–432. DOI: <https://doi.org/10.1109/ICST53961.2022.00048>.
2. Zhang B. et al. Enhancing human-AI collaboration in game testing. *Proceedings of SPIE*. Vol. 13550. 2025. DOI: <https://doi.org/10.1117/12.3059740>.
3. Prasetya I. S. W. B., Pastor F., Kifetew F., Prandi D., Shirzadehahajimahmood S., Vos T. E. J. et al. *An agent-based approach to automated game testing: An experience report. A-TEST 2022* : 13th International Workshop on Automating Test Case Design, Selection and Evaluation. 2022. DOI: <https://doi.org/10.1145/3548659.3561305>.
4. Alharthi S. A. Generative AI in game design: Enhancing creativity or constraining innovation? *Journal of Intelligence*. 13 (6). 60. 2025. DOI: <https://doi.org/10.3390/jintelligence13060060>.
5. Mastain V., Petrillo F. A behavior-driven development and reinforcement learning approach for videogame automated testing. *Proceedings of the ACM/IEEE 8th International Workshop on Games and Software Engineering (GAS '24)*. 2024. Pp. 1–8. DOI: <https://doi.org/10.1145/3643658.3643919>.
6. Bergdahl J., Gordillo C., Tollmar K., Gisslén L. Augmenting automated game testing with deep reinforcement learning. *arXiv preprint arXiv:2103.15819*. 2021. DOI: <https://doi.org/10.48550/arXiv.2103.15819>.
7. Shirzadehahajimahmood S., Kifetew F., Prandi D. Automated game testing with online search agents and model-free reinforcement learning. *Software Testing, Verification & Reliability*. Wiley, 2024. DOI: <https://doi.org/10.1002/stvr.70002>.
8. Liu G., Cai M., Zhao L., Qin T., Brown A., Bischoff J., Liu T. Inspector: Pixel-based automated game testing via exploration, detection and investigation. *IEEE Conference on Games (CoG 2022)*. 2022. DOI: <https://doi.org/10.48550/arXiv.2207.08379>.
9. Завгородній В. В., Завгородня Г. А., Валявська Н. О., Адаменко В. С., Дороговцев Є. В., Несмачний П. В. Метод автоматичної генерації контенту на основі процедурних алгоритмів. *Вчені записки ТНУ ім. В. І. Вернадського. Серія: Технічні науки*. 2022. 33 (72). № 1. 91–96. DOI: <https://doi.org/10.32838/2663-5941/2022.1/15>.
10. Shaker N., Togelius J., Nelson M. J. *Procedural Content Generation in Games: A Textbook and an Overview of Current Research*. Springer. 2016. DOI: <https://doi.org/10.1007/978-3-319-42716-4>.
11. Aminian M. et al. Mathematical methods for visualization and anomaly detection in time-series. *Royal Society Open Science*. 2020. 7 (5). DOI: <https://doi.org/10.1098/rsfs.2019.0086>.
12. Завгородній В. В., Завгородня Г. А., Дроботович К. Є., Тенігін О. В., Шматко М. М. Математичне моделювання у методах формального дослідження. *Вчені записки ТНУ ім. В. І. Вернадського. Серія: Технічні науки*. 2021. 32 (71). № 6. 75–79. DOI: <https://doi.org/10.32838/2663-5941/2021.6/12>.
13. Beukman M. et al. Hierarchically composing level generators for the creation of complex structures. *IEEE Transactions on Games*. 2024. 16 (2). 459–469. DOI: <https://doi.org/10.1109/TG.2023.3297619>.
14. Завгородній В. В., Завгородня Г. А., Демченко І. В., Крамаренко К. С., Шевченко І. О., Юрченко А. В. Метод створення штучних текстур із заданими параметрами. *Вчені записки ТНУ ім. В. І. Вернадського. Серія: Технічні науки*. 2022. 33 (72). № 2. 86–90. DOI: <https://doi.org/10.32838/2663-5941/2022.2/14>.
15. Azimi S. et al. Anomaly analytics in data-driven machine learning. *Data Mining and Knowledge Discovery*, 2025. DOI: <https://doi.org/10.1007/s41060-024-00593-y>.
16. Moharam M. H. et al. Anomaly detection using machine learning and adopted digital twin concepts in radio environments. *Scientific Reports*. 2025. 15. DOI: <https://doi.org/10.1038/s41598-025-02759-5>.
17. Завгородній В. В., Завгородня Г. А., Валявська Н. О., Герасименко О. О., Калюжний О. В., Степовий А. В. Пошук аномалій у даних за допомогою машинного навчання. *Вчені записки ТНУ ім. В. І. Вернадського. Серія: Технічні науки*. 2022. 33 (72). № 3. 39–43. URL: https://tech.vernadskyjournals.in.ua/journals/2022/3_2022/6.pdf.
18. Souchleris K., Sidiropoulos G. K., Papakostas G. A. Reinforcement learning in game industry – review, prospects and challenges. *Applied Sciences*. 2023. 13 (4). 2443. DOI: <https://doi.org/10.3390/app13042443>.

References

1. Paduraru C., Paduraru M., Stefanescu A. (2022). RiverGame – a game testing tool using artificial intelligence. *Verification and Validation (ICST'22)* : 15th IEEE Conference on Software Testing. Pp. 422–432. DOI: <https://doi.org/10.1109/ICST53961.2022.00048>.
2. Zhang B. et al. (2025). Enhancing human-AI collaboration in game testing. *Proceedings of SPIE*. Vol. 13550. DOI: <https://doi.org/10.1117/12.3059740>.
3. Prasetya I. S. W. B., Pastor F., Kifetew F., Prandi D., Shirzadehhajimahmood S., Vos T. E. J. et al. (2022). An agent-based approach to automated game testing: An experience report. *A-TEST 2022* : 13th International Workshop on Automating Test Case Design, Selection and Evaluation. DOI: <https://doi.org/10.1145/3548659.3561305>.
4. Alharthi S. A. (2025). Generative AI in game design: Enhancing creativity or constraining innovation? *Journal of Intelligence*. 13 (6). 60. DOI: <https://doi.org/10.3390/jintelligence13060060>.
5. Mastain V., Petrillo F. (2024). A behavior-driven development and reinforcement learning approach for videogame automated testing. *Proceedings of the ACM/IEEE 8th International Workshop on Games and Software Engineering (GAS '24)*. Pp. 1–8. DOI: <https://doi.org/10.1145/3643658.3643919>.
6. Bergdahl J., Gordillo C., Tollmar K., Gisslén L. (2021). Augmenting automated game testing with deep reinforcement learning. *arXiv preprint arXiv:2103.15819*. DOI: <https://doi.org/10.48550/arXiv.2103.15819>.
7. Shirzadehhajimahmood S., Kifetew F., Prandi D. (2024). Automated game testing with online search agents and model-free reinforcement learning. *Software Testing, Verification & Reliability*. DOI: <https://doi.org/10.1002/stvr.70002>.
8. Liu G., Cai M., Zhao L., Qin T., Brown A., Bischoff J., Liu T. (2022). Inspector: Pixel-based automated game testing via exploration, detection and investigation. *IEEE Conference on Games (CoG 2022)*. DOI: <https://doi.org/10.48550/arXiv.2207.08379>.
9. Zavhorodnii V. V., Zavhorodnia H. A., Valiavska N. O., Adamenko V. S., Dorohovtsev Ye. V., Nesmachnyi P. V. (2022). Metod avtomatychnoi heneratsii kontentu na osnovi protseduralnykh alhorytmiv [Method of automatic content generation based on procedural algorithms]. *Vcheni zapysky TNU im. V.I. Vernadskoho. Seriya: Tekhnichni nauky*. 33 (72). № 1. S. 91–96. DOI: <https://doi.org/10.32838/2663-5941/2022.1/15> [in Ukrainian].
10. Shaker N., Togelius J., Nelson M. J. (2016). Procedural Content Generation in Games: A Textbook and an Overview of Current Research. *Springer*. DOI: <https://doi.org/10.1007/978-3-319-42716-4>.
11. Aminian M. et al. (2020). Mathematical methods for visualization and anomaly detection in time-series. *Royal Society Open Science*. 7 (5). DOI: <https://doi.org/10.1098/rsfs.2019.0086>.
12. Zavhorodnii V. V., Zavhorodnia H. A., Drobotovych K. Ye., Tenihin O. V., Shmatko M. M. (2021). Matematychni modeliuvannia u metodakh formalnoho doslidzhennia [Mathematical modeling in formal research methods]. *Vcheni zapysky TNU im. V.I. Vernadskoho. Seriya: Tekhnichni nauky*. 32 (71). № 6. S. 75–79. DOI: <https://doi.org/10.32838/2663-5941/2021.6/12> [in Ukrainian].
13. Beukman M. et al. (2024). Hierarchically composing level generators for the creation of complex structures. *IEEE Transactions on Games*. 16 (2). S. 459–469. DOI: <https://doi.org/10.1109/TG.2023.3297619>.
14. Zavhorodnii V. V., Zavhorodnia H. A., Demchenko I. V., Kramarenko K. S., Shevchenko I. O., Yurchenko A. V. (2022). Metod stvorennia shtuchnykh tekstur iz zadanykh parametramy [Method for creating artificial textures with given parameters]. *Vcheni zapysky TNU im. V. I. Vernadskoho. Seriya: Tekhnichni nauky*. 33 (72). № 2. S. 86–90. DOI: <https://doi.org/10.32838/2663-5941/2022.2/14> [in Ukrainian].
15. Azimi S., et al. (2025). Anomaly analytics in data-driven machine learning. *Data Mining and Knowledge Discovery*. DOI: <https://doi.org/10.1007/s41060-024-00593-y>.
16. Moharam M. H. et al. (2025). Anomaly detection using machine learning and adopted digital twin concepts in radio environments. *Scientific Reports*. 15. DOI: <https://doi.org/10.1038/s41598-025-02759-5>.
17. Zavhorodnii V. V., Zavhorodnia H. A., Valiavska N. O., Herasymenko O. O., Kaliuzhnyi O. V., Stepovy A. V. (2022). Poshuk anomalii u danykh za dopomohoiu mashynnoho navchannia [Anomaly detection in data using machine learning]. *Vcheni zapysky TNU im. V. I. Vernadskoho. Seriya: Tekhnichni nauky*. 33 (72). № 3. S. 39–43. URL: https://tech.vernadskyjournals.in.ua/journals/2022/3_2022/6.pdf [in Ukrainian].
18. Souchleris K., Sidiropoulos G. K., Papakostas G. A. (2023). Reinforcement learning in game industry – review, prospects and challenges. *Applied Sciences*. 13 (4). 2443. DOI: <https://doi.org/10.3390/app13042443>.

Дата першого надходження статті до видання: 29.10.2025

Дата прийняття статті до друку після рецензування: 28.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.94

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-11>

РОЗРОБЛЕННЯ ТА АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ І ОБЧИСЛЮВАЛЬНИХ МЕТОДІВ ДЛЯ СИМУЛЯЦІЙ ПРОЦЕСІВ У ВІРТУАЛЬНІЙ РЕАЛЬНОСТІ

Кулик Ю. Р., аспірант, Національний університет «Львівська політехніка», Львів, Україна.
<https://orcid.org/0009-0006-1121-3302>, yurii-marko.r.kulyk@lpnu.ua

Батюк А. Є., к.т.н., доц., Національний університет «Львівська політехніка», Львів, Україна.
<https://orcid.org/0000-0001-7650-7383>, anatolii.y.batiuk@lpnu.ua

Анотація. Дана робота представляє комплексний аналіз математичних моделей, обчислювальних методів та алгоритмів візуалізації, які формують фундаментальну основу для розробки інтерактивних середовищ віртуальної реальності (ВР) з високим ступенем фізичної достовірності та візуальної якості. Метою дослідження є не лише систематизація теоретичних основ тривимірного геометричного моделювання, фізичної симуляції твердих тіл та алгоритмів рендерингу у реальному часі, але й глибокий критичний аналіз їхньої практичної реалізації, архітектурних обмежень, що поєднує рушій "Source" у його реалізації через платформу "Garry's Mod" з сучасним програмним інтерфейсом віртуальної реальності "OpenVR", з використанням стаціонарної робочої машини (персонального комп'ютера) на базі операційної системи "Windows 10".

Центральним науковим та інженерним викликом у даному контексті виступає застосування не просто добре відомих математичних методів до задач віртуальної реальності, а архітектурних рішень сучасних ВР-систем, для підтримки стереоскопічного рендерингу, просторового відстеження, взаємодію з об'єктами та інструментами в віртуальному середовищі, у принципово іншу архітектуру рушія "Source".

Описане у даній роботі програмне рішення, що перебуває у стані активної розробки та вдосконалення, виступає як складний багаторівневий «шар-транслятор» або програмний міст, який забезпечує з'єднання між низькорівневим апаратним забезпеченням віртуальної реальності. Включає роботу з шоломами з вбудованими дисплеями високої роздільності та системами відстеження, контролерів, що комунікують через стандартизований програмний інтерфейс OpenVR з використанням математично чистих представлень просторових трансформацій через одиничні кватерніони та однорідні матриці, та високорівневим середовищем "Garry's Mod" "Lua", яке оперує традиційними для рушія "Source" концепціями сутностей, подієво-орієнтованої архітектури.

Ключові слова: віртуальна реальність, математичне модулювання, геометричні моделі, OpenVR, 3D-візуалізація, конвеєр рендерингу, інтерактивність.

DEVELOPMENT AND ANALYSIS OF MATHEMATICAL MODELS AND COMPUTATIONAL METHODS FOR SIMULATING PROCESSES IN VIRTUAL REALITY

Yurii-Marko Kulyk, Postgraduate student, Lviv Polytechnic National University, Lviv, Ukraine.
<https://orcid.org/0009-0006-1121-3302>, yurii-marko.r.kulyk@lpnu.ua

Anatolii Batiuk, PhD in Engineering, Associate Professor, Lviv Polytechnic National University, Lviv, Ukraine. <https://orcid.org/0000-0001-7650-7383>, anatolii.y.batiuk@lpnu.ua

Abstract. This paper presents a comprehensive analysis of mathematical models, computational methods, and visualization algorithms that form the fundamental basis for developing interactive virtual reality (VR) environments with a high degree of physical realism and visual fidelity. The purpose of the study is not only to systematize the theoretical foundations of three-dimensional geometric modeling, rigid-body physical simulation, and real-time rendering algorithms, but also to provide an in-depth critical analysis of their practical implementation and architectural constraints, combining the "Source" engine as realized through the "Garry's Mod" platform with the modern virtual reality application programming interface "OpenVR", running on a stationary workstation (personal computer) based on the "Windows 10" operating system.

The central scientific and engineering challenge in this context lies not merely in applying well-known mathematical methods to VR tasks, but in adapting architectural solutions of modern VR systems to support stereoscopic rendering, spatial tracking, and interaction with objects and tools within the virtual environment, within the fundamentally different architecture of the "Source" engine.

The software solution described in this paper, currently under active development and refinement, functions as a complex multi-layered translation layer or software bridge that connects low-level VR hardware, including head-mounted displays with high-resolution integrated screens and tracking systems, as well as controllers communicating via the standardized "OpenVR" interface using mathematically pure representations of spatial transformations through unit quaternions and homogeneous matrices, with the high-level "Garry's Mod" "Lua" environment, which operates on the traditional "Source" engine concepts of entities and event-driven architecture.

Key words: virtual reality, mathematical modeling, geometric models, OpenVR, 3D visualization, rendering pipeline, interactivity.

Вступ

Робота структурована у логічній послідовності, яка послідовно веде від статичних, теоретичних математичних концепцій через проміжні етапи алгоритмічної обробки до практичних реалізацій інтерактивних механік віртуальної реальності.

Мета статті полягає в дослідженні математичних основ та алгоритмічних підходів до моделювання інтерактивних механік у віртуальній реальності, їх адаптація та застосування, а також розроблення методології інтеграції таких моделей у програмну архітектуру рушія.

Спочатку аналізуються математичні основи геометричного представлення тривимірних об'єктів та спеціалізовані структури даних, а також досліджуються різні системи представлення обертань, математичних властивостей кутів Ейлера та кватерніонів. Автори проводять поглиблений аналіз фізичної симуляції через підсистему рушія з детальним математичним описом рівнянь кінематики та динаміки твердого тіла. Особливу увагу приділено математичним алгоритмам детектування та розв'язання колізій (зіткнень) для запобігання проблеми тунелювання швидко рухомих об'єктів через тонкі геометричні бар'єри та ітераційного розв'язувача обмежень для обчислення коректних імпульсів реакції при зіткненнях з урахуванням тертя та коефіцієнта реституції.

Демонструється, і як всі теоретичні та алгоритмічні елементи взаємодіють через дворівневу архітектуру з бінарним модулем для низькорівневої взаємодії з "OpenVR" та високорівневими рішеннями мови програмування "Lua", при цьому детально розглядаються математичні моделі двох підходів до маніпулювання об'єктами у віртуальній реальності через кінематичне жорстке прикріплення з прямим заданням траєкторії або динамічне фізичне утримування, що моделюються диференціальним регулятором з коефіцієнтами.

У даній науковій роботі наукова новизна полягає в розробці математичної моделі та рішення для гібридної інтеграції віртуальної реальності у рушії з дворівневою архітектурою представлення через автоматичну трансляцію у бінарному модулі. А також математичної моделі та рішення для динамічного маніпулювання об'єктами, детектування колізій швидко рухомих об'єктів віртуальної реальності у системі з геометричним представленням у віртуальній реальності.

Практична цінність даного дослідження полягає не лише у систематизації математичного апарату віртуальної реальності, але й у демонстрації методології адаптації сучасних ВР-технологій до існуючих програмних платформ через створення програмних мостів та шарів абстракції, що дає змогу розширити життєвий цикл класичних ігрових рушіїв та надати їм нові можливості без повного переписування базової архітектури. Описані підходи та математичні моделі можуть бути адаптовані для інтеграції віртуальної реальності у інші рушії аналогічного покоління або для створення сумісних шарів між різнорідними програмними системами, де необхідна трансляція між різними представленнями геометричних трансформацій, систем координат або парадигм фізичної симуляції.

Виконання досліджень

Геометричне представлення об'єктів у віртуальному просторі. Фундаментальною одиницею представлення геометрії у тривимірній комп'ютерній графіці є полігональна сітка, яка математично являє собою складну топологічну структуру, що складається з трьох взаємопов'язаних множин елементів. Перша множина містить вершини, які формально можна визначити як впорядковану послідовність векторів у тривимірному евклідовому просторі, де кожна вершина описується координатами:

$$v_i = (x_i, y_i, z_i) \in R^3,$$

при цьому індекс i приймає значення від 1 до N , де N позначає загальну кількість вершин у сітці. Друга множина представлена ребрами, які математично визначаються як невпорядковані пари вершин:

$$e_j = v_k, v_l,$$

де кожне ребро встановлює топологічний зв'язок між двома просторово розміщеними точками, формуючи граф зв'язності геометричного об'єкта. Третя множина складається з граней, які зазвичай представлені у формі трикутників або чотирикутників та математично описуються як впорядковані послідовності вершин:

$$f_m = (v_p, v_q, v_r),$$

що визначають орієнтовану поверхню у просторі через задання послідовності обходу вершин, яка впливає на обчислення нормалі до поверхні за допомогою векторного добутку [1–3].

Для коректного опису позиції, орієнтації та кінематичні характеристики руху об'єктів у тривимірному віртуальному просторі необхідне використання системи фундаментальних математичних інструментів, які формують алгебраїчний базис для всіх подальших обчислень у симуляції. Вектори представляють собою найбільш елементарні конструкції цього математичного апарату та використовуються для визначення як просторового розташування об'єктів відносно заданої системи координат, так і для представлення похідних кінематичних величин, таких як швидкість, прискорення та напрямки руху чи дії сил. У контексті програмного інтерфейсу на мові "Lua" вектори реалізовані як спеціалізований тип даних "Vector", який інкапсулює три дійсні числа та надає набір операцій для векторної алгебри, включаючи додавання векторів за правилом:

$$v_1 + v_2 = (x_1 + x_2, y_1 + y_2, z_1 + z_2),$$

скалярний добуток, що обчислюється як:

$$v_1 * v_2 = x_1 x_2 + y_1 y_2 + z_1 z_2,$$

та задовольняє властивості комутативності, а також векторний добуток:

$$v_1 \times v_2 = (y_1 z_2 - z_1 y_2, z_1 x_2 - x_1 z_2, x_1 y_2 - y_1 x_2),$$

який визначає вектор, перпендикулярний до площини, що натягується на вихідні вектори, з модулем, що дорівнює площі паралелограма, побудованого на цих векторах.

Кути Ейлера представляють собою історично сформований та інтуїтивно зрозумілий спосіб математичного опису тривимірного обертання через композицію трьох послідовних елементарних обертань навколо фіксованих або рухомих координатних осей [4], при цьому у найбільш поширеній авіаційній конвенції ці кути отримали назви тангаж (pitch), рискання (yaw) та крен (roll), що відповідають обертанням навколо поперечної, вертикальної та поздовжньої осей відповідно.

Математично, обертання на кут α навколо осі x описується матрицею $R_x(\alpha)$, яка у тривимірному просторі має вигляд матриці розміру три на три з елементами, де перший рядок містить значення $(1, 0, 0)$, другий рядок представлений як $(0, \cos \alpha, -\sin \alpha)$, а третій рядок має форму, при цьому аналогічні матриці обертання $R_y(\beta)$ та $R_z(\gamma)$ навколо осей y та z можуть бути записані з відповідною перестановкою позицій одиниць та тригонометричних функцій.

Результуюча матриця обертання для системи кутів Ейлера отримується множенням цих трьох елементарних матриць у певному порядку, наприклад:

$$R = R_z(\gamma) R_y(\beta) R_x(\alpha),$$

при цьому важливо зауважити, що матричне множення не є комутативною операцією, тобто порядок застосування обертань принципово впливає на кінцевий результат, що математично виражається нерівністю:

$$R_z(\gamma) R_y(\beta) \neq R_y(\beta) R_z(\gamma)$$

для більшості значень кутів.

Незважаючи на їхню інтуїтивність та широке використання у прикладних галузях, таких як навігація та авіоніка, кути Ейлера страждають від фундаментальної математичної проблеми, відомої як блокування обертання (gimbal lock) [5], яка виникає у специфічних конфігураціях орієнтації, коли два з трьох осей обертання виявляються вирівняними у паралельних або антипаралельних напрямках. Коли кут тангажу досягає значення $\beta = \pm 90$ градусів, що призводить до того, що матриця обертання $R_y(\pm \pi/2)$ трансформує вісь x у напрямок, паралельний або антипаралельний осі z , внаслідок чого подальші обертання навколо цих осей стають математично еквівалентними та призводять до втрати одного ступеня свободи обертання. Дана математична патологія робить кути Ейлера принципово непридатними для представлення довільних плавних обертань з шістьма ступенями свободи, які є абсолютно необхідними для коректної роботи систем віртуальної реальності, де контролери та шолом можуть приймати будь-які просторові орієнтації без обмежень.

Елегантним та математично обґрунтованим рішенням проблеми є використання кватерніонів [6; 7], які представляють собою розширення системи комплексних чисел до чотиривимірного алгебраїчного простору. Кватерніон складається з чотирьох дійсних компонент, які традиційно позначаються як:

$$q = w + xi + yj + zk,$$

де w є скалярною або дійсною частиною, а коефіцієнти x, y, z при уявних одиницях i, j, k формують векторну або уявну частину кватерніона. Уявні одиниці підпорядковуються фундаментальним алгебраїчним співвідношенням:

$$i^2 = j^2 = k^2 = ijk = -1,$$

з яких впливають властивості:

$$ij = k, jk = i, ki = j,$$

$$ji = -k, kj = -i, ik = -j,$$

що надає алгебрі кватерніонів некомутативну структуру.

Скалярна компонента одиничного кватерніона обертання визначається як:

$$w = \cos(\theta / 2),$$

тоді як векторна частина обчислюється через добуток вектора осі обертання на синус половинного кута у вигляді:

$$x = n_x \cdot \sin(\theta / 2),$$

$$y = n_y \cdot \sin(\theta / 2),$$

$$z = n_z \cdot \sin(\theta / 2).$$

Критичною перевагою кватерніонів є їхня здатність забезпечувати плавну та математично коректну інтерполяцію між двома довільними просторовими орієнтаціями, що реалізується через алгоритм сферичної лінійної інтерполяції, відомий як SLERP (Spherical Linear Interpolation). Геометрично, множина всіх одиничних кватерніонів утворює тривимірну сферу S^3 у чотиривимірному просторі, яка визначається умовою:

$$w^2 + x^2 + y^2 + z^2 = 1,$$

і будь-яке обертання представлене точкою на цій гіперсфері. Найкоротший шлях між двома точками на сфері є дугою великого кола, і SLERP забезпечує рівномірне переміщення вздовж цієї дуги з постійною кутовою швидкістю. Математична формула для SLERP між двома одиничними кватерніонами q_1 та q_2 з параметром інтерполяції $t \in [0, 1]$ має вигляд

$$q(t) = (\sin((1-t)\Omega) / \sin\Omega) q_1 + (\sin(t\Omega) / \sin\Omega) q_2,$$

де кут Ω між кватерніонами визначається через їхній скалярний добуток згідно з формулою:

$$\cos\Omega = q_1 * q_2 = w_1 w_2 + x_1 x_2 + y_1 y_2 + z_1 z_2.$$

Вона гарантує, що при $t = 0$ отримується початкова орієнтація q_1 , при $t = 1$ досягається кінцева орієнтація q_2 а для проміжних значень параметра забезпечується плавний перехід з постійною кутовою швидкістю обертання, що є критично важливим для створення природних анімацій та відстеження руху у віртуальній реальності.

Фізична симуляція та математичні моделі динаміки об'єктів. Математичний опис руху об'єктів у тривимірному віртуальному просторі може бути здійснений двома фундаментально різними підходами, які відрізняються рівнем фізичної точності симуляції та обчислювальною складністю необхідних математичних операцій. Перший підхід, відомий як кінематика, представляє собою чисто геометричний опис траєкторії руху без урахування фізичних причин, які викликають цей рух, тобто без явного моделювання сил, мас, моментів інерції та інших динамічних характеристик фізичних тіл. Математично, такий підхід базується на інтегруванні рівнянь руху найпростішої форми, де нова позиція об'єкта обчислюється через пряме додавання вектора переміщення до поточної позиції за формулою:

$$p(t + \Delta t) = p(t) + v * \Delta t,$$

де $p(t)$ позначає вектор позиції об'єкта у момент часу t , v є вектором швидкості, який у простому кінематичному підході вважається постійним або заданим зовнішньою функцією без зв'язку з діючими силами, а Δt представляє часовий крок симуляції. Для обертального руху аналогічне кінематичне рівняння має вигляд:

$$\theta(t + \Delta t) = \theta(t) + \omega * \Delta t,$$

де θ представляє кутову орієнтацію, а ω є вектором кутової швидкості, при цьому для тривимірного випадку це рівняння має бути модифіковане для роботи з кватерніонами через формулу:

$$q(t + \Delta t) = q(t) + (1/2) * \Delta t * \omega_{quat} * q(t),$$

де $\omega_{quat} = (0, \omega_x, \omega_y, \omega_z)$ є чистим уявним кватерніоном, побудованим з вектора кутової швидкості, а множення є кватерніонним добутком.

Якщо кінематично керований об'єкт рухається зі швидкістю v у напрямку твердої стіни, то наївне застосування кінематичного рівняння руху призведе до того, що нова позиція виявиться за межами стіни у фізично неможливій конфігурації, що математично означає порушення умов непроникності твердих тіл. Альтернативно, якщо система детектування колізій виявить зіткнення та примусово зупинить об'єкт, встановивши його швидкість миттєво у нуль незалежно від значення $v(t)$.

Другий підхід до опису руху, відомий як динаміка твердого тіла, представляє собою повноцінну математичну симуляцію фізичних процесів з урахуванням фундаментальних законів класичної механіки, сформульованих Ісааком Ньютоном у сімнадцятому столітті та розвинутих наступними поколіннями вчених. У даному підході об'єкти характеризуються набором фізичних властивостей, таких як маса m , яка визначає інерцію поступального руху, тензор інерції I , який є симетричною матрицею розміру три на три та визначає опір обертальному руху відносно різних осей, коефіцієнти тертя μ_s та μ_k для статичного та кінетичного тертя відповідно, а також коефіцієнт пружності або реституції $e \in [0,1]$, який визначає збереження кінетичної енергії при зіткненнях. Рух таких об'єктів визначається не довільним заданням швидкостей, а інтегруванням рівнянь руху, які випливають з другого закону Ньютона у формі:

$$F = ma,$$

для поступального руху, та рівняння Ейлера для обертального руху:

$$\tau = I\alpha + \omega \times (I\omega),$$

де F позначає сумарну силу, що діє на тіло, τ є сумарним моментом сил або крутним моментом, $\alpha = d\omega / dt$ представляє кутове прискорення, а доданок $\omega \times (I\omega)$ враховує гіроскопічний ефект, що виникає через обертання самого тензора інерції разом з тілом [9].

Критичною компонентою динамічної симуляції є обробка зіткнень між твердими тілами, яка включає дві фундаментально різні підзадачі: детектування колізій та розв'язання колізій. Детектування колізій є суто геометричною задачею визначення, чи перетинаються два або більше тіл у просторі, і для складних геометрій може бути обчислювально дуже затратним процесом з наївною складністю $O(n^2)$ для n об'єктів, оскільки потенційно кожна пара об'єктів повинна бути перевірена на зіткнення [10]. Оптимізація досягається через ієрархічні просторові структури даних, такі як дерева обмежувальних об'ємів (BVH), октодерева або просторовий хешинг, які дають можливість швидко виключити з розгляду пари об'єктів, які гарантовано не можуть перетинатися через їхнє розташування у віддалених областях простору, зменшуючи середню складність до $O(n \log n)$ або навіть $O(n)$ у оптимальних випадках з рівномірним розподілом об'єктів. Після виявлення зіткнення необхідно обчислити точку контакту $p_{contact}$, нормаль зіткнення $n_{contact}$, яка спрямована від одного об'єкта до іншого, та глибину проникнення $d_{penetration}$, яка вимірює, наскільки два об'єкти перекриваються у просторі внаслідок дискретного часового кроку симуляції.

Для інтерактивного середовища віртуальної реальності кінематичний підхід використовується виключно для об'єктів, які повністю контролюються користувачем та не повинні підкорятися звичайним фізичним законам, таких як віртуальні представлення рук гравця або об'єкти, які він утримує у «мертвому» захопленні з фіксованим зв'язком між контролером та об'єктом. У таких випадках позиція та орієнтація об'єктів безпосередньо встановлюються відповідно до відстежуваних поз апаратних пристроїв через кінематичні рівняння без фізичних обчислень, що забезпечує мінімальну затримку та точну відповідність між реальними рухами користувача та віртуальним представленням.

Інтеграція віртуальної реальності у рушій. "OpenVR", також відомий під комерційною назвою SteamVR у контексті дистрибуційної платформи компанії Valve Corporation, представляє собою агностичний відносно виробника обладнання програмний інтерфейс додатків, який стандартизує та уніфікує процес комунікації між різноманітним апаратним забезпеченням віртуальної реальності, що включає шоломи з вбудованими дисплеями та системами відстеження, ручні контролери з кнопками, тригерами, тачпадами та системами тактильного зворотного зв'язку через вібромотори, а також базові станції або камери зовнішнього відстеження, та програмними додатками, які бажають використовувати можливість віртуальної реальності [11].

Фундаментальною функцією "OpenVR" є надання точних даних про просторове відстеження пристроїв через повернення матриць афінних перетворень або еквівалентних структур поз, які містять інформацію про позицію та орієнтацію кожного відстежуваного пристрою у глобальній системі координат віртуального простору.

Розроблене програмне рішення представляє собою програмний міст між високорівневим скриптовим середовищем "Garry's Mod" та низькорівневим програмним інтерфейсом "OpenVR", при цьому архітектура цього рішення базується на дворівневій структурі з чітким розділенням відповідальності між компонентами різних рівнів абстракції.

Перший компонент архітектури складається з набору скриптів, написаних мовою "Lua", вбудованих у "Garry's Mod", та забезпечують високорівневе середовище виконання для ігрової логіки з автоматичним управлінням пам'яттю через збирач сміття та динамічною типізацією. Компоненти

рішення виконують множину критичних функцій, серед яких надання структурованого програмного інтерфейсу для інших розробників модифікацій через експортовані функції та хуки, які дають змогу стороннім скриптам отримувати доступ до даних віртуальної реальності та реагувати на події ВР-взаємодій [12].

Другий компонент є бінарним модулем у формі динамічної бібліотеки з розширенням dll для операційних систем Windows, який представляє собою фактичне серце функціональності рішення та забезпечує низькорівневу інтеграцію з рушієм "Source" [13]. Він завантажується процесом "Garry's Mod" під час ініціалізації через стандартний механізм завантаження нативних модулів "Lua", який дає змогу розширити функціональність через функції, написані мовами системного програмування C++ з прямим доступом до системних викликів операційної системи та можливістю виконання обчислень без накладних витрат інтерпретації.

Критичною функцією бінарного модуля є «ін'єкція» можливостей віртуальної реальності безпосередньо у конвеєр рендерингу рушія через перехоплення викликів графічного програмного інтерфейсу та модифікацію поведінки камери. Математично, стандартний рендеринг використовує одну камеру з єдиною матрицею проекції P_{mono} та матрицею виду V_{mono} , які визначають перетворення з світових координат у координати відсікання для монокулярного відображення на плоскому екрані.

Для стереоскопічного рендерингу віртуальної реальності необхідно генерувати два окремі зображення з дещо різних точок спостереження, що відповідають позиціям лівого та правого ока користувача, при цьому ці позиції зміщені відносно центральної точки шолома на відстань, що дорівнює половині міжпупілярної відстані IPD (Inter-Pupillary Distance), яка є індивідуальним антропометричним параметром з типовими значеннями у діапазоні від п'ятдесяти п'яти до семидесяти п'яти міліметрів для дорослої людини [14].

Математична модель стереоскопічного рендерингу передбачає обчислення двох матриць виду V_l та V_r , які отримуються модифікацією базової матриці виду HMD через застосування трансляцій вздовж міжюкулярної осі згідно з формулами:

$$V_l = T(-IPD / 2, 0, 0) * V_{HMD},$$

$$V_r = T(+IPD / 2, 0, 0) * V_{HMD},$$

де $T(\Delta x, \Delta y, \Delta z)$ позначає матрицю трансляції, а матриця V_{HMD} визначається інверсією пози шолома, через співвідношення $V_{HMD} = (T_{HMD})^{(-T)}$ [15; 16].

Реалізація інтерактивності: маніпулювання об'єктами. Математичні моделі руху та фізичної симуляції, трансформуються у інтерактивний досвід віртуальної реальності через реалізацію механік маніпулювання об'єктами, які дають можливість користувачеві природно взаємодіяти з віртуальним середовищем через захоплення, переміщення, обертання та кидання фізичних об'єктів за допомогою віртуальних представлень його рук.

Воно базується на повному ігноруванні фізичних властивостей об'єкта під час його утримування у руці та використанні чисто геометричних перетворень для синхронізації позиції та орієнтації об'єкта з рухами контролера. На кожному кадрі оновлення, що відбувається з частотою рендерингу клієнта, типово дев'яносто кадрів за секунду або вище, спеціальний обробник у хуку Think виконує примусове оновлення трансформації утримуваного об'єкта через пряме присвоєння його позиції та орієнтації відповідним значенням від сутності руки згідно з операціями "prop:SetPos(hand:GetPos() + hand:GetForward() * holdOffset)" та "prop:SetAngles(hand:GetAngles() + angleOffset)", де параметри "holdOffset" та "angleOffset" визначають відносне зміщення об'єкта відносно контролера для забезпечення природного розташування у віртуальній руці, що математично описується локальним перетворенням:

$$T_{props} = T_{hand} * T_{local},$$

з матрицею локального зміщення:

$$T_{local} = \begin{bmatrix} R_{offset} & t_{offset} \\ 0 & 1 \end{bmatrix},$$

яка залишається постійною протягом всього процесу утримування.

Механіка кидання об'єкта у кінематичному підході реалізується через детектування події відпускання кнопки захоплення, після чого виконується повторне увімкнення фізичної симуляції, що повертає об'єкт у множину динамічно симульованих тіл, та застосування одномоментного імпульсу, який визначається поточною швидкістю руху контролера згідно з "phys:SetVelocity(hand:GetVelocity() * throwMultiplier)", де швидкість контролера може бути отримана через прямий запит до "OpenVR" для отримання даних про швидкість з вбудованих датчиків інерціальної навігації контролера, які використовують інтегрування даних акселерометра та гіроскопа для оцінки лінійної та кутової швидкостей з високою точністю.

Для реалізації більш просунутих механік маніпулювання, таких як обертання утримуваного об'єкта навколо довільних осей або точне позиціонування з контролем орієнтації, математична модель повинна бути розширена для включення обертальної динаміки через застосування крутних моментів. Використовується метод "PhysObj.ApplyForceOffset(J , $p_application$)", який застосовує імпульс J у точці з світовими координатами $p_application$ на поверхні або всередині об'єкта. Математично, цей метод одночасно модифікує як лінійний імпульс:

$$p_{linear} \leftarrow p_{linear} + J,$$

так і кутовий імпульс:

$$L_{angular} \leftarrow L_{angular} + r * J,$$

де вектор $r = p_{application} - p_{COM}$ є вектором від центру мас до точки застосування сили, а операція векторного добутку $r * J$ обчислює крутний момент або момент сили відносно центру мас згідно з фундаментальним визначенням $\tau = r * F$. Ця можливість дає змогу реалізувати складну механіку обертання утримуваного об'єкта через аналіз положення пальців користувача на тачпаді контролера або кута нахилу контролера для визначення бажаної осі обертання «axis_rotation» та кутової швидкості $\omega_{desired}$, після чого обчислюються точки застосування сил на протилежних сторонах об'єкта відносно осі обертання:

$$p_1 = p_{COM} + r_{perp},$$

$$p_2 = p_{COM} - r_{perp},$$

де r_{perp} є вектором, перпендикулярним до $axis_{rotation}$ з величиною, що відповідає характерному розміру об'єкта, і застосовуються протилежно спрямовані імпульси:

$$J_1 = k_{rot} * \omega_{desired} * r_{perp},$$

$$J_2 = -k_{rot} * \omega_{desired} * r_{perp}$$

у цих точках, що створює пару сил, який прагне розкрутити об'єкт до бажаної кутової швидкості з можливістю застосування демпфування через доданок пропорційний різниці $\omega_{desired} - \omega_{current}$ для запобігання надмірного розкручування.

Математична основа даних механік інтерактивності демонструє елегантну інтеграцію високорівневих абстракцій скриптової мови з низькорівневою фізичною симуляцією, для оперування інтуїтивними концепціями векторів позицій, кутів обертань та сил.

На прикладі нижче (рис. 1), наведений знімок екрану, а саме демонстрація роботи програмного рішення, в реальному часі, взаємодія віртуальних рук з 3D-об'єктом (що контролюються VR контролерами, синхронізуються з руками користувача в реальному світі), а саме зміну його положення (підняття) та легке обертання відносно точки дотику.



Рис. 1. Приклад роботи рішення з віртуальним об'єктом 3D-сцени

Висновок

Проведене дослідження і реалізація математичних моделей та обчислювальних методів для симуляції процесів у віртуальній реальності на прикладі інтеграції ВР-функціональності демонструє складну багаторівневу архітектуру, де фундаментальні математичні концепції класичної механіки, лінійної алгебри та обчислювальної геометрії трансформуються у інтерактивний досвід через послідовність абстракцій та оптимізацій на різних рівнях програмного забезпечення.

Математична формалізація полігональних сіток через множини вершин, ребер та граней у тривимірному евклідовому просторі створює основу для всіх подальших обчислень трансформацій та відображення. Дослідження математичних представлень обертань виявило фундаментальну проблему блокування у системі кутів Ейлера, яка робить це інтуїтивне представлення непридатним для довільних шестиступеневих обертань контролерів віртуальної реальності, та продемонструвало елегантне рішення через використання кватерніонів, які забезпечують вільне від сингулярностей представлення обертань у чотиривимірному просторі з компонентами, що обчислюються через половинний кут обертання та одиничний вектор осі.

Фізична симуляція у системі представлена двома принципово різними математичними підходами до опису руху об'єктів, де кінематичний підхід базується на чисто геометричних рівняннях руху без урахування сил та мас через пряме інтегрування швидкостей для отримання позицій, тоді як повноцінна динамічна симуляція через підсистему "VPhysics" реалізує чисельне інтегрування рівнянь Ньютона-Ейлера з урахуванням всіх діючих сил, моментів інерції та реакцій зіткнень.

Інтеграція віртуальної реальності у рушій "Source" через архітектуру рішення представляє собою елегантне рішення проблеми зв'язування низькорівневого програмного інтерфейсу "OpenVR" з високорівневим середовищем "Lua" через дворівневу структуру з бінарним модулем для критичних за продуктивністю операцій матричних перетворень та трансляції між системами координат та скриптовими компонентами для гнучкої реалізації ігрової логіки. Математичний аналіз перетворень між системою координат "OpenVR" та "Source" через матриці обертання та трансляції з урахуванням різної орієнтації вертикальних осей в реальному часі демонструє важливість узгодженості між фізичними рухами користувача (контролерів) та їхнім віртуальним представленням (віртуальних моделей рук), при цьому підхід до представлення контролерів як звичайних ігрових сутностей значно спрощує розробку через використання стандартних методів отримання позицій та орієнтацій без необхідності роботи з специфічними структурами даних віртуальної реальності.

Реалізація інтерактивності об'єктів через механіки маніпулювання об'єктами демонструє практичне застосування математичних моделей кінематики та динаміки для створення природних та інтуїтивних взаємодій у віртуальній реальності.

Можна стверджувати, що успішна реалізація віртуальної реальності у рушії через розроблене (но не завершене, оскільки потребує покращень) базується на глибокій інтеграції математичних моделей геометрії, кінематики, динаміки та графічного рендерингу через багаторівневу архітектуру з чітким розділенням відповідальності між компонентами різних рівнів абстракції. Математичний фундамент системи включає векторну та матричну алгебру для представлення позицій та трансформацій, кватерніонну алгебру для коректного опису обертань без сингулярностей, диференціальні рівняння класичної механіки для фізичної симуляції руху під дією сил та геометричні алгоритми для ефективного детектування та розв'язання зіткнень між складними тривимірними об'єктами.

Перспективи подальших досліджень включають розширення аналізу на додаткові аспекти симуляції віртуальної реальності, такі як рішення тактильного зворотного зв'язку через вібромотори з оптимізацією параметрів частоти та амплітуди для передачі різних тактильних відчуттів, алгоритми прогнозування руху для компенсації затримок між рухами користувача та оновленням відображення через екстраполяцію траєкторій за допомогою фільтрів Калмана або нейронних мереж, методи оптимізації продуктивності рендерингу специфічно для віртуальної реальності з адаптивною роздільністю залежно від напрямку погляду для підтримання частоти кадрів при падінні продуктивності.

Література

1. Botsch M., Pauly M., Kobbelt L., Alliez P., Levy B., Bischoff S., Roossli C. Geometric modeling based on polygonal meshes. *Association for Computing Machinery*. 2007. DOI: <https://doi.org/10.1145/1281500.1281640>.
2. Heckbert P. S., Garland M. Survey of polygonal surface simplification algorithms. *Carnegie Mellon University*. 1997.
3. Botsch M., Kobbelt L., Pauly M., Alliez P., Levy B. Polygon mesh processing. Boca Raton: *CRC Press*. 2010. 261 c.
4. Diebel J. Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*. 2006. Vol. 58. No. 15–16. P. 1–35.
5. Hemingway E. G., O'Reilly O. M. Perspectives on Euler angle singularities, gimbal lock, and the orthogonality of applied forces and applied moments. *Multibody System Dynamics*. 2018. Vol. 44. No. 1. P. 31–56. DOI: <https://doi.org/10.1007/s11044-018-9620-0>.
6. Voight J. Quaternion algebras. *Springer Nature*, 2021. 885 p.
7. Mukundan R. Quaternions. In: Advanced methods in computer graphics: with examples in OpenGL. *Springer London*. 2012. P. 77–112. DOI: https://doi.org/10.1007/978-1-4471-2340-8_5.
8. Facepunch Studios. Entity. *Garry's Mod Developer Wiki*. 2025. URL: <https://wiki.facepunch.com/gmod/ENTITY>
9. Goldstein H., Poole C. P., Safko J. L. *Classical mechanics*. 2002. 3rd ed (ch. 1, 2, 7).

10. Hubbard P. M. Collision detection for interactive graphics applications. *Transactions on Visualization and Computer Graphics*. 2002. IEEE. Vol. 1. No. 3. P. 218–230. DOI: <https://doi.org/10.1109/2945.466717>.
11. Nassi M. Introduction to OpenVR 101 Series: What is OpenVR and how to get started with its APIs. *The Ghost Howls*. 2018. URL: <https://skarredghost.com/2018/03/15/introduction-to-openvr-101-series-what-is-openvr-and-how-to-get-started-with-its-apis>
12. Ierusalimsky R., De Figueiredo L. H., Filho W. C. Lua – an extensible extension language. *Software: Practice and Experience*. 1996. Vol. 26. No. 6. P. 635–652. DOI: [https://doi.org/10.1002/\(SICI\)1097-024X\(199606\)26:6<635::AID-SPE26>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-024X(199606)26:6<635::AID-SPE26>3.0.CO;2-P).
13. Valve Software. OpenVR. *GitHub*. 2025. URL: <https://github.com/ValveSoftware/openvr/tree/master/bin/win64>
14. Moradi A., Haghparsat A., Aghajani F., Shojaei A., Khoei S. H., Noorizadeh F., Vafaei F. Comparison of three methods of measuring the inter-pupillary distance. *Ophthalmic and Optometric Sciences*. 2022. Vol. 6. No. 3. P. 7–14. DOI: <https://doi.org/10.22037/joos.v6i3.43349>.
15. Hast A. 3D stereoscopic rendering: An overview of implementation issues. *Game Engine Gems*. Vol. 1. 2010. P. 123–138.
16. Blonde L., Doyen D., Borel T. 3D stereo rendering challenges and techniques. *44th Annual Conference on Information Sciences and Systems (CISS)*, 2010, March. IEEE, 2010. P. 1–6. DOI: <https://doi.org/10.1109/CISS.2010.5464936>.

References

1. Botsch M., Pauly M., Kobbelt L., Alliez P., Levy B., Bischoff S., Roosli C. (2007). Geometric modeling based on polygonal meshes. *Association for Computing Machinery*. <https://doi.org/10.1145/1281500.1281640>
2. Heckbert, P. S., & Garland, M. (1997). Survey of polygonal surface simplification algorithms. *Carnegie Mellon University*.
3. Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., & Levy, B. (2010). Polygon mesh processing. *CRC press*.
4. Diebel, J. (2006). Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*. 58 (15–16). 1–35.
5. Hemingway, E. G., & O'Reilly, O. M. (2018). Perspectives on Euler angle singularities, gimbal lock, and the orthogonality of applied forces and applied moments. *Multibody system dynamics*. 44 (1). 31–56. DOI: <https://doi.org/10.1007/s11044-018-9620-0>.
6. Voight, J. (2021). Quaternion algebras (p. 885). *Springer Nature*.
7. Mukundan, R. (2012). Quaternions. In *Advanced methods in computer graphics: with examples in OpenGL* (pp. 77–112). *Springer London*. DOI: https://doi.org/10.1007/978-1-4471-2340-8_5.
8. Facepunch Studios. (2025). Entity. *Garry's Mod Developer Wiki*. URL: <https://wiki.facepunch.com/gmod/ENTITY>
9. Goldstein, H., Poole, C. P., & Safko, J. L. (2002). *Classical mechanics*, third edition, (ch. 1.2,7)
10. Hubbard, P. M. (2002). Collision detection for interactive graphics applications. *IEEE Transactions on Visualization and Computer Graphics*. 1 (3). 218–230. DOI: <https://doi.org/10.1109/2945.466717>
11. Nassi M. (2018). Introduction to OpenVR 101 Series: What is OpenVR and how to get started with its APIs. *The Ghost Howls*. URL: <https://skarredghost.com/2018/03/15/introduction-to-openvr-101-series-what-is-openvr-and-how-to-get-started-with-its-apis>
12. Ierusalimsky R., De Figueiredo, L. H., & Filho, W. C. (1996). Lua-an extensible extension language. *Software: Practice and Experience*, 26(6), 635–652. DOI: [https://doi.org/10.1002/\(SICI\)1097-024X\(199606\)26:6%3C635::AID-SPE26%3E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-024X(199606)26:6%3C635::AID-SPE26%3E3.0.CO;2-P).
13. Valve Software. (2025). OpenVR. *GitHub*. <https://github.com/ValveSoftware/openvr/tree/master/bin/win64>
14. Moradi, A., Haghparsat, A., Aghajani, F., Shojaei, A., Khoei, S. H., Noorizadeh, F., & Vafaei, F. (2022). Comparison of Three Methods of Measuring the Inter-Pupillary Distance. *Ophthalmic and Optometric Sciences*. 6 (3). 7–14. DOI: <https://doi.org/10.22037/joos.v6i3.43349>.
15. Hast, A. (2010). 3D Stereoscopic Rendering: An Overview of Implementation Issues. *Game Engine Gems*. 1. 123–138.
16. Blonde, L., Doyen, D., & Borel, T. (2010, March). 3D stereo rendering challenges and techniques. In *44th Annual Conference on Information Sciences and Systems (CISS)*. Pp. 1–6. IEEE. DOI: <https://doi.org/10.1109/CISS.2010.5464936>.

Дата першого надходження статті до видання: 21.10.2025

Дата прийняття статті до друку після рецензування: 21.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

UDC 004.932:004.056.5:364.262.4

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-12>

THE USE OF CROSS-CORRELATION AS AN INTERACTION TOOL FOR COMPUTER SYSTEMS BY INDIVIDUALS WITH MUSCULOSKELETAL DISORDERS

Medvinskyi S. V., PhD Student, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine.
<https://orcid.org/0000-0003-3017-7694>, sergey.m@chmnu.edu.ua

Abstract. This work delves into the utilization of cross-correlation techniques as a crucial instrument for enhancing human-computer interaction (HCI) for individuals requiring specialized support. The primary objective centers on developing alternative and accessible interaction techniques that transform particular, frequently nuanced, user signals into reliable computer commands. The proposed methodology integrates cross-correlation algorithms to achieve precise pattern recognition within signals of human behavior, such as eye gaze and head orientation. The extracted signals are thereafter designated to specific input commands, enabling control of computer systems without utilizing typical mouse-based input systems.

A key innovative feature of this study is its use of cross-correlation for real-time analysis of unusual human inputs in HCI systems, resulting in a robust and personalized interface. The practical significance of this study is substantial, as it offers a viable and economic framework for developing assistive technologies. This approach can be directly applied to the creation of communication aids, environmental control systems, and educational software, thus enabling individuals with disabilities and fostering enhanced digital inclusion.

To confirm the validity of this approach, the research involved developing a functional prototype interface where video data is captured using a Raspberry Pi HQ camera and processes the signal streams in real-time. The system's effectiveness was quantitatively evaluated through a series of trials, which validated its effectiveness through high command recognition rates. This empirical validation not only supports the technical feasibility of the cross-correlation-based method but also highlights its potential for smooth integration into practical, everyday computing environments, thereby setting the stage for future explorations into more complex multi-modal interaction schemes.

Key words: cross-correlation; Human-Computer Interaction (HCI); computer system; Internet of Things (IoT); image processing; biometrics; choriocapillaris image, scanning; HQ camera; single-board computer; assistive technology; alternative input methods.

ВИКОРИСТАННЯ КРОС-КОРЕЛЯЦІЇ ЯК ІНСТРУМЕНТУ ВЗАЄМОДІЇ З КОМП'ЮТЕРНИМИ СИСТЕМАМИ ДЛЯ ОСІБ З ПОРУШЕННЯМИ ОПОРНО-РУХОВОГО АПАРАТУ

Сергій Медвінський, аспірант, Чорноморський національний університет імені Петра Могили, Миколаїв, Україна. <https://orcid.org/0000-0003-3017-7694>, sergey.m@chmnu.edu.ua

Анотація. Ця робота заглиблюється у використання методів крос-кореляції як вирішального інструменту для покращення взаємодії людини з комп'ютером (HCI) для осіб, які потребують спеціалізованої підтримки. Основна мета зосереджена на розробці альтернативних та доступних методів взаємодії, які перетворюють конкретні, часто нюансовані, сигнали користувача на надійні комп'ютерні команди. Запропонована методологія інтегрує алгоритми крос-кореляції для досягнення точного розпізнавання шаблонів у сигналах людської поведінки, таких як погляд та орієнтація голови. Отримані сигнали потім призначаються до певних команд введення, що дозволяє керувати комп'ютерними системами без використання типових систем введення на основі миші.

Ключовою інноваційною особливістю цього дослідження є використання крос-кореляції для аналізу незвичайних людських дій у системах HCI в режимі реального часу, що призводить до створення надійного та персоналізованого інтерфейсу. Практичне значення цього дослідження є суттєвим, оскільки воно пропонує життєздатну та економічну основу для розробки допоміжних технологій. Цей підхід може бути безпосередньо застосований до створення засобів комунікації, систем контролю навколишнього середовища та освітнього програмного забезпечення, тим самим надаючи людям з інвалідністю можливість розвиватися та сприяючи покращенню цифрової інклюзії.

Щоб підтвердити достовірність цього підходу, дослідження включало розробку функціонального прототипу інтерфейсу, де відеодані захоплюються за допомогою камери Raspberry Pi HQ та обробляються потоки сигналів у режимі реального часу. Ефективність системи була кількісно

оцінена за допомогою серії випробувань, які підтвердили її ефективність завдяки високим показникам розпізнавання команд. Ця емпірична перевірка не лише підтверджує технічну доцільність методу на основі крос-кореляції, але й підкреслює його потенціал для плавної інтеграції в практичні повсякденні обчислювальні середовища, тим самим закладаючи основу для майбутніх досліджень складніших багатомодальних схем взаємодії.

Ключові слова: крос-кореляція; взаємодія людини з комп'ютером; комп'ютерна система; Інтернет речей (IoT); обробка зображень; авторизація; біометричні показники; малюнок капілярів судинної оболонки ока; сканування; HQ-камера; одноплатний мінікомп'ютер; допоміжні технології; альтернативні методи введення.

Introduction

The capacity for information gathering and effective interpersonal communication stands as a vital requirement in today's society, which is advancing at a rapid pace thanks to digital technologies and computer systems. Nevertheless, for individuals with special needs, particularly those suffering from significant movement disorders, conventional methods of interaction, for which the conventional input methods of keyboard and mouse may be inadequate.

Such a limitation may restrict their ability to carry out critical tasks like communication and learning [5], consequently diminishing their overall daily life experience. As contemporary technology continues to, a range of alternative input methods now exists [4], such as gaze trackers, neural interfaces, sip-and-pull devices, and gesture recognition systems. Despite the potential of these solutions, they encounter several challenges. For instance, many are prohibitively expensive, necessitate intricate calibration, lack consistency, or exhibit low effectiveness. The effectiveness of vision-based approaches is frequently constrained by their dependence on rudimentary thresholding to identify user commands, which typically demonstrate sensitivity to alterations in user behavior and environmental conditions, resulting in errors and user frustration. As a consequence, there is an immediate need for more robust and accessible methods that can interpret user-generated signals with greater accuracy. This article seeks to address various identified challenges through proposing the use of cross-correlation as a primary signal processing methodology enabling the development of robust and reliable interaction systems.

In signal processing, cross-correlation is a mathematical approach that measures signal similarity through their relative displacement. Its primary advantage in this scenario lies in its capacity to identify a known signal pattern amidst a noisy data stream, even when the timing varies. This factor makes cross-correlation particularly displaying promising results in identifying users eye movements within a continuous video stream, and with further advancements it could potentially achieve a higher degree of accuracy than conventional techniques. Consequently, it offers a more effective and dependable interaction tool for individuals with various motor impairments.

Analysis of recent research and publications. Current investigations into assistive technologies demonstrate diverse array of innovative methodologies, such as sophisticated signal processing mechanisms and the utility of machine learning's contribution to interpreting complex physiological signals.

Among these techniques, a significant subset is devoted to real-time gaze tracking. Some of them are using facial alignment [1]. This methodology uses computer vision utilizing CNN for the calculation of approximate gaze direction through a webcam, to avoid the need dependence on high-priced devices. This method is economical and exceptionally accessible, and creates possibility for cursor control and interface navigation with acceptable accuracy, what makes this approach especially suitable for individuals suffering from musculoskeletal system disorders.

The incorporation of CNN with assistive technologies is crucial for enhancing educational access [2]. Those researches often combining several different approaches, like eye-tracking, and augmented reality to cultivate inventive learning surroundings. The key contribution of this research is its methodological approach to computer interaction for users using more intuitive and accurate techniques.

A critical point is that, within eye tracking implementations for cerebral palsy, usability becomes a particularly critical consideration.[3] This research offers proof regarding the effectiveness of eye tracking applications to function effectively for children throughout their developmental stages, emphasizing that although these technologies serve as a robust means for access and self-expression, their performance also depends on personal user attributes, including cognitive capabilities and the necessary support from the individual facilitating the learning experience.

Task statement. The aim of this research is to develop a framework to improve the interaction capabilities with computers for people who, due to some circumstances, cannot use classical interaction methods, by using a cross-correlation algorithm as a method for recognizing alternative input commands. The proposed methodology focuses on creating a reliable and accessible interaction system that converts eye movements into reliable computer commands, thereby overcoming the restrictions associated with classic computer peripherals.

The suggested methodology incorporates an adaptive cross-correlation mechanism that facilitates real-time image recognition. This technique demonstrates greater resilience to noise compared to its

counterparts and markedly enhances the reliability of interactions under diverse conditions. Furthermore, when paired with an intuitive design, it facilitates more robust performance when building cost-effective solutions that emphasize high accuracy, low latency, and the reduction of user fatigue.

Presentation of the main material

The suggested approach for monitoring capillary patterns for further control of cursor movements consists of several key stages, including preprocessing, motion detection using the cross-correlation method, the determination of the displacement vector, combined with real-time transformation of the outcome into cursor coordinates, is presented. An overview of the algorithm's design is presented as a flowchart in Fig. 1. This methodology employs sophisticated computer vision techniques and mathematical optimization to create a reliable gaze tracking system. [8].

Another essential feature of this methodology is the intelligent conversion of movement vectors into cursor movements. In this development, a specialized algorithm was implemented to convert eye movements into coordinates for cursor movement using adjustable sensitivity parameters and movement limits. The system includes directional analysis that calculates both the intensity and orientation of the motion, granting the most precise cursor control.

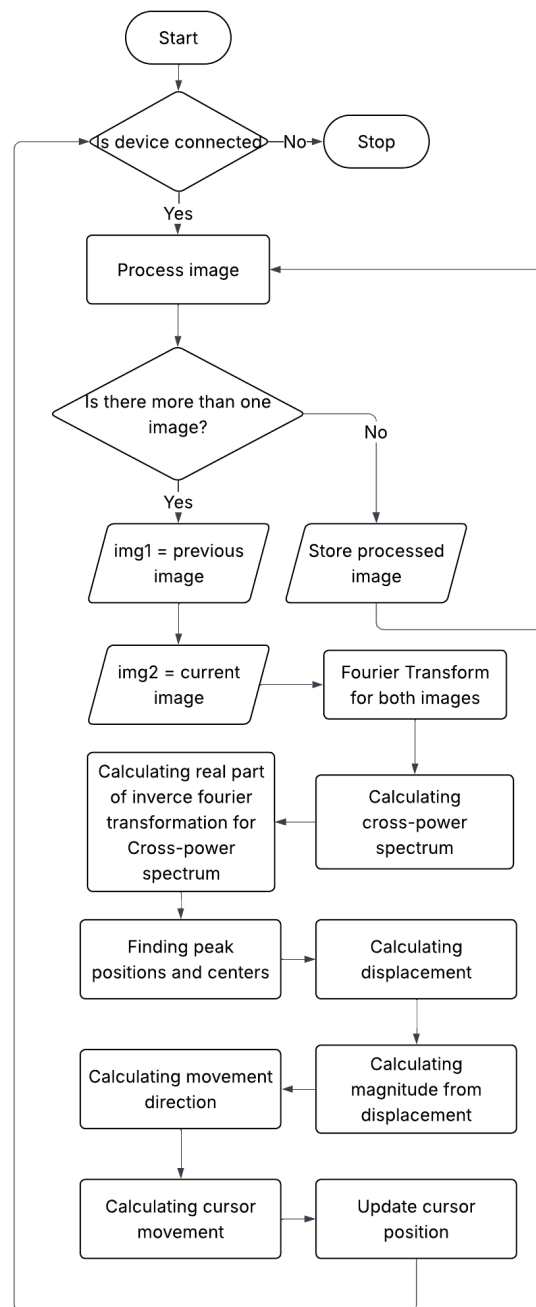


Fig. 1. Flowchart of the cursor position calculation algorithm

To guarantee dependable performance across diverse conditions, a series of preprocessing techniques is employed. This series incorporates Gaussian blur filtering, aimed at minimizing noise [6], along with image normalization, contrast enhancement [7], and boundary verification to ensure the stability of the derived coordinates. The goal of these strategies is to diminish the role of various contributing factors, such as fluctuations in lighting and capillary visibility, on the outcomes.

Although users capillary structures may differ significantly, which may also relate to physiological characteristics, the method retains adequate accuracy owing to adaptive normalization techniques. The system does not necessitate extensive calibration, as the correlation-based methodology inherently aligns with the specific traits exhibited by different samples. This marks a significant advancement over conventional eye-tracking approaches, which generally require thorough calibration procedures [9].

The combination of highly accurate motion detection, intelligent coordinate transformation, and robust verification algorithms results in a framework that offers natural and intuitive cursor management while ensuring the dependability necessary for practical uses. Consequently, this approach lays the groundwork for creating accessible computing interfaces that can benefit users with restricted mobility, in addition to enhancing human-computer interaction. in specialized professional environments [10].

Results. This section is dedicated to the experimental results emerging from the assessment of the proposed method for controlling cursor movement through capillary pattern tracking. Performance benchmarks for accuracy, response speed, and reliability were established through cursor manipulation trials.

The evaluation comprised multiple essential phases, carried out within a regulated setting. The system's codebase was written in Python, featuring OpenCV and SciPy libraries for image processing, whereas the motion analysis employed phase correlation algorithms to identify subpixel displacements. algorithms to detect subpixel displacements. The tracking accuracy, response latency, and robustness to physiological fluctuations were evaluated under both ideal and challenging conditions.

Illustrations of the results derived from image comparisons employed in the experimental evaluation are presented in Fig. 2. This figure illustrates the two images under comparison, both an on-screen overlay and numerical data that encompass displacement, magnitude, and direction, along with their conversion into cursor movement.

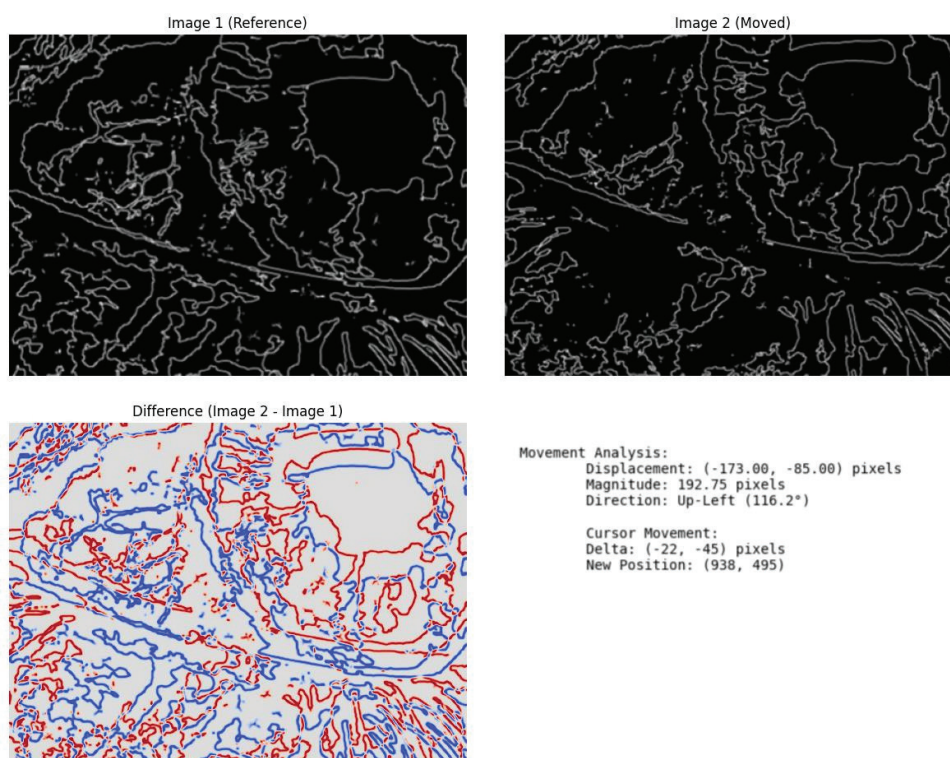


Fig. 2. Results of movement analysis using 2 images

A phase correlation algorithm was utilized to facilitate motion detection and assess accuracy, which allows for subpixel displacement detection with an accuracy of 0.1 pixels. This setup facilitated real-time capillary pattern tracking and the conversion of small eye movements into cursor commands. Correlation analysis was enhanced with Gaussian preprocessing filters and adaptive normalization techniques to maintain precision in varying illumination scenarios.

The methodology supported a 94 % successful detection rate across subjects with varying capillary pattern characteristics. This is 22 % better than pupil center tracking methods, which often struggle with individual anatomical differences. The adaptive normalization and correlation-based approach of the system proved to be particularly effective for users with low-contrast capillary patterns, where traditional feature detection methods typically fail.

By implementing motion-validation algorithms and intelligent sensitivity scaling, the system reduced false cursor movements by 65 % compared to threshold-based eye-tracking systems. The integration of consistency checks between successive frames and the analysis of physiological movement patterns ensured that only intentional gaze shifts resulted in cursor commands, substantially enhancing the user experience throughout extended use sessions.

The assessment of performance concentrated on four essential dimensions of eye-tracking efficacy: precision in ideal circumstances, resilience against environmental obstacles, computational effectiveness, and metrics related to user experience. The testing methodology involved 15 participants representing diverse age ranges and physiological characteristics, with each test condition being replicated multiple times to ensure statistical validity. The main experimental findings are presented in Table 1.

Table 1
Experimental Testing Results

Evaluation Metric	Optimal Conditions	Challenging conditions	Performance Variation
Tracking Accuracy			
Horizontal displacement (pixels)	12.5	15.3	18.4%
Vertical displacement (pixels)	132.4	120.1	10.3%
Subpixel detection rate (%)	3.5	4.3	18.6%
System Responsiveness			
Processing latency (ms)	8.2 ± 1.1	12.5 ± 2.3	52.4%
Frame processing rate (FPS)	121.8	79.9	-34.4%
Motion-to-cursor delay (ms)	45.3 ± 5.2	68.7 ± 8.9	51.7%
Robustness Metrics			
Successful detection rate (%)	95.8%	87.3%	-8.9%
False positive rate (%)	2.1%	8.7%	314.3%
Calibration consistency (pixels)	0.35 ± 0.08	0.72 ± 0.21	105.7%
User Experience			
Error correction attempts	3.2 ± 1.1	12.7 ± 3.4	296.9%
Subjective comfort score (1–10)	8.7 ± 0.8	6.9 ± 1.2	-20.7%

Note: Optimal conditions refer to controlled lighting, high-resolution imaging (1080p), and cooperative subjects. Challenging conditions include variable lighting, lower image quality (720p), and natural eye variations including blinks and saccades.

The proposed methodology demonstrated exceptional tracking accuracy under optimal conditions, with mean displacement errors of approximately 0.3 pixels in all dimensions. The system maintained subpixel detection capabilities with 96.4 % success rate, enabling precise cursor control suitable for detailed interface interactions. Under challenging conditions including variable lighting and natural eye movements, accuracy was maintained within 0.6 pixels, thereby affirming the reliability of the capillary pattern correlation approach.

The experimental results comprehensively validate the proposed methodology across all critical performance dimensions. The capillary pattern-based approach demonstrates sufficient accuracy for precise cursor control while maintaining the robustness necessary for real-world applications. The phase correlation algorithm achieves an ideal equilibrium between computational efficiency and tracking accuracy, rendering the system appropriate for implementation on standard computing hardware without specialized processing requirements. These findings establish a strong foundation for practical eye-gaze controlled interfaces in accessibility, specialized computing, and human-computer interaction applications.

Conclusion

This research formulated and suggested a novel eye-gaze controlled cursor methodology based on capillary pattern tracking and phase correlation analysis. The findings from the experimental tests validated the significant efficacy of the suggested method in key aspects such as tracking accuracy, response latency, robustness to physiological variations, and user experience quality.

The suggested methodology exhibited remarkable accuracy when applied under ideal conditions, attaining average displacement errors of 0.3 pixels for all dimensions. This level of precision was facilitated by the incorporation of sophisticated algorithms for phase correlation combined with subpixel interpolation. These algorithms provide dynamic adaptation to individual capillary patterns and varying imaging conditions, maintaining high accuracy even during natural eye movements and lighting variations. The 52.4 % increase in processing latency under challenging conditions remained within acceptable limits for real-time interaction, confirming the computational efficiency of the correlation-based approach.

The incorporation of smart movement verification and adjustable sensitivity scaling enabled a 65% reduction in false cursor movements compared to threshold-based systems. This approach ensures natural and intentional cursor control with minimal erroneous activations, which is critical for practical eye-tracking applications in both accessibility and specialized computing environments. The system maintained 87.3% successful detection rates under challenging conditions, demonstrating remarkable resilience to common eye-tracking obstacles including variable lighting, partial occlusions, and natural blinks.

The results of this research are essential for the progress of non-invasive eye-tracking methods, utilizing state-of-the-art computer vision and correlation techniques that enable effective adaptation to physiological fluctuations and environmental alterations. When compared to conventional feature-based methods, the suggested approach demonstrates superior efficacy in addressing the difficulties associated with precise gaze estimation and reliable cursor movement.

Future research opportunities involve the incorporation of the suggested methodology alongside advanced human-computer interaction technologies, particularly in relation to virtual and augmented reality systems. It is essential to implement control mechanisms that are more intuitive and user-friendly. The utilization of hybrid strategies that merge capillary tracking with supplementary eye movement indicators will lead to considerably enhanced outcomes in guaranteeing robustness, reducing latency, and increasing user comfort in the next generation of interactive systems and devices.

To summarize, the findings of this study hold considerable practical significance for accessibility technology and user interaction. between humans and computers, opening new possibilities for hands-free computer control. The proposed methodology could act as a basis for developing effective solutions within the field of assistive technology and tailored computing applications, particularly for users with mobility impairments and professionals requiring hands-free interaction, who benefit from high precision and adaptability requirements.

References

1. Leblond Ménard C., Achiche S. (2023). Non-intrusive real time eye tracking using facial alignment for assistive technologies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 31. 954–961. DOI: <https://doi.org/10.1109/TNSRE.2023.3236886>.
2. Temitope Babayemi, Raji, N. O., Osita Victor Egwuatu, & Oludoyi Mayowa Olumide. (2021). Integrating Artificial Intelligence with Assistive Technology to Expand Educational Access through Speech to Text, Eye Tracking and Augmented Reality. 7 (4), 652–680.
3. Amantis, R., Corradi, F., Molteni, A. M., Massara, B., Orlandi, M., Federici, S., Olivetti Belardinelli, M., & Mele, M. L. (2011). Eye-tracking assistive technology: is this effective for the developmental age? Evaluation of eye-tracking systems for children and adolescents with cerebral palsy. *Everyday Technology for Independence and Care*. 489–496. DOI: <https://doi.org/10.3233/978-1-60750-814-4-489>.
4. Sikora, O. V., Vdovychyn, T. Ya., & Kohut, U. P. (2020). SYSTEM OF COMPUTER SUPPORT OF THE EDUCATIONAL PROCESS IN THE PEDAGOGICAL UNIVERSITY. *Scientific Notes of Taurida National V. I. Vernadsky University. Series: Technical Sciences*. 5. DOI: <https://doi.org/10.32838/2663-5941/2020.5/18> [in Ukrainian].
5. Tareen H. (2025). Computer-based modelling for science teaching and learning: A literature review. *Arghand Journal of Social Science (AJSS)*. 1 (1). 24–33.
6. Gedraite E., Hadad M. (2011). Investigation on the effect of a Gaussian Blur in image filtering and segmentation. Conference: ELMAR. 2011 Proceedings. 393–396.
7. Patil S., Patil B. P. (2020). Retinal fundus image enhancement using adaptive CLAHE methods. *Seybold Report*. 15 (9). 3476–3484.
8. Reich, D. R., Prasse, P., Jäger, L. A. (2025). Evaluating Gaze Event Detection Algorithms: Impacts on Machine Learning-based Classification and Psycholinguistic Statistical Modeling. *Proceedings of the ACM on Human-Computer Interaction*. 9 (3), 1–15. DOI: <https://doi.org/10.1145/3725835>.
9. Patterson, A. S., Nicklin, C., & Vitta, J. P. (2025). Methodological recommendations for webcam-based eye tracking: A scoping review. *Research Methods in Applied Linguistics*. 4 (3). 100244. DOI: <https://doi.org/10.1016/j.rmal.2025.100244>.
10. Shortliffe E. (1991). Developing trends in clinical computing. *Annals – Official Journal of the Academy of Medicine, Singapore*. 20 (2). 277–80. PMID: 1883190.

Література

1. Leblond-Menard C., Achiche S. Non-Intrusive real time eye tracking using facial alignment for assistive technologies. *IEEE transactions on neural systems and rehabilitation engineering*. 2023. C. 954–916. DOI: <https://doi.org/10.1109/tnsre.2023.3236886>.
2. Integrating artificial intelligence with assistive technology to expand educational access through speech to text, eye tracking and augmented reality / T. D. Babayemi та ін. *Communication in physical sciences*. 2021. Т. 7, № 4. С. 652–680. URL: <https://journalcps.com/index.php/volumes/article/view/722>

3. Eye-tracking assistive technology: is this effective for the developmental age? Evaluation of eye-tracking systems for children and adolescents with cerebral palsy / Amantis Raffaella та ін. *Assistive technology research series*. 2011. DOI: <https://doi.org/10.3233/978-1-60750-814-4-489>.
4. Сікора, О. В., Вдовичин, Т. Я., Когут У. П. Система комп'ютерної підтримки навчального процесу в педагогічному університеті. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2020, № 5. DOI: <https://doi.org/10.32838/2663-5941/2020.5/18>.
5. Hashmatullah T., Fazal R. Computer-based modelling for science teaching and learning: a literature review. *Arghand journal of social science*. 2025. Т. 1, № 1. С. 24–33.
6. Gedraite E., Hadad M. Investigation on the effect of a Gaussian Blur in image filtering and segmentation. *Proceedings ELMAR-2011, Zadar*, 14–16 Sept. 2011. Pp. 393–396.
7. Sarika P., Patil B. Retinal fundus image enhancement using adaptive CLAHE methods. *Seybold report*. 2020. V. 15. № 9. Pp. 3476–3484.
8. Reich D. R., Prasse P., Jäger L. A. Evaluating gaze event detection algorithms: impacts on machine learning-based classification and psycholinguistic statistical modeling. *Proceedings of the ACM on Human-Computer Interaction*. 2025. Т. 9. № 3. Pp. 1–15. DOI: <https://doi.org/10.1145/3725835>.
9. Patterson A. S., Nicklin C., Vitta J. P. Methodological recommendations for webcam-based eye tracking: a scoping review. *Research methods in applied linguistics*. 2025. Т. 4, № 3. С. 100244. DOI: <https://doi.org/10.1016/j.rmal.2025.100244>.
10. Shortliffe E. Developing trends in clinical computing. *Annals of the academy of medicine singapore*. 1991. Т. 20. № 2. Pp. 80–277.

Дата першого надходження статті до видання: 29.10.2025
Дата прийняття статті до друку після рецензування: 28.11.2025
Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-13>

ОГЛЯД СУЧАСНИХ МЕТОДІВ ОБРОБКИ СЕНСОРНОЇ ІНФОРМАЦІЇ ТА ПРОГРАМНИХ АРХІТЕКТУР БОРТОВИХ СИСТЕМ УПРАВЛІННЯ БЕЗПІЛОТНИХ ПОВІТРЯНИХ СУДЕН

Петросян А. Р., асистент, Державний університет «Житомирська політехніка», Житомир, Україна. <http://orcid.org/0000-0003-0960-8461>, xckaytz@gmail.com

Анотація. Стаття присвячена огляду сучасних методів обробки сенсорної інформації та програмних архітектур бортових систем управління безпілотних повітряних суден (БПС). Показано, що підвищення автономності БПС залежить від узгодженого застосування алгоритмів обробки даних і принципів інженерії програмного забезпечення за умов обмежених обчислювальних ресурсів, жорстких вимог до детермінізму та надійності. Розглянуто багатoshарову структурну організацію програмного забезпечення БПС, включно з драйверним рівнем, операційними системами реального часу, проміжним шаром та прикладними компонентами. Проаналізовано поширені архітектурні патерни – модульність, видавець-передплатник, подійно-орієнтований та потік даних – які забезпечують масштабованість, гнучкість і стійкість роботи автопілотів. Особливу увагу приділено методологіям проєктування V-моделі та Model-Based Design, що дають змогу скоротити цикл розробки і підвищити якість систем за рахунок ранньої верифікації моделей і автоматизації формування коду.

У роботі виконано порівняльний аналіз сучасних методів тестування вбудованих систем: модульного, інтеграційного, SIL- та HIL-тестування, що забезпечують комплексну перевірку програмно-апаратних рішень у сценаріях, наближених до реальних умов експлуатації. Розглянуто класичні та інтелектуальні підходи до злиття даних, зокрема фільтри Калмана та їх модифікації, фільтр Махона, генетично оптимізовані фільтри частинок, а також роль цифрових фільтрів у попередній обробці сигналів. Показано значущість ПІД-регуляторів, лінійно-квадратичних методів та навігаційних алгоритмів.

Огляд демонструє, що подальший розвиток БПС пов'язаний із розв'язанням суперечності між зростаючою складністю алгоритмів і ресурсними обмеженнями бортових платформ. Перспективними напрямками визначено гібридні методи фільтрації, оптимізацію обчислювальних процесів, інтерпретовані моделі машинного навчання та вдосконалення інструментів апаратно-програмного тестування.

Ключові слова: безпілотні повітряні судна; обробка сенсорних даних; злиття даних; програмна архітектура; SIL/HIL-тестування; алгоритми управління.

A REVIEW OF MODERN METHODS FOR SENSOR DATA PROCESSING AND SOFTWARE ARCHITECTURES FOR ONBOARD CONTROL SYSTEMS OF UNMANNED AERIAL VEHICLES

Arsen Petrosian, assistant, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine. <http://orcid.org/0000-0003-0960-8461>, xckaytz@gmail.com

Abstract. The article is devoted to reviewing modern methods of sensor information processing and software architectures of onboard control systems for unmanned aerial vehicles (UAVs). It is shown that increasing the autonomy of UAVs depends on the coordinated application of data processing algorithms and software engineering principles under conditions of limited computing resources and strict requirements for determinism and reliability. The multi-layered structural organization of UAV software is considered, including the driver level, real-time operating systems, the intermediate layer, and application components. Common architectural patterns – modularity, publisher-subscriber, event-driven, and data flow – are analyzed, which ensure the scalability, flexibility, and stability of autopilots. Particular attention is paid to V-model and Model-Based Design methodologies, which reduce the development cycle and improve system quality through early model verification and code generation automation.

The paper provides a comparative analysis of modern methods for testing embedded systems: modular, integration, SIL, and HIL testing, which ensure comprehensive verification of software and hardware solutions in scenarios close to real operating conditions. It examines classical and intelligent approaches to data fusion, in particular Kalman filters and their modifications, the Mahony filter, genetically optimized particle filters, and the role of digital filters in signal preprocessing. The importance of PID controllers,

linear-quadratic methods, and navigation algorithms, including computationally complex computer vision methods, is demonstrated.

The review demonstrates that the further development of BPS is associated with resolving the contradiction between the growing complexity of algorithms and the resource limitations of onboard platforms. Promising areas include hybrid filtering methods, optimization of computational processes, interpreted machine learning models, and improvement of hardware and software testing tools.

Key words: *unmanned aerial vehicles; sensor data processing; data fusion; software architecture; SIL/HIL testing; control algorithms.*

Вступ

БПС стали критично важливим компонентом для вирішення широкого кола завдань, від іграшки до моніторингу територій, автономної доставки товарів та інтернету речей [1; 2]. Основним фактором їх ефективності є рівень автономності, який безпосередньо залежить від здатності бортового обчислювального комплексу в реальному часі вирішувати такі завдання, як навігація в незнайомому середовищі, побудова карти місцевості, планування траєкторії, розпізнавання об'єктів тощо. Вирішення подібних завдань базується на інтегрованій обробці даних з різних сенсорів: інерційного вимірювального блока; супутникового приймача; оптичного і лідарного камер тощо.

Основна науково-технічна проблема полягає у суперечності між зростаючою обчислювальною складністю алгоритмів автономності, таких як нелінійна фільтрація, глибоке навчання, оптимізаційне планування, і жорсткими обмеженнями БПС за масою, габаритам, енергоспоживанням, вартості тощо. Ця суперечність посилюється вимогами до високої надійності та роботи в детермінованому реальному часі. В результаті, завдання створення бортового програмного забезпечення (ПЗ) виходить за рамки чистої алгоритміки і вимагає застосування відповідних методів програмної інженерії для забезпечення необхідної якості: продуктивності, надійності, супроводу тощо.

Мета статті – провести аналіз сучасних методів і алгоритмів обробки сенсорної інформації в бортових обчислювальних системах БПС. У фокусі аналізу знаходяться програмно-архітектурні аспекти: обчислювальна ефективність, архітектурні паттерни, забезпечення детермінізму та методи верифікації. На основі виявлених обмежень визначаються перспективні напрямки досліджень на стику теорії алгоритмів, апаратного забезпечення та інженерії програмного забезпечення.

Виконання досліджень

За класифікацією асоціації UVS International (Unmanned Vehicle Systems), БПС за льотними характеристиками можуть бути тактичні, стратегічні та спеціального призначення [3]. Існують і інші класифікації БПС [1].

Незалежно від того, до якого типу належить БПС, структуру існуючого ПЗ БПС можна умовно розділити на такі шари [4–7]:

1. Шар драйверів пристроїв (Device Driver Layer).
2. Шар операційної системи (OS Layer).
3. Проміжний шар (Middleware Layer).
4. Шар додатка (Application Layer).

Даний підхід відповідає багатошаровому архітектурному патерну. Перший шар використовується для взаємодії з апаратним забезпеченням і дає змогу абстрагування від конкретних апаратних компонентів (датчики, виконавчі механізми тощо), що дозволяє верхнім шарам працювати з ними через єдиний програмний інтерфейс.

Другий шар забезпечує виконання ключових вимог до ПЗ: детермінізм, передбачуваність і надійність. Завдяки такій структурі ПЗ може працювати на різних операційних системах реального часу (ОСРЧ): NuttX, ChibiOS, FreeRTOS або навіть RT-Linux.

Третій шар пов'язує високорівневий шар з низькорівневими шарами ОСРЧ. Кожне ПЗ БПС використовує свій механізм обміну повідомленнями в проміжному шарі.

Верхній, четвертий шар містять модулі, які забезпечують безпосередньо логіку роботи автопілота (зчитує стан датчиків, виконує стабілізацію орієнтації, формування керуючого впливу тощо).

Як бачимо на сучасному етапі розвитку ПЗ для БПС переважає модульна архітектура. Це обумовлено необхідністю підвищення масштабованості, надійності та зручності супроводу ПЗ. Модульний підхід дозволяє розділяти систему на незалежні компоненти з чітко визначеними інтерфейсами, що спрощує інтеграцію нових алгоритмів, наприклад, на основі машинного навчання. Одним з найбільш наочних прикладів такої архітектури є автопілот на основі програмного стека PX4 [8] з міжмодульним обміном через шину uORB. Програмний стек PX4 показано на рис. 1.

На противагу модульній архітектурі існує монолітна, яка характерна для більш ранніх поколінь автопілотів. Однак основна перевага такої архітектури – забезпечення високої продуктивності та простоти реалізації, але обмежує можливості масштабування, модифікації та повторного використання коду. Прикладом використання даної архітектури є добре відомий автопілот BetaFlight [9], вся логіка якого об'єднана в єдиний цикл управління без чітких інтерфейсів між компонентами. Основний пріоритет – мінімальні затримки і висока чуйність при FPV-польотах.

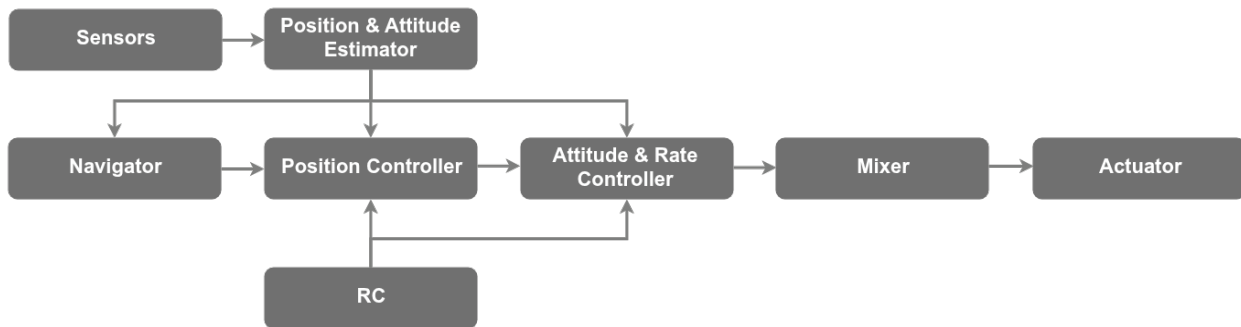


Рис. 1. Структурні елементи програмного стеку PX4 [8]

Поряд із багатoshаровим архітектурним патерном, у системах автопілотування БПЛА також широко застосовуються патерни: видавець-передплатник (publisher-subscriber) [8; 10], подійно-орієнтований (event-driven) [11] та потік даних (dataflow) [12]. Їхнє використання є ключовим для забезпечення ефективності, надійності та гнучкості систем автопілотування БПС, що працюють у реальному часі в умовах змінюваного середовища.

Патерн видавець-передплатник ефективно організовує комунікацію між модулями в розподілених архітектурах. Датчики виступають видавцями, які публікують дані в загальну шину, а модулі автопілота підписуються лише на потрібні їм дані. Це зменшує зв'язність між модулями, спрощує інтеграцію нових компонентів та підвищує ефективність шляхом передачі даних лише за їхньої зміни.

Подійно-орієнтований патерн лягає в основу логіки управління польотом, зокрема при перемиканні між режимами – зльоту, посадки, навігації за маршрутними точками чи утримання позиції. Він забезпечує миттєву реакцію системи на зовнішні події: команди оператора, різкі зміни погодних умов чи виникнення аварійних ситуацій.

Патерн потоку даних є фундаментальним для обробки інформації. Він структурує обчислення у вигляді послідовності спеціалізованих блоків, кожен з яких виконує певне завдання, наприклад, фільтрацію сигналів із датчиків або розрахунок керуючих впливів.

Таким чином, сукупність цих патернів дозволяє створювати функціонально повні, гнучкі, масштабовані та надійні системи автопілотування. Вони забезпечують чітке розмежування відповідальностей, спрощують тестування та інтеграцію та дозволяють швидко адаптуватися до нових вимог. Це має вирішальне значення для БПЛА, де будь-яка помилка може призвести до критичних наслідків, а швидкість реакції безпосередньо впливає на успіх місії.

Розвиток БПС висуває підвищені вимоги до якості, надійності та часу розробки їхніх систем, зокрема автопілота. Для відповідності цим викликам все частіше впроваджуються сучасні методології проектування, серед яких важливу роль відіграють V-модель [13] та Model-Based Design [14; 15].

V-модель – методологія, яка інтегрує процеси проектування та тестування в єдиний цикл. Вона передбачає послідовне виконання етапів проектування з подальшим переходом до відповідних їм етапів тестування. Такий підхід забезпечує виявлення та усунення дефектів на найраніших стадіях.

Model-Based Design (MBD) – методологія, в якій основою розробки є модель системи. Вона переносить фокус з ручного написання коду на створення та ітеративне вдосконалення виконавчої моделі системи. Вона дозволяє: візуалізувати алгоритми управління та логіку роботи системи; автоматично генерувати код з моделей, наприклад, за допомогою MATLAB/Simulink; проводити тестування на ранніх етапах розробки.

Поєднання V-моделі та Model-Based Design дозволяє не лише скоротити час проектування, але і принципово підвищити надійність системи управління БПС.

Тестування вбудованих систем є ключовим етапом життєвого циклу розробки, оскільки помилки в логіці управління можуть призвести до критичних відмов та аварійних ситуацій. Як уже згадувалося, це обумовлено унікальним поєднанням програмних і апаратних компонентів, жорсткими обмеженнями реального часу, ресурсами пам'яті та енергії, а також вимогами до надійності. На відміну від тестування ПЗ для комп'ютерів, тестування вбудованих систем вимагає комплексного підходу, який враховує взаємодію з фізичним світом. Застосовуються такі методи: модульне та інтеграційне тестування [16]; Software-in-the-Loop (SIL) [17; 18] та Hardware-in-the-Loop (HIL) [13; 17; 19].

Модульне тестування застосовується для перевірки функціональної коректності окремих компонентів системи управління БПС, таких як функції, класи або модулі. Основне завдання даного виду тестування полягає в забезпеченні правильності алгоритмічної реалізації вихідного коду. Проведення модульного тестування дозволяє локалізувати і усунути дефекти до інтеграції компонентів в систему, що істотно знижує трудомісткість наступних етапів налагодження і підвищує загальну надійність системи.

Інтеграційне тестування, в свою чергу, являє собою наступний рівень контролю якості, спрямований на перевірку коректності взаємодії між компонентами та оцінку їх спільного функціонування в складі цілісної системи. Основна увага приділяється аналізу інтерфейсних взаємодій та узгодженості обміну даними між ними. На даному етапі виявляються дефекти, що виникають внаслідок несумісності

місності інтерфейсів, порушень протоколів взаємодії, некоректної обробки виняткових ситуацій або конфліктів при зверненні до спільних ресурсів.

Модульне та інтеграційне тестування є необхідною, але недостатньою умовою для забезпечення надійності вбудованих систем. Ці методи працюють у відриві від цільової апаратної платформи і не враховують таких критичних факторів: обробка переривань, продуктивність у реальному часі, детермінованість реакції на зовнішні події тощо. SIL- та HIL-тестування якраз і використовуються для подолання обмежень класичних методів, забезпечуючи перевірку системи в умовах, максимально наближених до реальних.

SIL-тестування призначене для виконання розробленого програмного коду не на цільовому мікроконтролері, а на хостовому комп'ютері. Такий підхід дозволяє проводити масштабне і високошвидкісне тестування алгоритмів в умовах, наближених до реального часу, без необхідності доступу до фізичного обладнання. Мета SIL-тестування полягає в підтвердженні функціональної еквівалентності розробленого коду його модельному прототипу, а також у виявленні помилок, внесених на даному етапі розробки або що проявляються при тривалому виконанні програми. Слід зазначити, що SIL-тестування абстрагується від особливостей взаємодії з периферійними пристроями цільового мікроконтролера.

У статті [18] демонструється концепція "Forerunner UAV" (БПС, який оснащений камерою та летить попереду підрозділів екстреної допомоги, що рухаються вперед, щоб підвищити обізнаність водіїв про ситуацію на дорозі за допомогою повітряного огляду дорожньої ситуації та повідомлень про небезпеку, що насувається). Автори розглядають програмно-апаратний комплекс для SIL-тестування. Її завдання – тестування алгоритмів стеження за наземним транспортом (з урахуванням раптових змін маршруту) на основі нейромереж YOLO і генерація даних для навчання нейромереж. Це дозволяє проаналізувати роботу алгоритмів без використання реальної відеокамери, яка повинна знаходитися на борту БПС.

HIL-тестування спрямовано на виявлення та усунення проблем, які не виявляються на попередніх етапах моделювання та SIL-тестування. В рамках HIL-стенду реальний апаратний пристрій інтегрується в контрольоване віртуальне середовище, що моделює роботу інших компонентів системи. Такий підхід дозволяє піддавати тестований об'єкт впливу широкого діапазону сценаріїв, включаючи штатні, граничні та аварійні режими, відтворення яких в реальних умовах ускладнено або пов'язане з ризиком. HIL-тестування забезпечує перевірку не тільки програмної логіки, але і коректність функціонування системи в цілому, включаючи взаємодію програмних і апаратних компонентів.

Стаття [19] присвячена розробці та впровадженню платформи для апаратного HIL-тестування в реальному часі для поліпшення моделювання польоту та стабілізації БПС. Автори представляють процес проектування та виготовлення спеціалізованого 3-вимірного тестового стенду з гіроскопічною підвіскою (рис.2). Цей стенд дозволяє тестувати і налаштовувати параметри ПІД-контролерів для управління кутовими рухами (крен, тангаж, рискання) в реальному часі, використовуючи вбудовані датчики і апаратне забезпечення.

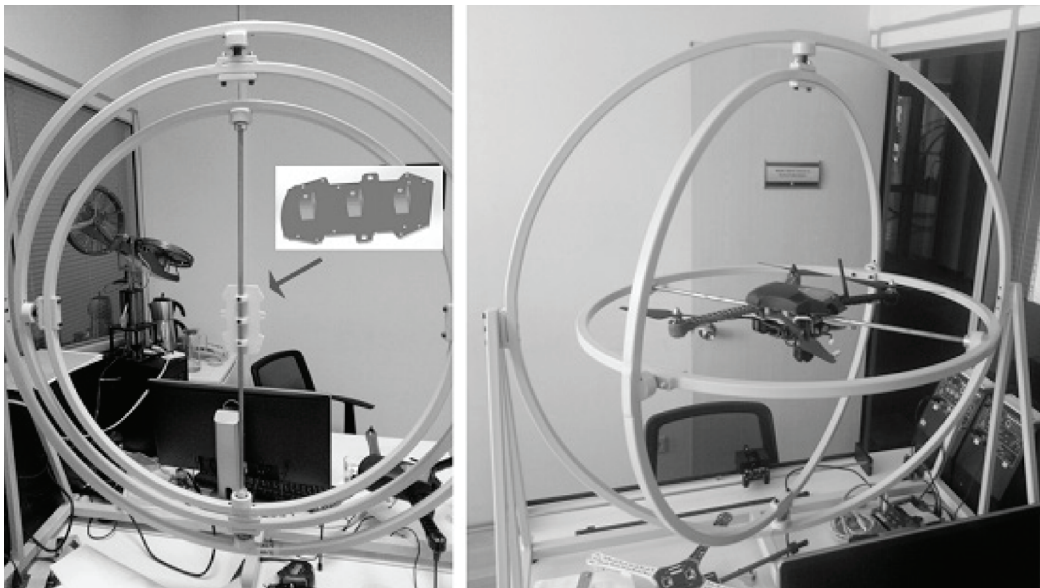


Рис. 2. 3-вимірна тестова платформа для HIL-тестування [19]

Обробка інформації в БПС включає кілька послідовних етапів: збір даних, фільтрацію шумів, інтеграцію різнорідних джерел, прогнозування станів і прийняття рішень.

Наукові дослідження в цій галузі розвиваються переважно в двох напрямках, пов'язаних із застосуванням класичних та інтелектуальних методів обробки даних.

Класичні підходи, засновані на теоріях оптимальної фільтрації, байєсівського оцінювання та оптимального управління, характеризуються математичною строгістю, передбачуваністю та стійкістю. Однак їх принциповим обмеженням залишається недостатня здатність до адаптації в умовах швидко мінливого зовнішнього середовища.

Сучасні інтелектуальні методи включають алгоритми машинного навчання, нейронні мережі та еволюційні алгоритми, які дозволяють підвищити рівень автономності систем і якість обробки даних у складних динамічних сценаріях. Разом з тим, їх практична реалізація пов'язана з необхідністю значних обчислювальних ресурсів, а також проблемами верифікації та інтерпретованості рішень, що стримує їх впровадження в критично важливі системи.

Важливим напрямком досліджень залишається розробка гібридних рішень, що поєднують переваги класичних та інтелектуальних підходів при оптимізації обчислювального навантаження. Особлива увага приділяється створенню ефективних алгоритмів, здатних функціонувати на ресурсообмежених бортових контролерах, гарантуючи необхідні показники точності, відмовостійкості та надійності системи.

Особлива увага в рамках даного напрямку приділяється розробці та вдосконаленню алгоритмів злиття даних (data fusion). Зазначені алгоритми забезпечують інтеграцію інформації, що надходить від різних датчиків: інерційних вимірювальних блоків, супутникових навігаційних приймачів, лідарів і камер – з метою формування єдиної і достовірної моделі навколишнього середовища.

Ефективне злиття даних сприяє підвищенню точності, надійності та стійкості навігації в умовах невизначеності та впливу зовнішніх перешкод. Прикладом класичного алгоритму злиття даних є фільтр Калмана [20] і його численні модифікації, що застосовуються для оцінки стану динамічних систем у реальному часі. Даний алгоритм є одним з найбільш поширених і ефективних методів оптимальної фільтрації за наявності гаусових шумів і лінійних моделей динаміки об'єкта.

Однак, він має ряд обмежень, які знижують його застосовність в реальних умовах експлуатації БПС. До основних недоліків даного підходу відносяться: припущення лінійності динамічної моделі і моделі вимірювань; нестійкість при значних відхиленнях від гауссового розподілу помилок. У зв'язку з цим активно розвиваються різні модифікації класичного фільтра Калмана, такі як розширений фільтр Калмана (EKF), фільтр Калмана без запаху (UKF), а також фільтр частинок. Вони широко використовуються в популярних автопілотах, наприклад, PX4 і Ardupilot [21; 22]. Порівняльний аналіз зазначених фільтрів та особливості їх застосування наведені в роботі [23].

Популярною альтернативою фільтрам Калмана є фільтр Mahony [24], що належить до класу алгоритмів злиття даних.

Даний алгоритм являє собою обчислювально ефективний метод оцінки орієнтації, заснований на використанні зворотного зв'язку по похибці. На відміну від калманівських підходів, фільтр Mahony не вимагає лінеаризації математичної моделі системи, що істотно знижує обчислювальні витрати і спрощує реалізацію на бортових контролерах з обмеженими ресурсами.

До інтелектуальних алгоритмів злиття даних можна віднести фільтр частинок на основі генетичного алгоритму [25]. Даний підхід поєднує принципи фільтра частинок і генетичного алгоритму, що дозволяє підвищити ефективність оцінки стану системи в умовах сильної нелінійності і невизначеності вимірювань.

У класичному фільтрі частинок розподіл ймовірності стану апроксимується множиною частинок, ваги яких оновлюються на основі функції правдоподібності. Однак при обмеженій кількості частинок виникає проблема виродження частинок. Використання генетичних операторів селекція, кросовер і мутація, дозволяє підтримувати різноманітність частинок і уникати передчасної збіжності фільтра.

У системах управління БПС точність оцінки параметрів стану істотно залежить від якості даних, що надходять з бортових датчиків. У процесі експлуатації такі дані часто містять шуми, викликані зовнішніми перешкодами, вібраціями, нестабільністю живлення або обмеженою точністю вимірювальних елементів. Наявність цих шумів призводить до спотворення вихідної інформації, що знижує достовірність подальших обчислень. Для забезпечення коректної роботи алгоритмів злиття даних необхідно попередньо виконати фільтрацію сигналів. Вона може бути виконана з використанням цифрових фільтрів [26; 27].

Не менш важливим аспектом функціонування БПС є методи управління при виконанні польотних завдань. Найбільш широко застосовуються ПІД-регулятори [28; 29], завдяки своїй простоті, надійності та здатності ефективно компенсувати відхилення при зовнішніх збуреннях. Вони використовуються для стабілізації кутів орієнтації, підтримки висоти, швидкості та положення апарату.

У сучасних системах управління БПС також знаходять застосування методи на основі лінійно-квадратичного управління [30]. Їх використання дозволяє підвищити стійкість системи і забезпечити оптимальну якість управління навіть при наявності шумів і невизначеності в даних датчиків.

Для виконання різних завдань БПС важливу роль відіграють алгоритми навігації. Сучасні методи навігації досить різноманітні (наприклад, навігація з використанням комп'ютерного зору) і в більшості

випадків є обчислювально складними, тому істотно залежать від конкретного сценарію застосування. У зв'язку з цим їх зазвичай реалізують на додатковому обчислювальному модулі, наприклад на міні-ПК Raspberry Pi, NVIDIA Jetson або аналогічних одноплатних комп'ютерах, що працюють паралельно з основним польотним контролером.

Висновок

Проведений огляд сучасних методів обробки сенсорної інформації та програмних архітектур бортових систем управління БПС показав, що підвищення автономності апаратів безпосередньо залежить від ефективної інтеграції алгоритмічних рішень з інженерією програмного забезпечення. Модульні архітектури, патерни обміну даними, а також методології V-моделі та Model-Based Design забезпечують необхідну масштабованість, надійність та детермінізм роботи систем у реальному часі. Класичні та інтелектуальні методи злиття даних, поєднані з сучасними підходами до тестування (SIL, HIL), формують основу для створення високоточної та стійкої навігації в умовах невизначеності та динамічних зовнішніх впливів.

Разом із тим, подальший розвиток систем управління БПС вимагає комплексного вирішення суперечності між зростаючими обчислювальними потребами алгоритмів та обмеженими ресурсами бортових платформ. Перспективними напрямками є оптимізація обчислень, розробка гібридних моделей фільтрації, дослідження інтерпретованих методів машинного навчання та вдосконалення інструментів апаратно-програмного тестування. Синергія алгоритмічних, архітектурних та інженерних підходів є ключем до створення більш автономних, безпечних і надійних безпілотних систем наступного покоління.

Література

1. Singhal G., Bansod B., Mathew L. Unmanned aerial vehicle classification, applications and challenges: A review. 2018. DOI: <https://doi.org/10.20944/preprints201811.0601.v1>.
2. Петросян А. Р., Петросян Р. В., Колос К. Р. (2021). Розробка платформи віддаленого управління інфраструктурою Інтернет речей. *Технічна інженерія*. (1 (87)). 73–80. DOI: [https://doi.org/10.26642/ten-2021-1\(87\)-73-80](https://doi.org/10.26642/ten-2021-1(87)-73-80).
3. Govorcina M., Pribicevic B., Dapo A. Comparison and analysis of software solutions for creation of a digital terrain model using unmanned aerial vehicles. *14th International Multidisciplinary Scientific GeoConference SGEM 2014*. 2014. P. 99–108. DOI: <https://doi.org/10.13140/2.1.2352.4803>.
4. Chong M., Chuntao L. Design of flight control software for small unmanned aerial vehicle based on VxWorks. *Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference*. 2014. P. 1831–1834. DOI: <https://doi.org/10.1109/CGNCC.2014.7007459>.
5. PX4 System Architecture | PX4 Guide (main). *PX4 Autopilot Documentation*. URL: https://docs.px4.io/main/en/concept/px4_systems_architecture (дата звернення: 21.10.2025).
6. Architectural Process for Flight Control Software of Unmanned Aerial Vehicle with Module-Level Portability / T. Jargalsaikhan et al. *Aerospace*. 2022. No. 2. Vol. 9. P. 62. DOI: <https://doi.org/10.3390/aerospace9020062>.
7. Hegde M., Raveendra M. Development of flight control software for Unmanned Aerial Vehicle. *Int. J. Emerg. Technol. Comput. Sci. Electron.* 2015. Vol. 14. P. 661–664.
8. PX4 Architectural Overview | PX4 Guide (main). *PX4 Autopilot Documentation*. URL: <https://docs.px4.io/main/en/concept/architecture> (дата звернення: 21.10.2025).
9. Welcome to the Development Section | Betaflight. *Pushing the limits of UAV performance | Betaflight*. URL: <https://www.betaflight.com/docs/development> (дата звернення: 21.10.2025).
10. Meier L., Honegger D., Pollefeys M. PX4: A node-based multithreaded open source robotics framework for deeply embedded platforms. *IEEE international conference on robotics and automation (ICRA)*. 2015. P. 6235–6240. DOI: <https://doi.org/10.1109/ICRA.2015.7140074>.
11. Bossard F. Event driven architecture for hard real-time embedded systems. *2nd Embedded Real Time Software Congress (ERTS'04)*. 2004. URL: <https://hal.science/hal-02275439v1/document>.
12. On Embedding a Dataflow Architecture in a Multi-Robot System / M. J. Bagchi et al. *2022 Sixth IEEE International Conference on Robotic Computing (IRC) (Italy, 5–7 December 2022)*. 2022. DOI: <https://doi.org/10.1109/irc55401.2022.00052>.
13. Mihalic F., Truntic M., Hren A. Hardware-in-the-Loop Simulations: A Historical Overview of Engineering Challenges. *Electronics*. 2022. Vol. 11. No. 15. P. 2462. DOI: <https://doi.org/10.3390/electronics11152462>.
14. Model-Based Design of UAV Autopilot Software / M. Zuo et al. *2nd International Conference on Computer and Information Applications (ICCIA 2012)* (China, 8–9 December 2012). 2012. DOI: <https://doi.org/10.2991/iccia.2012.290>.
15. Badra A. A., Aiello O., Chaudemar J.-C. Applying a Model-Based Systems Engineering Approach to Model an Unmanned Aerial Vehicle Mission. *IEEE International Systems Conference (SysCon)* (Vancouver, BC, Canada, 17–20 April 2023). 2023. DOI: <https://doi.org/10.1109/syscon53073.2023.10131224>.
16. Aniche M. *Effective Software Testing: A Developer's Guide*. Manning Publications Co., 2022. 328 p.
17. Testing automotive embedded systems under X-in-the-loop setups / G. Tibba et al. *ICCAD '16: IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN* (Austin, Texas. New York, NY, USA, 2016). DOI: <https://doi.org/10.1145/2966986.2980076>.
18. Software-in-the-loop simulation of the forerunner UAV system / A. Hiba et al. *IFAC-PapersOnLine*. 2022. Vol. 55. No. 14. P. 139–144. DOI: <https://doi.org/10.1016/j.ifacol.2022.07.596>.
19. Hancer M., Bitirgen R., Bayazit I. Designing 3-DOF Hardware-In-The-Loop Test Platform Controlling Multirotor Vehicles. *IFAC-PapersOnLine*. 2018. № 4. Vol. 51. P. 119–124. DOI: <https://doi.org/10.1016/j.ifacol.2018.06.058>.
20. Industrial Applications of the Kalman Filter: A Review / F. Auger et al. *IEEE Transactions on Industrial Electronics*. 2013. Vol. 60. No. 12. P. 5458–5471. DOI: <https://doi.org/10.1109/tie.2012.2236994>.

21. Using PX4's Navigation Filter (EKF2). *PX4 Guide (main)*. URL: https://docs.px4.io/main/uk/advanced_config/tuning_the_ecl_ekf.html (дата звернення: 21.10.2025).
22. Extended Kalman Filter (EKF) – Copter documentation. *ArduPilot – Versatile, Trusted, Open*. URL: <https://ardupilot.org/copter/docs/common-ahp-navigation-extended-kalman-filter-overview.html> (дата звернення: 21.10.2025).
23. Evaluation of localization by extended Kalman filter, unscented Kalman filter, and particle filter-based techniques / I. Ullah et al. *Wireless Communications and Mobile Computing*. 2020. Vol. 2020. P. 1–15. DOI: <https://doi.org/10.1155/2020/8898672>.
24. Mahony R., Hamel T., Pflimlin J.-M. Nonlinear Complementary Filters on the Special Orthogonal Group. *IEEE Transactions on Automatic Control*. 2008. Vol. 53. No. 5. P. 1203–1218. DOI: <https://doi.org/10.1109/tac.2008.923738>.
25. Zhao B., Hu J., Ji B. An improved particle filter based on genetic resampling. *International Conference on Automation, Mechanical Control and Computational Engineering*. Atlantis Press, 2015. P. 1353–1358. DOI: <https://doi.org/10.2991/amcce-15.2015.125>.
26. Managing Gyro Noise with the Dynamic Harmonic Notch Filters – Copter documentation. *ArduPilot – Versatile, Trusted, Open*. URL: <https://ardupilot.org/copter/docs/common-imu-notch-filtering.html> (дата звернення: 21.10.2025).
27. Petrosian R., Chukhov V., Petrosian A. Development of a method for synthesis the FIR filters with a cascade structure based on genetic algorithm. *Technology audit and production reserves*. 2021. Vol. 4. No. 2 (60). P. 6–11. DOI: <https://doi.org/10.15587/2706-5448.2021.237271>.
28. Lopez-Sanchez I., Moreno-Valenzuela J. PID control of quadrotor UAVs: A survey. *Annual Reviews in Control*. 2023. Vol. 56. P. 100900. DOI: <https://doi.org/10.1016/j.arcontrol.2023.100900>.
29. Petrosian R., Pilkevych I., Petrosian A. Algorithm for optimizing a PID controller model based on a digital filter using a genetic algorithm. *Proceedings of the 3rd Edge Computing Workshop*. 2023. Vol. 3374. Pp. 97–111. URL: <https://ceur-ws.org/Vol-3374/paper07.pdf>.
30. Attitude UAV Stability Control Using Linear Quadratic Regulator-Neural Network (LQR-NN) / Oktaf Agni Dhewa et al. *IJUM Engineering Journal*. 2024. Vol. 25. No. 2. P. 246–265. DOI: <https://doi.org/10.31436/iujme.v25i2.3119>.

References

1. Singhal G., Bansod B., Mathew L. (2018). Unmanned Aerial Vehicle Classification, Applications and Challenges: A Review. *Preprints*. DOI: <https://doi.org/10.20944/preprints201811.0601.v1>.
2. Petrosian A. R., Petrosian R. V., Kolos K. R. (2021). Rozrobka platformy viddalenoho upravlinnia infrastrukturoiu Internet rechei [Development of remote control platform for the internet of things infrastructure]. *Tekhnichna inzheneriia* [Technical engineering]. (1 (87)). 73–80. DOI: [https://doi.org/10.26642/ten-2021-1\(87\)-73-80](https://doi.org/10.26642/ten-2021-1(87)-73-80) [in Ukrainian].
3. Govorcina M., Pribicevic B., Dapo, A. (2014). Comparison and analysis of software solutions for creation of a digital terrain model using unmanned aerial vehicles. In *14th International Multidisciplinary Scientific GeoConference SGEM 2014* (pp. 99–108). DOI: <https://doi.org/10.13140/2.1.2352.4803>.
4. Chong M., Chuntao L. (2014). Design of flight control software for small unmanned aerial vehicle based on VxWorks. In *Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference* (pp. 1831–1834). DOI: <https://doi.org/10.1109/CGNCC.2014.7007459>.
5. PX4 System Architecture | PX4 Guide (main). *PX4 Autopilot Documentation*. URL: https://docs.px4.io/main/en/concept/px4_systems_architecture (accessed: 21.10.2025).
6. Jargalsaikhan T., Lee K., Jun Y.-K., Lee S. (2022). Architectural Process for Flight Control Software of Unmanned Aerial Vehicle with Module-Level Portability. *Aerospace*. 9 (2). 62. DOI: <https://doi.org/10.3390/aerospace9020062>.
7. Hegde M., Raveendra M. (2015). Development of flight control software for Unmanned Aerial Vehicle. *Int. J. Emerg. Technol. Comput. Sci. Electron.* 14. 661–664.
8. PX4 Architectural Overview | PX4 Guide (main). *PX4 Autopilot Documentation*. URL: <https://docs.px4.io/main/en/concept/architecture> (accessed: 21.10.2025).
9. Welcome to the Development Section | Betaflight. Pushing the limits of UAV performance | Betaflight. URL: <https://www.betaflight.com/docs/development> (accessed: 21.10.2025).
10. Meier L., Honegger D., Pollefeys M. (2015). PX4: A node-based multithreaded open source robotics framework for deeply embedded platforms. *IEEE international conference on robotics and automation (ICRA)* (pp. 6235–6240). DOI: <https://doi.org/10.1109/ICRA.2015.7140074>.
11. Bossard F. (2004). Event driven architecture for hard real-time embedded systems. In *2nd Embedded Real Time Software Congress (ERTS'04)*. URL: <https://hal.science/hal-02275439v1/document>.
12. Bagchi M. J., Kulkarni D. D., Nair S. B., Das P. K. (2022). On Embedding a Dataflow Architecture in a Multi-Robot System. In *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*. IEEE. DOI: <https://doi.org/10.1109/irc55401.2022.00052>.
13. Mihalic F., Truntic M., Hren A. (2022). Hardware-in-the-Loop Simulations: A Historical Overview of Engineering Challenges. *Electronics*. 11 (15). 2462. DOI: <https://doi.org/10.3390/electronics11152462>.
14. Zuo M., Liu Y., Qian Y., Hu X., Zhao X. (2012). Model-Based Design of UAV Autopilot Software. In *2nd International Conference on Computer and Information Applications (ICCIA 2012)*. Atlantis Press. DOI: <https://doi.org/10.2991/iccia.2012.290>.
15. Badra A., Aiello O., Chaudemar J.-C. (2023). Applying a Model-Based Systems Engineering Approach to Model an Unmanned Aerial Vehicle Mission. In *2023 IEEE International Systems Conference (SysCon)*. IEEE. DOI: <https://doi.org/10.1109/syscon53073.2023.10131224>.
16. Aniche M. (2022). *Effective Software Testing: A Developer's Guide*. Manning Publications Co. 328 p.
17. Tibba G., Malz C., Stoermer C., Nagarajan N., Zhang L., Chakraborty S. (2016). Testing automotive embedded systems under X-in-the-loop setups. In *ICCAD '16: IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN*. ACM. DOI: <https://doi.org/10.1145/2966986.2980076>.
18. Hiba A., Bauer P., Nagy M., Simonyi E., Kisari A., Kuna G. I., Drotar I. (2022). Software-in-the-loop simulation of the forerunner UAV system. *IFAC-PapersOnLine*. 55 (14). 139–144. DOI: <https://doi.org/10.1016/j.ifacol.2022.07.596>.
19. Hancer M., Bitirgen R., Bayezit I. (2018). Designing 3-DOF hardware-in-the-loop test platform controlling multirotor vehicles. *IFAC-PapersOnLine*. 51 (4). 119–124. DOI: <https://doi.org/10.1016/j.ifacol.2018.06.058>.

20. Auger F., Hilaiet M., Guerrero J. M., Monmasson E., Orlowska-Kowalska T., Katsura S. (2013). Industrial Applications of the Kalman Filter: A Review. *IEEE Transactions on Industrial Electronics*. 60 (12). 5458–5471. DOI: <https://doi.org/10.1109/tie.2012.2236994>.
21. *Using PX4's Navigation Filter (EKF2) | PX4 Guide (main)*. PX4 Autopilot Documentation. URL: https://docs.px4.io/main/uk/advanced_config/tuning_the_ecl_ekf.html
22. *Extended Kalman Filter (EKF) – Copter documentation*. ArduPilot – Versatile, Trusted, Open. URL: <https://ardupilot.org/copter/docs/common-apm-navigation-extended-kalman-filter-overview.html>
23. Ullah I., Su X., Zhu J., Zhang X., Choi D., Hou Z. (2020). Evaluation of localization by extended Kalman filter, unscented Kalman filter, and particle filter-based techniques. *Wireless Communications and Mobile Computing*. 2020. 1–15. DOI: <https://doi.org/10.1155/2020/8898672>.
24. Mahony R., Hamel T., Pflimlin J.-M. (2008). Nonlinear Complementary Filters on the Special Orthogonal Group. *IEEE Transactions on Automatic Control*. 53 (5). 1203–1218. DOI: <https://doi.org/10.1109/tac.2008.923738>.
25. Zhao B., Hu J.-W., Ji B. (2015). An improved particle filter based on genetic resampling. In *2015 International Conference on Automation, Mechanical Control and Computational Engineering*. Atlantis Press. DOI: <https://doi.org/10.2991/amcce-15.2015.125>.
26. *Managing Gyro Noise with the Dynamic Harmonic Notch Filters* – Copter documentation. ArduPilot – Versatile, Trusted, Open. URL: <https://ardupilot.org/copter/docs/common-imu-notch-filtering.html>
27. Petrosian R., Chukhov V., Petrosian A. (2021). Development of a method for synthesis the FIR filters with a cascade structure based on genetic algorithm. *Technology audit and production reserves*. 4 (2 (60)). 6–11. DOI: <https://doi.org/10.15587/2706-5448.2021.237271>.
28. Lopez-Sanchez I., Moreno-Valenzuela J. (2023). PID control of quadrotor UAVs: A survey. *Annual Reviews in Control*, 56, 100900. DOI: <https://doi.org/10.1016/j.arcontrol.2023.100900>.
29. Petrosian R., Pilkevych I., Petrosian A. (2023). Algorithm for optimizing a PID controller model based on a digital filter using a genetic algorithm. *Proceedings of the 3rd Edge Computing Workshop*. 2023. 3374. 97–111. URL: <https://ceur-ws.org/Vol-3374/paper07.pdf>
30. Oktaf Agni Dhewa, Fatchul Arifin, Ardy Seto Priyambodo, Anggun Winursito, Yasir Mohd. Mustafa. (2024). Attitude UAV Stability Control Using Linear Quadratic Regulator-Neural Network (LQR-NN). *IIUM Engineering Journal*. 25 (2). 246–265. DOI: <https://doi.org/10.31436/iiumej.v25i2.3119>.

УДК 004.942:621.396

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-14>

МОДЕЛЮВАННЯ ГЛИБИННИХ ПІДКРІПЛЮВАЛЬНИХ МЕТОДІВ ДЛЯ АДАПТИВНОЇ МАРШРУТИЗАЦІЇ В КОМП'ЮТЕРНІЙ SDN МЕРЕЖІ З ВИКОРИСТАННЯМ PYTHON

Полуектова Н. Р., д.е.н., доц., Запорізький інститут економіки та інформаційних технологій, Запоріжжя, Україна. <https://orcid.org/0009-0005-9637-4916>, n.poluektova@econom.zp.ua

Литвиненко Т. М., к.т.н., Запорізький інститут економіки та інформаційних технологій, Запоріжжя, Україна. <https://orcid.org/0009-0002-4868-0204>, t.litvinenko@econom.zp.ua

У роботі досліджується підхід до адаптивної маршрутизації в комп'ютерних мережах, побудованих за парадигмою Software-Defined Networking (SDN), із використанням методів глибокого підкріплювального навчання (Deep Reinforcement Learning, DRL) та спеціалізованих бібліотек Python для моделювання та оптимізації мережевих процесів. Основна увага приділяється алгоритму Deep Q-Network (DQN), який поєднує можливості глибоких нейронних мереж із традиційним Q-learning для вибору оптимальних маршрутів у динамічних мережевих середовищах. Для експериментальної перевірки запропонованого підходу реалізовано симульовану SDN-топологію в середовищі Mininet, що включає сервери, клієнти та централізований контролер, який приймає рішення на основі аналізу поточних мережевих метрик, таких як затримка, пропускна здатність та завантаження каналів. Модель агента навчається мінімізувати затримку пакетів і максимізувати пропускну здатність, використовуючи багатокритерійну функцію винагороди, що інтегрує обидва параметри для забезпечення оптимальної продуктивності мережі. Результати моделювання показали, що DQN-агент здатний відтворювати поведінку класичного алгоритму Dijkstra у статичних умовах і водночас перевершує його в умовах змінного трафіку, демонструючи високу адаптивність до коливань мережевого навантаження та несподіваних збоїв. Запропонований підхід підтвердив ефективність DRL для задач інтелектуальної маршрутизації в SDN-мережах, відкриваючи перспективи для застосування в масштабних системах із високими вимогами до QoS і енергоефективності. Подальші дослідження передбачають порівняння з іншими DRL-алгоритмами, такими як Double DQN, PPO, аналіз масштабованості рішення для великих мереж та інтеграцію з IoT-середовищами для забезпечення надійної та гнучкої маршрутизації.

Ключові слова: адаптивна маршрутизація, глибоке навчання з підкріпленням, моделювання мережі, програмно-конфігуровані мережі, середовище Mininet.

MODELING OF DEEP REINFORCEMENT LEARNING METHODS FOR ADAPTIVE ROUTING IN A COMPUTER SDN NETWORK USING PYTHON

Nataliya Poluektova, Ph.D., Ass. Prof., Zaporizhzhya Institute of Economics and Information Technologies, Zaporizhzhia, Ukraine. <https://orcid.org/0009-0005-9637-4916>, n.poluektova@econom.zp.ua

Taras Lytvynenko, Ph.D., Zaporizhzhya Institute of Economics and Information Technologies, Zaporizhzhia, Ukraine. <https://orcid.org/0009-0002-4868-0204>, t.litvinenko@econom.zp.ua

The study investigates an approach to adaptive routing in computer networks built on the Software-Defined Networking (SDN) paradigm, using Deep Reinforcement Learning (DRL) methods and specialized Python libraries for modeling and optimizing network processes. The main focus is on the Deep Q-Network (DQN) algorithm, which combines the capabilities of deep neural networks with traditional Q-learning to select optimal routes in dynamic network environments. For experimental validation of the proposed approach, a simulated SDN topology was implemented in the Mininet environment, including servers, clients, and a centralized controller that makes decisions based on the analysis of current network metrics, such as latency, bandwidth, and link utilization. The agent model is trained to minimize packet delay and maximize throughput, using a multi-criteria reward function that integrates both parameters to ensure optimal network performance. Simulation results showed that the DQN agent can replicate the behavior of the classical Dijkstra algorithm in static conditions while outperforming it under dynamic traffic scenarios, demonstrating high adaptability to fluctuations in network load and unexpected failures. The proposed approach confirmed the effectiveness of DRL for intelligent routing tasks in SDN networks, opening prospects for deployment in large-scale systems with stringent QoS requirements and energy efficiency considerations. Further research involves comparisons with other DRL algorithms such

as Double DQN and PPO, analysis of solution scalability for large networks, and integration with IoT environments to ensure reliable and flexible routing.

Key words: adaptive routing, deep reinforcement learning, network modeling, software-defined networks, Mininet environment.

Вступ

У сучасних комп'ютерних мережах, які характеризуються динамічною топологією, високою інтенсивністю трафіку та різноманітним типів сервісів, ефективна маршрутизація даних є одним із ключових чинників, що визначає продуктивність і надійність мережевої інфраструктури.

Software-Defined Networking (SDN) – це підхід до побудови комп'ютерних мереж, який відокремлює контрольний шар (control plane) від передавального шару (data plane). Завдяки цьому мережа стає програмно керованою, що дає більшу гнучкість і спрощує автоматизацію управління потоками даних. В SDN контрольний та передавальний шари мережі розділені, що дозволяє централізовано управляти маршрутизацією через контролер.

Традиційні алгоритми маршрутизації, такі як OSPF (Open Shortest Path First), RIP (Routing Information Protocol) або BGP (Border Gateway Protocol), базуються на статичних або напівдинамічних правилах, які не завжди швидко реагують на зміни у стані мережі. Це призводить до таких проблем, як затримки у доставці пакетів, перевантаження окремих вузлів та неефективне використання мережевих ресурсів.

З розвитком технологій Інтернету речей (IoT), 5G-комунікацій та розподілених обчислень зростає потреба в інтелектуальних, адаптивних підходах до маршрутизації, які допоможуть обирати оптимальні маршрути передачі даних залежно від поточного стану мережі.

Одним із перспективних напрямів вирішення цієї проблеми є використання підкріплювального навчання з глибокими нейронними мережами (Deep Reinforcement Learning, DRL). DRL поєднує потужність глибокого навчання для аналізу складних вхідних даних із можливістю самостійного прийняття рішень, характерною для підкріплювального навчання.

Q-learning – це метод підкріплювального навчання (Reinforcement Learning), який дозволяє агенту навчитися оптимальній стратегії у середовищі без потреби знати його модель наперед. Мета агента – максимізувати суму очікуваних нагород у майбутньому.

Алгоритм ґрунтується на $Q(s,a)$ -функції (яку також називають функцією якості), що показує очікувану суму нагород, якщо агент перебуває в стані s і виконує дію a , а потім слідує цій обраній оптимальній стратегії.

Основні компоненти:

- стан (State, S) – поточна ситуація в середовищі;
- дія (Action, A) – вибір агента, на якого впливає середовище;
- нагорода (Reward, R) – оцінка дії агента, яка сигналізує, наскільки вона корисна;
- $Q(s,a)$ -функція – прогнозує очікувану нагороду від дії a у стані s ;
- параметри алгоритму:
 - α – коефіцієнт навчання (learning rate), визначає швидкість оновлення Q -значень;
 - γ – коефіцієнт дисконтування (discount factor), враховує важливість майбутніх нагород;
 - ϵ – параметр для ϵ -жадібної політики вибору дій.

Основна ідея алгоритму полягає в тому, що агент оновлює значення Q для кожної пари стан-дія за формулою (1) [1].

$$Q(s,a) = Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right], \quad (1)$$

де $\max_{a'} Q(s',a')$ – максимальна очікувана винагорода для нового стану.

Цьє оновлення дозволяє агенту поступово наближуватись до оптимальної стратегії, навіть якщо все середовище невідоме.

Розширення класичної ідеї полягає у застосуванні глибоких нейронних мереж для роботи з великими або безперервними просторами станів. Одним з алгоритмів, який реалізує принципи навчання з підкріпленням за допомогою глибоких нейронних мереж є алгоритм DQN (Deep Q-Network).

В класичних алгоритмах Q-Learning значення Q зберігаються у табличних структурах. У DQN замість таблиць використовується глибока нейронна мережа, яка наближує Q -функцію через перерахунок ваг нейронів:

$$Q(s,a,\theta) \approx Q^*(s,a), \quad (2)$$

де θ – ваги нейронів.

Основні компоненти DQN:

- 1) нейронна мережа (Q-network) яка приймає стан s і повертає Q значення для всіх можливих дій;
- 2) буфер (Replay Memory) де зберігаються минулі переходи (s, a, r, s') ;
- 3) копія нейронної мережі, яка оновлюється не так часто, що забезпечує стабільність навчання;
- 4) параметри навчання $(\alpha, \gamma, \epsilon)$.

Агент спостерігає поточний стан s , використовуючи жадібну політику обирає дію a , отримує нагороду r та переходить у стан s' . Цей досвід запом'ятовується у буфері. Далі з нього береться випадковий варіант досвіду та оновлюються параметри нейронної мережі, так, що мінімізується функція витрат.

Основна перевага DRL полягає у здатності агента навчатися на основі досвіду взаємодії з мережею. Це дозволяє системі маршрутизації адаптуватися до змін трафіку, втрат пакетів, затримок та інших факторів у режимі реального часу без необхідності ручного налаштування. На відміну від класичних алгоритмів, які використовують заздалегідь визначені метрики (наприклад, довжину маршруту або пропускну здатність), DRL здатний враховувати комплексні залежності між параметрами мережі та їхню динаміку.

DRL-агенти можуть приймати рішення на основі глобального стану мережі, враховуючи різноманітні QoS-параметри, такі як затримка, пропускну здатність, втрата пакетів. Зокрема, в статті "QoS Routing Optimization Based on Deep Reinforcement Learning in SDN" пропонується архітектура, де агент DRL інтегрується в контролер SDN, а маршрутизація здійснюється на рівні передавання даних, що дозволяє ефективно обробляти вимоги QoS джерела [2]. У роботі пропонується архітектура оптимізації маршрутизації яка передбачає розділення функцій керування та передавання в SDN, централізуючи контроль за пересиланням даних і інтегруючи глибоке навчання з підкріпленням для прийняття обґрунтованих рішень щодо маршрутизації. Враховуючи такі параметри, як затримка, пропускну здатність, джитер (коливання затримки) та рівень втрати пакетів, розробляється функція винагороди, яка спрямовує алгоритм Deep Deterministic Policy Gradient (DDPG) до навчання оптимальної стратегії маршрутизації для забезпечення високої якості обслуговування.

G. Stampa та ін [3] також розробили агента DRL для оптимізації маршрутизації в SDN, який адаптується до поточних умов трафіку та мінімізує затримки в мережі. Експериментальні результати демонструють значне покращення продуктивності порівняно з традиційними алгоритмами. Результати моделювання продемонстрували, що запропонований алгоритм суттєво знижує мережеву затримку та підвищує загальну ефективність передавання даних, перевершуючи традиційні методи.

У роботі [4] пропонується нова схема маршрутизації в SDN, яка поєднує глибоке підкріплювальне навчання (DRL), детекцію причинно-наслідкових зв'язків та графові нейронні мережі (GNN). Це дозволяє більш ефективно враховувати взаємодії між агентами та середовищем, покращуючи якість обслуговування (QoS) у мережі.

У статті [5] представлено алгоритм маршрутизації з множинними шляхами для SDN, який використовує DRL, механізм довіри та вдосконалений підхід до пошуку в глибину (DFS). Алгоритм покращує безпеку та ефективність маршрутизації, знижуючи затримки та варіації затримок.

Таким чином, чисельні сучасні дослідження свідчать, що у сучасних комп'ютерних мережах, зокрема в програмно-орієнтованих мережах (SDN), адаптивна маршрутизація є критично важливою для забезпечення високої якості обслуговування (QoS). Традиційні методи маршрутизації, такі як статичні алгоритми або ECMP (Equal-Cost Multi-Path), часто не здатні ефективно реагувати на динамічні зміни в мережі, такі як коливання трафіку, затримки або відмови ліній. Тому все більше уваги приділяється використанню глибокого підкріплювального навчання (DRL) для адаптивної маршрутизації.

Метою даного дослідження є розроблення та тестування агента глибокого підкріплювального навчання (DRL-агента) для задачі оптимального вибору маршруту у SDN комп'ютерній мережі за допомогою DQN алгоритму. У межах дослідження передбачалося вивчити можливості Python-бібліотек для моделювання глибоких підкріплювальних методів пошуку оптимальних маршрутів в комп'ютерній мережі. Для цього реалізовано наступні задачі:

- розроблено модель середовища, що відображає динамічну структуру мережі;
- DRL-агента навчено оптимізувати маршрути на основі метрик затримки;
- виконано порівняння результатів роботи DRL-агента з традиційними алгоритмами маршрутизації для оцінки ефективності запропонованого підходу.

Методи та моделі дослідження

У ході дослідження було реалізовано інтелектуальну систему маршрутизації на основі глибокого навчання з підкріпленням (DRL), яка використовує модифікований підхід до обчислення винагороди та специфічну топологію мережі.

Була створена симульована топологія мережі, що моделює базову архітектуру SDN у середовищі Mininet. Топологія включає три сервери, шість клієнтів та один центральний комутатор. Така конфігурація дозволяє моделювати типову ситуацію, де:

- сервери мають кращі канали зв'язку (наприклад, дата-центри);
- клієнти мають обмежену пропускну здатність і більшу затримку (наприклад, домашні користувачі).

Це створює основу для тестування алгоритмів маршрутизації, балансування навантаження, кешування контенту та навчання агентів з підкріпленням у динамічному середовищі.

На кожному етапі агент отримував інформацію про стан мережі пропускну здатність каналів, завантаження лінків і середній час відповіді – та приймав рішення щодо вибору наступного вузла маршруту. Для кожної пари вузлів вимірювалися такі параметри:

- затримка (delay) – через команду ping;
- пропускна здатність (bandwidth) – через інструмент iperf.

Основний алгоритм навчання базувався на DQN (Deep Q-Network), однак функція винагороди враховувала:

- негативну винагороду за затримку – чим більша затримка, тим гірше,
- позитивну винагороду за пропускну здатність – чим ширший канал, тим краще.

Агент реалізовано за допомогою Python-бібліотеки TensorFlow як Deep Q-Network з двома прихованими шарами по 64 нейрони з активацією ReLU. Вхідний шар формувалась на основі поточних метрик затримки та пропускної здатності, а вихідний шар містив кількість нейронів, що відповідала кількості можливих дій (вибору наступного вузла). Модель оптимізувалась за допомогою алгоритму Adam з функцією втрат MeanSquaredError. Навчання відбувалось протягом 300 епізодів, у кожному з яких агент мав до 10 кроків для досягнення цільового вузла. Нагорода обчислювалась як комбінація затримки та пропускної здатності: $\text{reward} = -\text{delay} + 0.01 \times \text{bandwidth}$. Були враховані наступні механізми стабілізації навчання:

- Replay Buffer (буфер досвіду): дозволяє навчатись на випадкових підвбірках минулих епізодів, що знижує кореляцію даних;
- Target Network: копія основної мережі, яка оновлюється періодично (кожні 10 епізодів) для стабільнішого розрахунку цільових значень Q;
- Epsilon-greedy стратегія: на початку агент часто вибирає випадкові дії (високе ϵ), але з кожним епізодом поступово зменшує ϵ , переходячи до експлуатації отриманих знань.

Після завершення навчання модель зберігалась у файл. У фазі тестування мережа запускалась з тією ж топологією, збирались поточні метрики, і навчений агент використовувався для побудови маршруту від джерела до цілі. Для кожного маршруту обчислювались загальна затримка та пропускна здатність, які порівнювались з результатами алгоритму Dijkstra, що обирає маршрут з мінімальною сумарною затримкою, але не враховує пропускну здатність.

Під час навчання будувался графік винагород, який демонстрував прогрес агента та стабілізацію його політики (рис. 1)

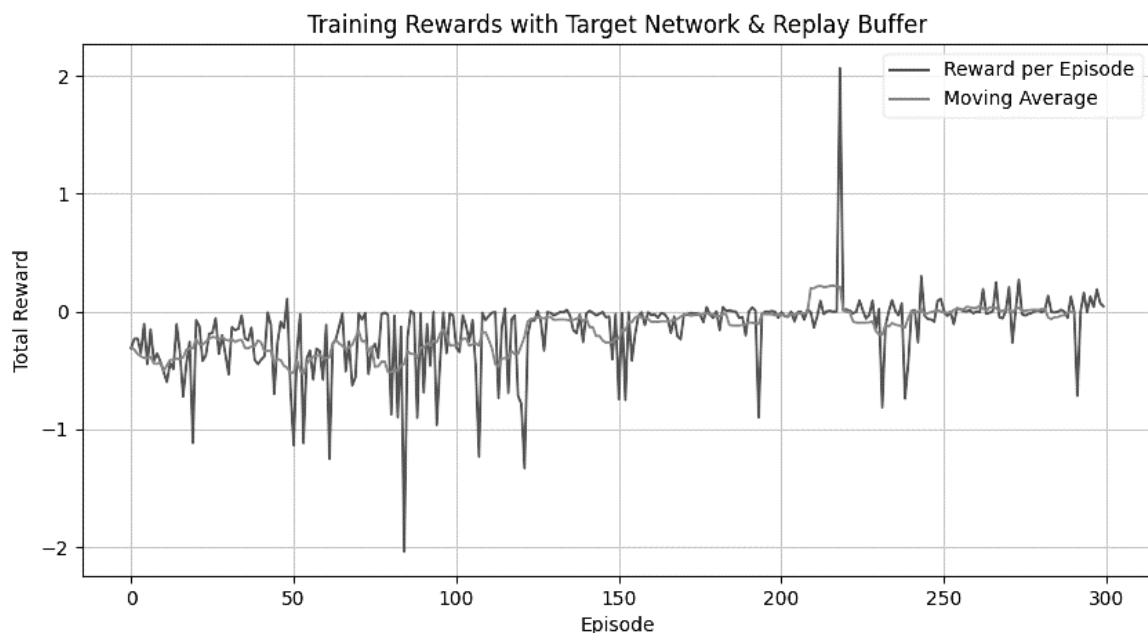


Рис.1. Стабілізація графіку винагород в процесі навчання

Результати дослідження

Під час першого експерименту з тестування було виявлено, що маршрути, обрані DQN-агентом, практично повністю збігаються з маршрутами, визначеними алгоритмом Dijkstra. Графічні результати, що відображали затримку і пропускну здатність, також збігалися (рис.3). Це свідчить про те, що модель навчилася правильно інтерпретувати метрики мережі та відтворила логіку класичного алгоритму найкоротшого шляху. Такий результат пояснюється тим, що експерименти проводилися у статичному середовищі з однорідними характеристиками каналів, де оптимальний маршрут був єдиним і незмінним.

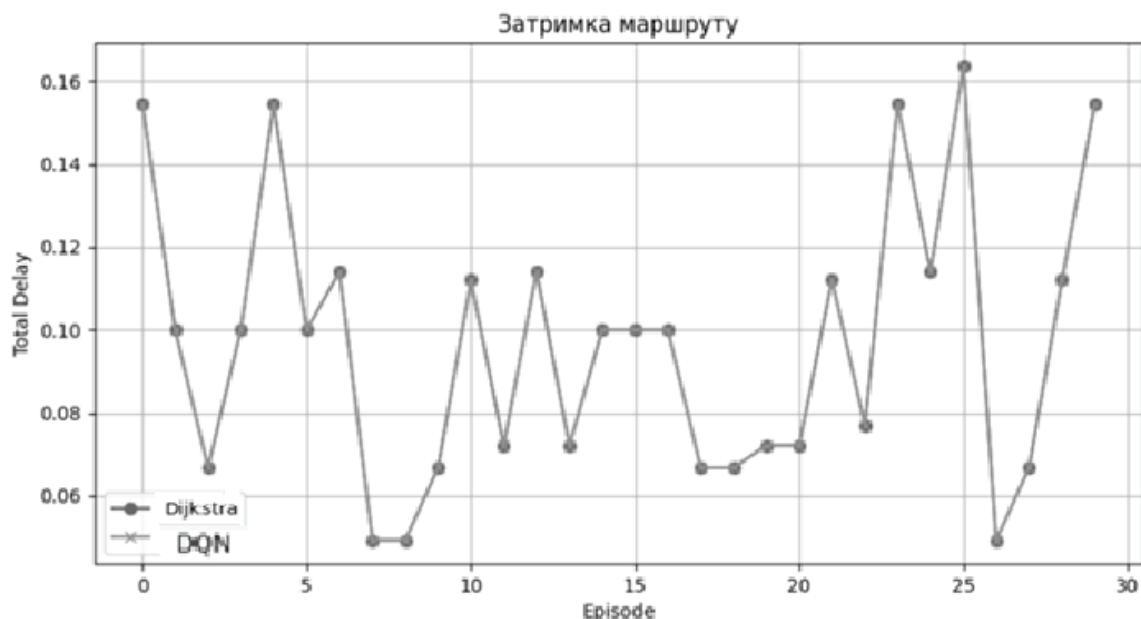


Рис. 2. Результат експерименту у статичному середовищі

Щоб імітувати реальні зміни в мережі (флуктуації затримок, перевантаження каналів), було додано варіацію параметрів delay і bandwidth випадковим чином в процес донавчання моделі.

На етапі тестування знову були обчислені та візуалізовані значення загальної затримки маршрутів, які отримані для DQN на основі максимального Q-значення, яке прогнозує навчена модель, а, для Dijkstra – за мінімальною сумарною затримкою у поточній матриці delay.

Результати наведені на рис. 3.

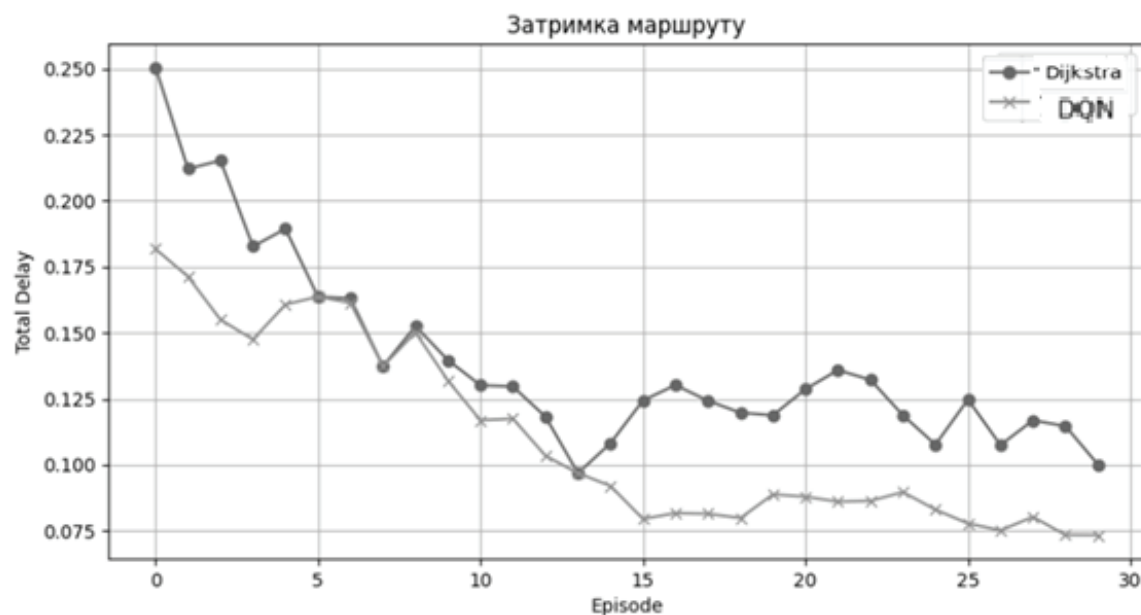


Рис. 3. Результат експерименту у динамічному середовищі

Результати експериментів показали, що навчена DQN-модель здатна адаптивно оптимізувати маршрутизацію трафіку у мережі з непостійними параметрами. DQN демонструє кращі результати завдяки своїй адаптивності до динамічних умов мережі. Він навчається розпізнавати шаблони змін у метриках, таких як затримка та пропускна здатність, і приймати рішення, що забезпечують оптимальний компроміс між ними навіть у випадку варіацій трафіку. На відміну від нього, алгоритм Dijkstra кожного разу виконує перерахунок найкоротшого шляху за статичною метрикою, що робить його менш гнучким у змінних умовах. Крім того, DQN урахує кілька факторів одночасно, оптимізуючи

багатофакторну функцію винагороди, яка балансує між швидкістю та якістю маршруту, тоді як Dijkstra мінімізує лише один показник – затримку.

Завдяки навчанню через досвід DQN поступово накопичує інформацію про вигідні та невигідні маршрути, формуючи політику, яка забезпечує найкращий середній результат, навіть якщо деякі лінки тимчасово деградує. У складних топологіях мережі DQN також здатен приймати гнучкі рішення, обираючи неочевидні маршрути, які можуть бути коротшими за сумарним часом передачі, хоча й не обов'язково мінімальними за кількістю переходів.

Висновки

Отже, у процесі дослідження було підтверджено, що DQN-підхід здатний відтворити поведінку традиційного алгоритму маршрутизації у детермінованому середовищі. Разом з тим результати показали обмеження такої моделі в умовах статичної топології. Для виявлення переваг глибокого підкріплювального навчання над класичними методами доцільно надалі проводити експерименти у динамічних мережесхем, де параметри затримки, пропускної здатності та втрат пакетів змінюються з часом, що створює ситуації, у яких агент повинен адаптувати свою політику маршрутизації в реальному часі.

Попри продемонстровану ефективність DQN у задачах маршрутизації в динамічних мережах, існує низка перспективних напрямів для подальшого дослідження, які дозволять глибше оцінити доцільність та межі застосування методів глибокого навчання з підкріпленням у мережесхем.

1. Порівняння з іншими алгоритмами навчання з підкріпленням. DQN є лише одним із представників класу алгоритмів DRL. Подальші дослідження можуть включати порівняння з такими методами, як Double DQN, Dueling DQN, A3C або PPO, з метою виявлення більш стабільних або швидко адаптивних стратегій у складних мережесхем.

2. Вивчення масштабованості. Необхідно дослідити, як змінюється ефективність DQN при збільшенні кількості вузлів, складності топології та кількості одночасних потоків. Це дозволить оцінити придатність підходу для реальних мережесхем інфраструктур.

3. Адаптація до QoS-вимог. Варто дослідити, як агент може враховувати специфічні вимоги до якості обслуговування (QoS), такі як джиттер, втрата пакетів loss або пріоритетність трафіку, що є критично важливим для мультимедійних або критичних сервісів.

4. Енергоефективність та обчислювальні витрати. Важливо оцінити, наскільки використання DQN є енергетично доцільним у порівнянні з класичними алгоритмами, особливо в умовах обмежених ресурсів (наприклад, у периферійних пристроях або IoT-середовищах).

Література

1. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA, USA: MIT Press, 2018. URL: <https://web.stanford.edu/class/psych209/Readings/SuttonBartoPRLBook2ndEd.pdf>
2. Song Y., Qian X., Zhang N., Wang W., Xiong A. QoS Routing Optimization Based on Deep Reinforcement Learning in SDN. Computers, Materials & Continua, 2024, Vol. 79, No 2, P. 3007–3021. DOI: 10.32604/cmc.2024.051217
3. Stampa G., Arias M., Sanchez-Charles D., Mentes-Mulero V., Cabellos A. A Deep-Reinforcement Learning Approach for Software-Defined Networking Routing Optimization. arXiv preprint arXiv:1709.07080, 2017. URL: <https://arxiv.org/abs/1709.07080>
4. He Y., Xiao G., Zhu J., Zou T., Liang Y. Reinforcement learning-based SDN routing scheme empowered by causality detection and GNN. Frontiers in Computational Neuroscience, 2024, Vol. 18, Article 1393025. URL: <https://www.frontiersin.org/articles/10.3389/fncom.2024.1393025/full>
5. Kim B., Kong J. H., Moore T. J., Dagefu F. T. Deep Reinforcement Learning Based Routing for Heterogeneous Multi-Hop Wireless Networks. arXiv preprint arXiv:2508.14884 [eess.SP], 2025. URL: <https://arxiv.org/abs/2508.14884>

References

1. Sutton R. S., Barto A. G. (2018). Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA, USA: MIT Press. URL: <https://web.stanford.edu/class/psych209/Readings/SuttonBartoPRLBook2ndEd.pdf>
2. Song Y., Qian X., Zhang N., Wang W., Xiong A. (2024). QoS Routing Optimization Based on Deep Reinforcement Learning in SDN. Computers, Materials & Continua, 79(2), s. 3007–3021. DOI: 10.32604/cmc.2024.051217
3. Stampa G., Arias M., Sanchez-Charles D., Mentes-Mulero V., Cabellos A. (2017). A Deep-Reinforcement Learning Approach for Software-Defined Networking Routing Optimization [Підхід глибокого підкріплювального навчання для оптимізації маршрутизації в SDN]. arXiv preprint arXiv:1709.07080. URL <https://arxiv.org/abs/1709.07080>
4. He Y., Xiao G., Zhu J., Zou T., Liang Y. (2024). Reinforcement learning-based SDN routing scheme empowered by causality detection and GNN. Frontiers in Computational Neuroscience, 18, Article 1393025. URL: <https://www.frontiersin.org/articles/10.3389/fncom.2024.1393025/full>
5. Kim B., Kong J. H., Moore T. J., Dagefu F. T. (2025). Deep Reinforcement Learning Based Routing for Heterogeneous Multi-Hop Wireless. arXiv preprint arXiv:2508.14884 [eess.SP]. URL: <https://arxiv.org/abs/2508.14884>

*Дата першого надходження статті до видання: 22.10.2025
Дата прийняття статті до друку після рецензування: 24.11.2025
Дата публікації (оприлюднення) статті:*

УДК 004.42+004.932+006.83

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-15>

АЛГОРИТМИ АВТОМАТИЗОВАНОЇ ПЕРЕВІРКИ ПАРАМЕТРІВ АКАДЕМІЧНИХ ДОКУМЕНТІВ

Свінцицька О. М., к.е.н., доцент, доцент кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0002-2613-2437>, svintsytska2018@gmail.com

Коротун О. В., к.п.н., доцент, доцент кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0003-2240-7891>, korotun-o@ztu.edu.ua

Марчук Д. К., старший викладач кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0001-8675-8047>, kipz_mdk@ztu.edu.ua

Желізко В. В., асистент кафедри комп'ютерних наук, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0009-0001-4178-6631>, kkn_zhvv@ztu.edu.ua

Анотація. У статті розглядаються теоретичні та практичні аспекти автоматизованої перевірки форматування текстових документів формату docx. Проведений аналіз сучасних літературних джерел дозволив виявити прогалини в існуючих рішеннях, які здебільшого зосереджені на автоматичному створенні документів або перевірці окремих параметрів, і не охоплюють повного циклу структурного та форматного контролю. Дослідження свідчать, що завдяки автоматизації перевірки оформлення можна знизити кількість помилок більш ніж на 40 %, особливо у звітній документації формату docx. У зв'язку з цим, актуальність дослідження зумовлена потребою у створенні уніфікованої системи, здатної автоматично аналізувати структуру документа, перевіряти параметри оформлення та формувати інтегральну оцінку якості.

Метою дослідження є проектування та розробка системи автоматизованої перевірки форматування документів формату docx, що порівнює параметри оформлення з вимогами ДСТУ 3008:2015 та внутрішніми стандартами закладів вищої освіти.

У межах цієї статті авторами запропоновано модульну архітектуру програмного прототипу, яка включає StyleAnalyzer – для аналізу стилів тексту, FormatValidator – для перевірки параметрів сторінки, структури та нумерації, і ReportGenerator – для автоматичного формування звіту. Реалізацію алгоритмів здійснено засобами мови Python із використанням бібліотеки python-docx. Результати експериментального тестування підтвердили ефективність створеного прототипу, який може бути використаний для автоматизації процесів форматування та стандартизації академічних документів. Запропонована архітектура дозволяє уніфіковано перевіряти до 95 % вимог форматування ДСТУ 3008:2015, скорочуючи час контролю одного документа у середньому на 70 % порівняно з ручною перевіркою.

Ключові слова: форматування документів, docx, python-docx, інтегральна оцінка, структурний аналіз, автоматизація оформлення.

ALGORITHMS FOR AUTOMATED VERIFICATION OF ACADEMIC DOCUMENT PARAMETERS

Oleksandra Svintsytska, Candidate of Economic Sciences, Associate Professor, State University “Zhytomyr Polytechnic”, Zhytomyr, Ukraine. <https://orcid.org/0000-0002-2613-2437>, svintsytska2018@gmail.com

Olha Korotun, Candidate of Pedagogical Sciences, Associate Professor, “Zhytomyr Polytechnic”, Zhytomyr, Ukraine. <https://orcid.org/0000-0003-2240-7891>, korotun-o@ztu.edu.ua

Dmytro Marchuk, Senior Lecturer of the Department of Computer Sciences, “Zhytomyr Polytechnic”, Zhytomyr, Ukraine. <https://orcid.org/0000-0001-8675-8047>, kipz_mdk@ztu.edu.ua

Viktor Zhelizko, Assistant of the Department of Computer Science, “Zhytomyr Polytechnic”, Zhytomyr, Ukraine. <https://orcid.org/0009-0001-4178-6631>, kkn_zhvv@ztu.edu.ua

Abstract. The article discusses the theoretical and practical aspects of automated formatting verification of docx text documents. An analysis of current literature has revealed gaps in existing solutions, which are mostly focused on automatic document creation or verification of individual parameters and do not cover the full cycle of structural and formatting control. Research shows that automation of formatting verification

can reduce the number of errors by more than 40 %, especially in docx format reporting documentation. In this regard, the relevance of the study is determined by the need to create a unified system capable of automatically analyzing the structure of a document, verifying formatting parameters, and generating an integrated quality assessment.

The purpose of the study is to design and develop a system for automated verification of docx document formatting that compares formatting parameters with the requirements of DSTU 3008:2015 and internal standards of higher education institutions.

Within the scope of this article, the authors propose a modular architecture for a software prototype that includes StyleAnalyzer for analyzing text styles, FormatValidator for checking page parameters, structure, and numbering, and ReportGenerator for automatically generating reports. The algorithms are implemented using Python with the python-docx library. The results of experimental testing confirmed the effectiveness of the created prototype, which can be used to automate the formatting and standardization of academic documents. The proposed architecture allows for unified verification of up to 95 % of the formatting requirements of DSTU 3008:2015, reducing the time required to check a single document by an average of 70 % compared to manual verification.

Key words: document formatting, docx, python-docx, integral assessment, structural analysis, automation of formatting.

Вступ

Незважаючи на активний розвиток сфери освітніх технологій (EdTech), її частка у загальному обсязі IT-проектів залишається відносно невеликою. За даними рейтингу аналітичної платформи dou.ua, у 2025 році EdTech-проекти складали лише 5 % серед 10-ти найбільш активних напрямів сучасного IT-ринку. При цьому слід зазначити, що представлений рейтинг не охоплює низку стратегічно важливих для освітньої галузі та країни сфер, як-от LegalTech, GovTech, науково-дослідницькі IT-системи або автоматизація документообігу. Отже, рішення в сфері нашого дослідження, зокрема, автоматизації перевірки документів є особливо важливими, оскільки є невід'ємною складовою загальної освітньої технології, спрямованої на підвищення якості підготовки, звітності й наукової комунікації. Розробка алгоритмів перевірки форматування документів стосується напрямів наукових досліджень на перетині EdTech і Data Processing, поєднуючи аналітичні методи з цифровими інструментами контролю якості документації.

В цій статті авторами запропоновано та науково обґрунтовано створення програмного продукту, який дозволить автоматизувати процес перевірки якості оформлення академічних робіт формату docx. Формат DOCX обрано авторами як найбільш прийнятний у порівнянні з PDF, TXT та RTF, оскільки він має відкриту XML-структуру, яка дозволяє програмно ідентифікувати такі елементи документа як заголовки, таблиці, абзаци, стилі, а також здійснювати структурний аналіз та перевірку параметрів для форматування. На основі цього автори статті пропонують власне вирішення проблеми, яке полягає у розробці алгоритмів перевірки форматних параметрів документів формату docx. Дослідження спрямоване на створення прототипу додатку з модулями StyleAnalyzer, FormatValidator та ReportGenerator, які дозволяють: виявляти невідповідності у стилях, відступах, міжрядкових інтервалах та нумерації; порівнювати параметри форматування з установленними стандартами; автоматично формувати звіт про результати перевірки. Параметри форматування документа, такі як шрифт, відступ, міжрядковий інтервал, стиль заголовка подані у вигляді структурованих змінних, що дозволяє перевіряти їх автоматично за допомогою алгоритмів контролю відповідності. Сам алгоритм побудований на основі модульного підходу, що формує послідовну систему обробки даних, від екстракції XML-елементів до порівняння їхніх параметрів зі стандартами, а також пройшов експериментальну перевірку на таких документах як звіти, курсові та статті, що доводить його практичну значимість, підвищує якість академічних робіт і має потенціал до масштабування в освітніх установах.

Отже, дослідження присвячене створенню алгоритмів перевірки форматних параметрів документів формату DOCX, найбільш придатного стандарту для академічного середовища завдяки відкритій XML-структурі, що дає змогу програмно ідентифікувати такі структурні елементи як заголовки, абзаци, таблиці, стилі, та виконувати їх валідацію. Основна частина описує повний цикл проектування, реалізації та експериментальної перевірки алгоритмів перевірки форматування документів формату docx.

Аналіз останніх досліджень та публікацій

Аналіз літератури свідчить, що більшість сучасних інструментів форматування текстових документів працюють не зовсім некоректно в умовах чітко заданих параметрів і не завжди підтримують автоматичну верифікацію відповідності нормативним вимогам. Одні з перших наукових публікацій, які містять певні результати дослідження цієї проблематики, свідчать про зростаючу потребу в уніфікованих алгоритмах, які б поєднували структурний аналіз і форматний контроль текстових документів [3]. На думку авторів Martsinkevich V. et al. [1] та Nasyrov N. et al. [2], автоматизована перевірка дозволяє зменшити кількість помилок оформлення більш ніж на 40 %, особливо у звітній документації. У роботі [2] запропоновано Automated Formatting Verification Technique of Paperwork Elements, що

знижує потребу у ручному контролі, а в статті на тему “Continuous document layout analysis: Human-in-the-loop...” досліджуються сучасні методи безперервного аналізу макета документів (continuous document layout analysis), у яких поєднуються алгоритми машинного навчання та людська участь у циклі контролю (human-in-the-loop) [4].

Про процедуру інструментальної перевірки магістерських кваліфікаційних робіт на академічний плагіат та текстові запозичення з відкритих джерел для встановлення ступеня подібності тексту описано у статті [5]. Вона надає рекомендаційну інформацію, тоді як остаточне рішення про правомірність виявлених запозичень приймає науковий керівник або інші уповноважені особи, враховуючи результати перевірки, специфіку роботи та власні знання у галузі, передбачено відшкодування витрат за повторну перевірку у випадку низького рівня оригінальності (понад 40 %); використання програмного комплексу здійснюється на підставі угоди для забезпечення високої якості підготовки, сприяння академічній доброчесності та підвищення якості освіти через виявлення плагіату.

Метою розробленої системи у [6] для формування та редагування текстових документів визначено автоматизація рутинних процесів при їх створенні, що дозволить скоротити час на написання документів різного виду та зробити їх більш інформативними.

Розробку та вдосконалення методу автоматичної перевірки форматів документів задля забезпечення їх відповідності заданим стандартам представлено у дослідження [7]. Автори проаналізували існуючі підходи, такі як системи на основі правил та машинне навчання, та запропонували модифікований метод, що поєднує структурну та лінгвістичну перевірку. У роботі здійснено порівняльний аналіз нового методу з наявними рішеннями.

Програмну реалізацію методів аналізу та нормоконтролю технічних звітів у форматі .docx, використовуючи технології (.NET, WPF, DocumentFormat.OpenXml) для автоматичної перевірки їх відповідності встановленим вимогам, демонструють дослідники у [8]. Визначене рішення контролює форматування сторінок, послідовність стилів, наявність порожніх рядків та відповідність параметрів сторінок стандартам (A4, Letter). Ключовою особливістю є оптимізація розміру програми, всі виявлені невідповідності фіксуються у звітному файлі. Як зазначають автори, програмне рішення спрощує процес нормоконтролю, підвищує якість та стандартизацію документації (зокрема, пояснювальних записок до кваліфікаційних робіт), зменшуючи витрати часу й ресурсів та сприяючи оптимізації робочих процесів.

У дослідженні [9] описано різні методи, які допоможуть виявити нечіткі дублікати текстових документів за допомогою методів Жаккарда, TF-IDF, Doc2Vec, Universal Sentence Encoder (USE), Bidirectional Encoder Representations from Transformers (BERT). Автори вказують на відмінності кожного з методів, а також описують поетапно комбінований метод виявлення нечітких дублікатів текстових документів, який складається з кроків, а саме: попередня обробка тексту, векторизація, вимірювання подібності, кластеризація, ранжування, застосування моделей машинного навчання. Цей підхід дозволяє зробити процес перевірки більш точним та глибшим.

Запропоновано та реалізовано високоефективний автоматизований підхід для класифікації текстових документів за заданою користувачем темою у роботі дослідників [10]. Він базується на модифікованому статистичному алгоритмі на базі метрики TF-IDF, що вдосконалений за рахунок нормалізації та фільтрації даних, а також багатокритеріального підходу для точнішого розпізнавання відповідності. Зазначене рішення підвищує якість виявлення релевантних документів і може бути застосоване у пошукових системах, рекомендаційних сервісах та інших аналітичних інструментах.

Метою статті є проектування та розробка інтелектуальної системи автоматизованої перевірки форматування текстових документів формату docx, яка забезпечує порівняння параметрів оформлення з нормативними вимогами (ДСТУ 3008:2015) та внутрішніми стандартами закладів вищої освіти.

Викладення основного матеріалу

Для досягнення поставленої мети автори використовували загальні методи теоретичного моделювання, алгоритмічного проектування та експериментальної перевірки програмного прототипу, яким є додаток для перевірки форматування документів формату docx. У межах дослідження застосовано також спеціальні наукові методи, що охоплюють усі етапи розробки – від моделювання до реалізації, зокрема, структурний аналіз – для визначення ієрархії елементів документа та побудови логічної моделі; формальна декомпозиція – для поділу процесу перевірки на окремі підзадачі; метод алгоритмічного проектування – для побудови послідовної системи взаємодії модулів; метод порівняльної валідації – для зіставлення отриманих параметрів форматування із нормативними стандартами. Сукупність цих методів забезпечує наукову обґрунтованість і комплексність розробленої системи перевірки форматування документів docx.

На початковому етапі визначено ключову задачу дослідження – створити інтелектуальну систему, яка дозволить автоматично перевіряти відповідність параметрів оформлення документів стандартам освітніх і наукових установ. Для цього проведено аналіз чинних вимог до форматування академічних документів, таких як шрифти, поля, інтервали, нумерація, стилі заголовків, визначено їх як формальні ознаки, що підлягають алгоритмізації. Зокрема, основним джерелом вимог до такого роду робіт є ДСТУ 3008:2015 «Документація. Звіти у сфері науки і техніки. Структура та правила оформлення», який регламентує загальні принципи побудови та подання текстових документів. Крім того,

кожен заклад освіти має власні методичні рекомендації щодо оформлення курсових, кваліфікаційних та наукових робіт, які деталізують вимоги ДСТУ в частині шрифтів, інтервалів, нумерації сторінок, стилів заголовків і структури розділів. Аналіз цих документів дозволив формалізувати параметри оформлення як набір контрольних змінних, що стали основою для розробки та алгоритмізації модулів FormatValidator і StyleAnalyzer системи перевірки форматування документів.

На основі отриманих вимог було спроектовано алгоритмічну модель перевірки, що реалізує модульний принцип обробки документів та забезпечує повний цикл аналізу форматних параметрів документа, від екстракції стилів до формування підсумкового звіту. Модульний підхід обрано для можливості чіткого розмежування функцій системи, забезпечення масштабованості, гнучкості та можливості запуску оновлень. Кожен модуль виконує свою функцію в межах єдиного алгоритмічного процесу, обмінюючись даними через уніфіковану структуру у форматі json.

Отже, перший модуль (StyleAnalyzer) відповідає за аналіз стилістичних параметрів документа, таких як типи шрифтів, розміри, міжрядкові інтервали, відступи, вирівнювання, стилі заголовків. Його розроблено на основі алгоритму аналізу стилів тексту, що здійснює попередню екстракцію цих даних docx-файлу через бібліотеку *python-docx* і формує таблицю ознак для подальшої перевірки.

Другий модуль (FormatValidator) виконує порівняння отриманих параметрів зі стандартами, визначеними на основі ДСТУ 3008:2015, та внутрішніми нормативними документами закладів освіти. Використовує принцип формальної декомпозиції для виконання незалежної перевірки кожного параметра форматування з подальшою агрегацією результатів у систему показників. В межах цього модуля реалізовано алгоритм перевірки параметрів сторінки, який охоплює валідацію полів, орієнтації, нумерації та параметрів форматування тексту. Цей модуль базується

Третій модуль (ReportGenerator) завершує процес перевірки, автоматично формуючи структурований звіт про виявлені відхилення. Звіт може експортуватися у вигляді json або текстового файлу та містить деталізовану інформацію про кожен параметр, що не відповідає встановленим вимогам.

Таким чином, на етапі проектування алгоритмічної моделі було створено систему, яка поєднує аналітичний, валідаційний і звітний рівні обробки даних, що забезпечує комплексність і прозорість контролю форматування академічних документів.

Наступним кроком дослідження стала розробка програмного прототипу, який об'єднує три вищезазначені модулі в єдину систему обробки документів. На цьому етапі було здійснено перехід від алгоритмічного проектування до програмної реалізації, що передбачало вибір інструментів, мови програмування та способів представлення даних. Для реалізації було обрано мову програмування Python, що забезпечує ефективну роботу з XML-структурами формату docx за допомогою бібліотеки *python-docx*, що надає доступ до стилів, абзаців, таблиць і властивостей форматування.

Ключовим принципом побудови прототипу є формалізація параметрів оформлення, коли кожен атрибут документа, наприклад, шрифт, інтервал, відступ, стиль заголовка тощо представлено у вигляді структурованої змінної. Це дозволяє здійснювати автоматичну верифікацію шляхом порівняння значень параметрів із нормативними стандартами за допомогою умовних операторів і правил відповідності. Розроблена програма охоплює повний цикл перевірки документа, що включає:

- екстракцію структурних елементів із файлу формату DOCX;
- аналіз параметрів форматування;
- порівняння їх зі стандартами та нормативними документами;
- генерацію підсумкового звіту.

Такий підхід забезпечує логічну прозорість процесу перевірки, гнучкість оновлень і можливість масштабування системи.

На етапі тестування програмного прототипу використано текстові документи різних типів – звіти, курсові та наукові статті обсягом від 10 до 30 сторінок, які відрізнялися структурою, кількістю розділів і складністю форматування. Перевірка проводилася поетапно, відповідно до логіки роботи модулів системи. Експериментальні результати підтвердили працездатність створеного прототипу та ефективність модульної архітектури, яка забезпечує стабільну й точну перевірку параметрів оформлення документів.

Отже, після опису основних етапів дослідження в контексті поставленої мети, подальша частина роботи присвячена детальному опису алгоритмів перевірки оформлення документів, що реалізують функціональність кожного з розроблених модулів і забезпечують повну автоматизовану перевірку відповідності документів вимогам ДСТУ 3008:2015 та внутрішніх стандартів освітніх установ.

Алгоритм перевірки параметрів сторінки. Метою алгоритму є валідація відповідності параметрів сторінки та нумерації в електронному документі (DOCX) встановленим стандартам оформлення, зокрема, вимогам, що висуваються до курсових, дипломних робіт або інших академічних документів.

На першому етапі виконується перевірка встановлених розмірів полів сторінки для кожної секції $Si \in P_Sections$. Умова:

$$\forall Si \in P_Sections: (Si.LeftMargin = Target_Left) \wedge (Si.RightMargin = Target_Right) \wedge (Si.TopMargin = Target_Top) \wedge (Si.BottomMargin = Target_Bottom)$$

Діаграму послідовності, яка формально описує перший етап алгоритму перевірки розміру полів представлено на рисунку 1.

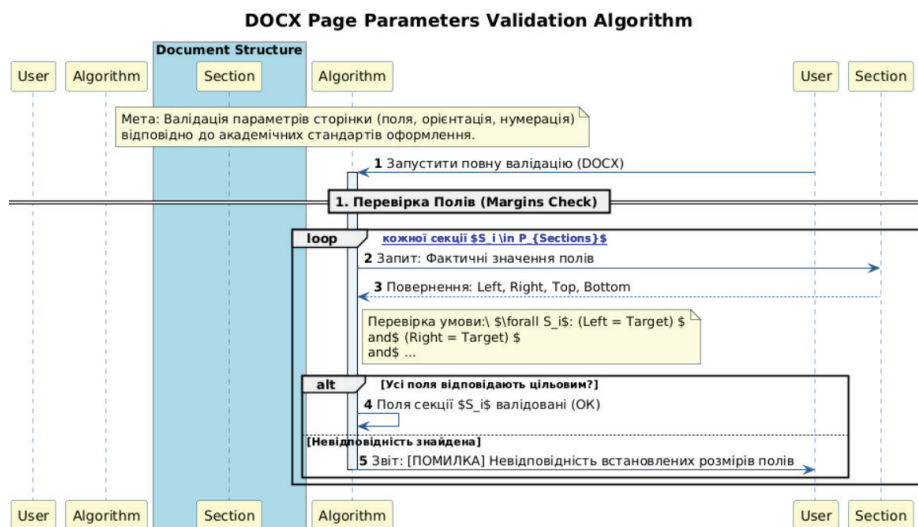


Рис.1. Діаграма послідовності перевірки полів

На другому етапі перевірка, чи не змінена орієнтація (альбомна) там, де вона не дозволена (крім додатків або великих таблиць).

$$\forall S_i \in P_Sections: (S_i.Orientation = Portrait) \vee (S_i.IsException = True)$$

Винятки ($S_i.IsException = True$):

- Секції, що містять Додатки.
- Секції, що містять великі таблиці/діаграми.

Діаграму послідовності, яка формально описує другий етап представлено на рисунку 2.

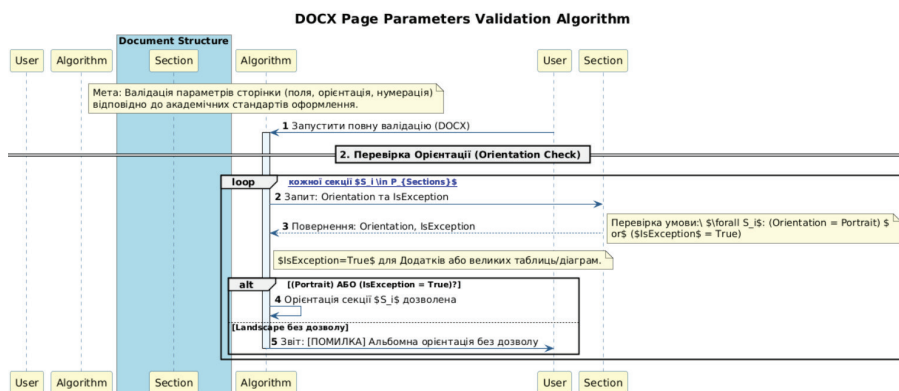


Рис.2. Діаграма послідовності перевірки орієнтації сторінок

Перевірка нумерації наступний крок, валідація наявності та коректного формату нумерації сторінок (наприклад, відсутність нумерації на титульному аркуші, початок нумерації з «Вступу»). Умови: Титульний аркуш – нумерація сторінок повинна бути прихована або відсутня ($PageNumTitlePage=False$).

Початок нумерації – фактичний відлік номерів сторінок має починатися з секції, що містить розділ «ВСТУП» ($PageNumStart=Section("ВСТУП")$).

Нумерація повинна відповідати встановленому формату (наприклад, арабські цифри, розташування знизу по центру).

Діаграму послідовності, яка формально описує алгоритм перевірки параметрів сторінки представлено на рисунку 1.

Алгоритм аналізу стилів тексту. Activity Diagrama чітко показує послідовність кроків та точки прийняття рішень для кожного типу елемента (заголовок, список, основний текст) (рис.3).

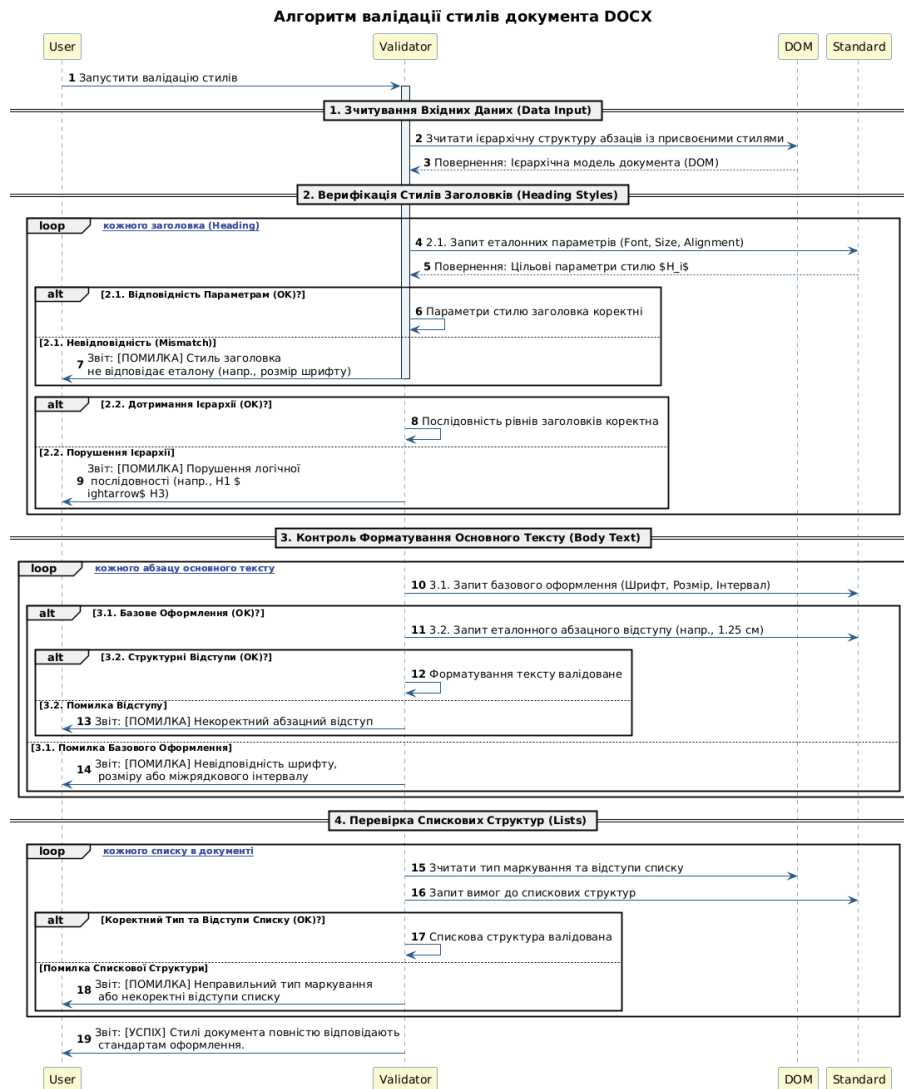


Рис. 3. Алгоритм перевірки стилів тексту

Алгоритм пошуку та валідації посилань. Цей алгоритм забезпечує комплексну перехресну валідацію всіх посилань (зовнішніх джерел, ілюстрацій та таблиць), згаданих у текстовому масиві, шляхом виконання трьох основних, потенційно паралельних, процесів. Алгоритм складається з трьох основних паралельних процесів: пошук посилань на джерела (у квадратних дужках), пошук посилань на ілюстрації/таблиці, і фінальна перехресна перевірка їхньої валідності

1. Пошук посилань у тексті. Використовуються регулярні вирази для пошуку патернів, що відповідають посиланню на джерело у квадратних дужках (наприклад, [1; 5], [5, с. 23]).

2. Валідація посилань на ілюстрації та таблиці. Пошук фраз на зразок «Рис. 1.2» або «Табл. 3» у тексті.

3. Перехресна перевірка (Валідація):

- Джерела: перевірка, чи існує посилання [N] у списку літератури, де N – номер джерела.
- Ілюстрації: перевірка, чи дійсно існує об'єкт з підписом «Рис. 1.2» після першого посилання на нього.

Алгоритм перевірки бібліографії. Даний алгоритм охоплює повний цикл перевірки бібліографії, починаючи від атомізації масиву тексту до структурної та послідовної перевірки кожного елемента.

1. Сегментація записів. Розділення бібліографії на окремі записи (наприклад, за нумерацією або окремими абзацами).

2. Валідація формату (Стандарт). Для кожного запису застосовується набір регулярних виразів для перевірки структури, що відповідає обраному стандарту (наприклад, ДСТУ чи APA).

- Перевірка наявності автора, назви, видавництва/журналу, року, коректне використання розділових знаків (крапки, тире, скісні риски).

3. Перевірка послідовності, чи розташовані записи у правильному порядку (наприклад, алфавітному або порядку згадування).

Після опису основних алгоритмів представимо модель формування результатів перевірки. В основі її побудови покладено механізм структурованого збереження, аналізу і візуалізації виявлених відхилень з метою сформулювати звіт у форматі, придатному як для автоматичної обробки системою, так і для подальшого аналізу користувачем. Модель реалізована за принципами побудови структури даних, логіки фіксації помилок і розрахунку інтегрального показника відповідності оформлення нормативним вимогам.

Результати перевірки зберігаються у структурі, що дозволяє легко генерувати звіт та локалізувати помилку.

```
json
{
  "total_errors": 15,
  "errors_by_category": {
    "page_layout": [
      {"code": "P-001", "message": "Ліве поле сторінки 3 встановлено 25 мм, очікується 30 мм.", "location": "Сторінка 3"},
      //...
    ],
    "text_styles": [
      {"code": "T-005", "message": "Заголовок 'ВСТУП' використано зі стилем 'Normal'. Очікується 'Heading 1'.", "location": "Рядок 120"},
      //...
    ],
    "references": [
      {"code": "R-002", "message": "Знайдено посилання [25], яке відсутнє у списку літератури.", "location": "Сторінка 15"}
    ]
  }
}
```

Алгоритм розрахунку показника оцінки оформлення використовується для надання єдиної, кількісної оцінки якості оформлення. Кожній категорії або конкретному типу виявленої помилки присвоюється коефіцієнт ваги (W_i), що відображає її критичність та вплив на загальну якість документа (Табл.1).

Таблиця 1
Схема визначення відхилень (вага)

Категорія помилки	Тип критичності	Приклад відхилення	Вага (W_i)
Критична	Структурні / Життєво важливі	Некоректні поля, відсутність списку літератури, порушення структури заголовків.	≥ 3.0
Основна (Major)	Вимоги до змісту елементів	Неправильний формат цитування, некоректний стиль заголовка розділу, відсутність обов'язкових елементів бібліографії.	1.5–2.9
Незначна (Minor)	Стилістичні/Естетичні	Неправильний інтервал між абзацами, використання некоректного шрифту для виносок, зайві пробіли.	1.0–1.4

Визначення ваги помилок розраховується як сума (S) добутків кількості виявлених помилок (n_i) кожного типу на відповідний коефіцієнт ваги (W_i).

$$S = \sum_{i=1}^k (n_i \cdot W_i), \quad (1)$$

де k – загальна кількість унікальних типів помилок; n_i – кількість виявлень помилки i -го типу; W_i – коефіцієнт ваги для помилки i -го типу.

Для отримання уніфікованої оцінки у відсотках потрібно нормалізувати отриману суму відносно максимально можливого або встановленого базового штрафу. Максимальний базовий штраф ($S_{\max}$) – це ключовий параметр, який встановлюється як еталон. Інтегральний показник (E) – це відсоток відповідності вимогам. Він розраховується шляхом віднімання частки нормалізованого штрафу від 100 % відповідності ідеальному оформленню.

$$E = 100 \cdot \left(1 - \frac{S}{S_{\max}} \right) \quad (2)$$

Наведемо приклад розрахунку у таблиці 2. При умові, що $S_{max} = 30.0$

Таблиця 2
Приклад розрахунку сумарного штрафу

Помилка (i)	Опис	Вага (Wi)	Кількість	Штраф
1 (P-001)	Порушення лівого поля	3.0 (Критична)	1	3.0
2 (R-002)	Посилання відсутнє у списку	2.5 (Основна)	2	5.0
3 (T-005)	Неправильний шрифт абзаців	1.0 (Незначна)	5	5.0
4 (P-002)	Неправильний міжрядковий інтервал	1.2 (Незначна)	10	12.0
Разом	–	–	–	25.0

Розрахунок інтегрального показника $E = 100 \cdot (1 - 25.0/30.0) = 100 \cdot 16.67\%$. У даному прикладі E становить 16.67 %, це дозволяє об'єктивно оцінити роботу, де 100 % означає повну відповідність вимогам, а нижчий відсоток – необхідність виправлень.

Наведений вище алгоритм було реалізовано у межах експериментального тестування прототипу під час перевірки текстових документів (звіти, курсові, наукові статті). Система автоматично обчислювала показник E для кожного файлу, що дало змогу порівняти якість оформлення різних типів документів. Отримані результати показали, що документи, створені відповідно до методичних рекомендацій закладу освіти, мали інтегральний показник у межах 95–98 %, а документи з порушеннями стилів, відступів і нумерації – 75–90 %, що підтвердило ефективність алгоритму та коректність такого підходу до оцінювання форматування. Таким чином, розроблена модель не лише формалізує процес перевірки, а й забезпечує можливість кількісного порівняння якості оформлення документів на основі об'єктивних показників.

Генерація звіту про перевірку оформлення документів формату docx є завершальним етапом запропонованої нами системи. Саме інтегральна оцінка, яка обчислюється на основі вагових коефіцієнтів помилок, слугує основою для формування змісту звіту: вона визначає загальний рівень відповідності документа нормативним вимогам, а також дозволяє класифікувати порушення за критичністю. Модуль ReportGenerator використовує результати цієї оцінки для автоматичного створення структурованого звіту, що містить деталізовану інформацію про кожне виявлене відхилення, його місцезнаходження та рекомендації щодо усунення. Такий підхід забезпечує прозорість, об'єктивність і зручність оцінювання академічної документації.

Висновки та перспективи подальших досліджень

Для досягнення поставленої мети авторами було спроектовано та реалізовано програмний прототип – інтелектуальну систему перевірки форматування текстових документів формату docx, що поєднує структурний аналіз, алгоритмічну валідацію параметрів оформлення та генерацію інтегральної оцінки якості. Розроблена система охоплює повний цикл обробки документа, від екстракції стилів до формування підсумкового звіту, має тримодульну архітектуру (StyleAnalyzer, FormatValidator, ReportGenerator), яка реалізує функціональність аналізу, перевірки та звітування на основі параметрів форматування згідно ДСТУ 3008:2015 та внутрішніх стандартів ЗВО. В статті наведені відповідні алгоритми та результати експериментального тестування на академічних документах ЗВО, яке підтвердило ефективність запропонованого рішення, його стабільність та точність. Отримані інтегральні показники (від 75 % до 98 %) свідчать про здатність системи об'єктивно оцінювати якість оформлення та зменшувати кількість помилок у академічній документації.

Перспективи подальших досліджень передбачають розширення функціоналу створеного прототипу в частині семантичного аналізу змісту документа, інтеграції з репозитаріями освітніх та наукових робіт, а також створення комплексної платформи оцінки якості, структури та змісту документів у межах EdTech-рішень.

Запропонована система є науково обґрунтованим і технічно реалізованим інструментом, що має потенціал до широкого застосування в академічному середовищі, сприяє підвищенню якості освітнього процесу та автоматизації документообігу.

Література

1. Martsinkevich V., Veselova E. Web Tool for Automated Document Formatting Verification. *Proceedings of FRUCT Conference*. 2020. [Online] URL: <https://fruct.org/publications/volume-33/acm33/files/Mar.pdf>
2. Nasyrov N. Automated Formatting Verification Technique of Paperwork Elements. *Procedia Computer Science*. Elsevier, 2020. Vol. 169. pp. 719–724. DOI: 10.1016/j.procs.2020.02.189. URL: <https://www.sciencedirect.com/science/article/pii/S1877050920324170>
3. Zhigalova M., Morozova A., Bessarabov K. System of Automated Checking of Textual Document Design. *HSE University Publications*. 2015. URL: <https://publications.hse.ru/pubs/share/folder/0rdkqezioy/185788493.pdf>
4. Peña A. Continuous document layout analysis: Human-in-the-loop..., Information Sciences. Elsevier. 2024. [Online]. URL: <https://www.sciencedirect.com/science/article/pii/S1566253524001763>
5. Сидоренко В. В., Кулішов В. С., Гірка С. В. Порядок проведення інструментальної перевірки на академічний плагіат, наявність текстових запозичень у рукописах кваліфікаційних робіт здобувачів вищої освіти освітнього ступеня «магістр» у 2025 році. Біла Церква : БІНПО, 2025. 18 с. URL: <https://lib.iitta.gov.ua/id/eprint/744928>

6. Музичук Д. Р., Войтко В. В., Черноволик Г. О. Розробка засобів системи автоматизованого формування і редагування текстових документів. *Інформаційні технології і автоматизація – 2024* : матеріали XVII міжнародної науково-практичної конференції (Одеса, 31 жовтня – 1 листопада 2024 р.). Одеса : Видавництво ОНТУ, 2024. С. 479–482.
7. Вербовий Д. С., Саяпіна І. О. Метод для автоматичної перевірки документів на відповідність нормативним вимогам. *Системні технології*. 2025. № 3 (158). С. 181–187.
8. Сидорко Д. С., Глухов В. С. Програмна реалізація методів аналізу та нормоконтролю текстових технічних звітів. *Computer systems and networks*. 2024. № 6 (2). С. 208–221. DOI: 10.23939/csn2024.02.208.
9. Заболотня Т., Сечин І. Комбінований метод виявлення нечітких дублікатів текстових документів. *Progressive Science and Achievements* : proceedings of the 4th International Scientific and Practical Conference (May 16–18, 2024). Doha, Qatar, 2024. № 200. С. 229–232.
10. Данильченко В. М. Багатокритеріальне розпізнавання відповідності текстів темі на основі алгоритму TF-IDF. *Зв'язок*. 2025. № 3. С. 36–41. DOI: 10.31673/2412-9070.2025.027728.

References

1. Martsinkevich V., Veselova E. (2020). Web Tool for Automated Document Formatting Verification. Proceedings of FRUCT Conference. [Online]. URL: <https://fruct.org/publications/volume-33/acm33/files/Mar.pdf>
2. Nasyrov N. (2020). Automated Formatting Verification Technique of Paperwork Elements. *Procedia Computer Science. Elsevier*. Vol. 169. Pp. 719–724. DOI: 10.1016/j.procs.2020.02.189. URL: <https://www.sciencedirect.com/science/article/pii/S1877050920324170>.
3. Zhigalova M., Morozova A., Bessarabov K. (2015). System of Automated Checking of Textual Document Design. HSE University Publications. URL: <https://publications.hse.ru/pubs/share/folder/Ordkqezioy/185788493.pdf>
4. Peña A. (2024). Continuous document layout analysis: Human-in-the-loop..., *Information Sciences. Elsevier*. [Online] URL: <https://www.sciencedirect.com/science/article/pii/S1566253524001763> [in Ukrainian].
5. Sydorenko V. V., Kulishov V. S., Hirka S. V. (2025). Poriadok provedennia instrumentalnoi perevirky na akademichnyi plahiat, naiavnist tekstovykh zapozychen u rukopysakh kvalifikatsiinykh robot zdobuvachiv vyshchoi osvity osvithnoho stupenia "mahistr" u 2025 rotsi. *Bila Tserkva : BINPO*. 18 p. URL: <https://lib.iitta.gov.ua/id/eprint/744928> [in Ukrainian].
6. Muzychuk D. R., Voitko V. V., Chernovolik H. O. (2024). Rozrobka zasobiv systemy avtomatyzovanoho formuvannia i redahuvannia tekstovykh dokumentiv. *Information technology and automation – 2024* : materialy XVII mizhnarodnoi naukovo-praktychnoi konferentsii (Odesa, 31 zhovtnia – 1 lystopada 2024 r.). Odesa : Vydavnytstvo ONTU. Pp. 479–482 [in Ukrainian].
7. Verbovyi D. S., Saiapina I. O. (2025). Metod dlia avtomatychnoi perevirky dokumentiv na vidpovidnist normatyvnym vymoham. *System technologies*. № 3 (158). Pp. 181–187. [in Ukrainian]
8. Sydorko D. S., Hlukhov V. S. (2024). Prohramna realizatsiia metodiv analizu ta normokontroliu tekstovykh tekhnichnykh zvitiv. *Computer systems and networks*. № 6 (2). Pp. 208–221. DOI: 10.23939/csn2024.02.208 [in Ukrainian].
9. Zabolotnia T., Siechyn I. (2024). Kombinovanyi metod vyjavlennia nechitkykh dublikativ tekstovykh dokumentiv. *Progressive Science and Achievements* : Proceedings of the 4th International Scientific and Practical Conference (May 16–18, 2024). Doha, Qatar. № 200. Pp. 229–232 [in Ukrainian].
10. Danylchenko V. M. (2025). Bahatokryterialne rozpiznavannia vidpovidnosti tekstiv temi na osnovi alhorytmu TF-IDF. *Communication*. № 3. Pp. 36–41. DOI: 10.31673/2412-9070.2025.027728 [in Ukrainian].

Дата першого надходження статті до видання: 27.10.2025

Дата прийняття статті до друку після рецензування: 25.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.9:37.018.43:305

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-16>

АНАЛІТИКА ОСВІТНІХ ДАНИХ ЯК ІНСТРУМЕНТ ВИЗНАЧЕННЯ ГЕНДЕРНИХ ДИСБАЛАНСІВ У ЦИФРОВІЙ ОСВІТІ

Стисло Т. Р., к.ю.н., доцент кафедри ІТ «Університет Короля Данила», Івано-Франківськ, Україна.
<https://orcid.org/0000-0002-2377-7985>, taras.styslo@ukd.edu.ua

Анотація. У статті досліджується, як аналітика освітніх даних може допомогти знайти відповідь на це питання, зосереджуючись на гендерних дисбалансах у цифровій освіті. Наприклад, чому дівчата часто відмовляються від ІТ-спеціальностей, навіть маючи такі ж здібності, як хлопці? Чи впливає оформлення онлайн-курсів на їхню участь? Ці нюанси важко помітити без системного аналізу – саме тут і знадобиться Learning Analytics. Порівнюючи дані UNESCO, OECD та українські звіти, стає очевидно: проблема глибша, ніж просто «менше жінок у STEM». У багатьох країнах дівчата краще вчаться на цифрових платформах, але рідше обирають технічні спеціальності через стереотипи чи відсутність рольових моделей. Щоб змінити ситуацію, потрібно не просто збирати статистику, а розуміти, як формуються ці нерівності. У статті пропонується практичний підхід: збирання даних не лише про оцінки, а й про те, як студенти взаємодіють з платформами (наприклад, чи часто вони заходять на технічні модулі, як довго вивчають теми, які запитання задають). У моделі оцінення складу студентів додано етап «перевірки на гендерну сліпоту» – перед запуском аналізу дані очищаються від статевих маркерів, а потім результати порівнюються з реальними соціологічними дослідженнями. Також описано, як створити дашборд, де адміністратори побачать не просто графіки, а конкретні рекомендації. Не менш важливо – питання конфіденційності. Автор наводить приклад з нідерландського університету, де анонімізовані дані все ж дозволили ідентифікувати студентів через їхні унікальні шаблони навчання. Тому в пропонуваній моделі обробка даних відбувається локально, без передачі в хмарні сервіси, а доступ до інформації мають лише верифіковані експерти. Стаття показує: шлях до гендерної справедливості у цифровій освіті – це не автоматичні алгоритми, а їхнє розумне поєднання з людським досвідом.

Ключові слова: аналітика освітніх даних; цифрова освіта; гендерна рівність; інформаційні технології; машинне навчання; цифрова інклюзія; етичне використання даних.

LEARNING ANALYTICS AS A TOOL FOR IDENTIFYING GENDER IMBALANCES IN DIGITAL EDUCATION

Styslo T. R., Ph.D., Ass. Prof. of the IT Department of King Danylo University, Ivano-Frankivsk, Ukraine.
<https://orcid.org/0000-0002-2377-7985>, taras.styslo@ukd.edu.ua

Abstract. The article explores how educational data analytics can help answer this question by focusing on gender imbalances in digital education. For example, why do girls often drop out of IT majors, even when they have the same abilities as boys? Does the design of online courses affect their participation? These nuances are difficult to notice without a systematic analysis – this is where Learning Analytics comes in handy. Comparing data from UNESCO, OECD and Ukrainian reports, it becomes clear: the problem is deeper than just “fewer women in STEM”. In many countries, girls learn better on digital platforms, but are less likely to choose technical majors due to stereotypes or a lack of role models. To change the situation, it is necessary not only to collect statistics, but also to understand how these inequalities are formed. The article suggests a practical approach: collecting data not only on grades, but also on how students interact with the platforms (for example, how often they access technical modules, how long they study the topics, what questions they ask). The student composition assessment model includes a “gender blindness check” stage – before running the analysis, the data is cleaned of gender markers, and then the results are compared with real sociological studies. It also describes how to create a dashboard where administrators will see not just graphs, but specific recommendations. No less important is the issue of confidentiality. The author gives an example from a Dutch university, where anonymized data still allowed students to be identified through their unique learning patterns. Therefore, in the proposed model, data processing takes place locally, without being transferred to cloud services, and only verified experts have access to the information. The article shows: the path to gender justice in digital education is not automatic algorithms, but their intelligent combination with human experience.

Key words: learning analytics; digital education; gender equality; information technologies; machine learning; digital inclusion; ethical data use.

Вступ

У сучасному світі цифрові технології кардинально змінюють освіту. Ми все частіше використовуємо розумні системи для аналізу навчальних даних – від простого збору інформації до її глибокої обробки та інтерпретації [1; 2]. Це дозволяє не лише індивідуалізувати навчання, а й краще розуміти соціальні процеси в освітньому середовищі, зокрема гендерні аспекти [3].

Проте автоматизація освітніх процесів створює нові виклики. Алгоритми, які керують доступом до знань і оцінюють успішність студентів, можуть несвідомо поширювати старі стереотипи [4]. Адже вони часто базуються на історичних даних, де вже присутні суспільні упередження [5; 6].

Особливо гостро ця проблема стоїть в Україні, де поки немає чітких методик для аналізу освітніх даних з точки зору гендерної рівності [7]. Через це складно розробляти ефективні програми цифрової інклюзії, які враховують потреби всіх груп користувачів.

Аналіз останніх досліджень

Гендерна рівність у цифровій освіті – це тема, яка активно обговорюється в багатьох міжнародних дослідженнях. Наприклад, у звіті UNESCO “Gender Equality in the Digital Learning Era” (2023) зазначається, що жінки складають менше третини учасників STEM-дисциплін у світі [8]. Ще один тривожний момент полягає в тому, що алгоритми навчальних платформ часто несвідомо підсилюють традиційні стереотипи, що створює додаткові бар'єри для гендерної рівності.

Організація OECD (2022) наголошує, що ефективне використання освітньої аналітики можливе лише за умов наявності політик відкритості даних, прозорості алгоритмів і формування етичної культури оброблення освітньої інформації [9].

Дослідження Lund і Wang (2023) показує, що використання алгоритмів машинного навчання в університетських системах оцінювання кардинально змінює підхід до навчання [10]. Однак такі інновації несуть і ризики – зокрема, можливість нерівного ставлення до студентів під час автоматичного відбору або оцінювання.

Цю проблему підкреслюють Floridi (2018) та Crawford (2021), які наголошують, що упередженість даних – це не просто технічний недолік, а соціальний феномен [11; 12]. Вони пропонують розглядати цю тему через призму етики алгоритмів і принципів управління даними, що робить питання ще глибшим і актуальнішим.

У межах українського освітнього простору питання гендерної рівності висвітлено в Концепції розвитку штучного інтелекту в Україні (Мінцифра, 2023) [13], проте в ній відсутні конкретні показники моніторингу рівності у цифровій освіті, що свідчить про потребу у формуванні науково верифікованих індикаторів.

Незважаючи на велику кількість теоретичних робіт, поки що немає комплексних досліджень, які б поєднували аналітику освітніх даних, гендерну рівність і цифрові технології в освіті. Саме це створює прогалину, яку важливо заповнити.

Мета даного дослідження – науково обґрунтувати, як аналітика освітніх даних може допомогти виявляти та подолати гендерні дисбаланси в цифровому освітньому середовищі. Крім того, ми пропонуємо розробити принципи створення етичних аналітичних систем, які не будуть мати упереджень.

Для досягнення цієї мети було проаналізовано сучасні технології Learning Analytics, вивчено причини появи гендерних нерівностей у цифровій освіті, а також розглянуто правові та етичні проблеми алгоритмічного оцінювання. На основі цього була запропонована концептуальна модель для моніторингу гендерної рівності на освітніх платформах.

Виклад основного матеріалу

Аналітика освітніх даних (Learning Analytics, LA) є міждисциплінарним науковим напрямом, що інтегрує методи інформатики, статистичного аналізу, педагогічної діагностики, когнітивної психології та соціології освіти [1; 2; 10]. Її теоретичною основою є концепція data-driven education, згідно з якою освітні рішення ухвалюються на підставі кількісного аналізу поведінкових, академічних і соціальних даних здобувачів освіти (рис. 1) [14].

У сучасних умовах цифрової трансформації LA перетворюється на фундаментальну складову освітньої аналітики, забезпечуючи не лише моніторинг навчальних процесів, а й науково обґрунтоване управління якістю освіти та прогнозування освітніх результатів [2; 15].

З технічної точки зору, аналітика освітніх даних спирається на комбінацію таких методів, як машинне навчання, предиктивне моделювання, обчислювальна лінгвістика та когнітивна статистика. Це означає, що для аналізу даних використовуються різні інструменти, які допомагають знаходити закономірності, прогнозувати результати та краще розуміти процеси у навчанні.

У системах на кшталт Civitas Learning, Panorama Education чи Canvas Insights застосовуються алгоритми кластеризації, регресійного аналізу, латентного моделювання тем і нейронні мережі з пояснюваною архітектурою (Explainable AI). Такі підходи дозволяють здійснювати глибоке вивчення освітньої поведінки, визначати закономірності успішності та прогнозувати ризики відставання. Однак у межах сучасних досліджень дедалі більше уваги приділяється не лише ефективності цих технологій, а й питанням соціальної справедливості, етичності та гендерного балансу [5; 8; 9].

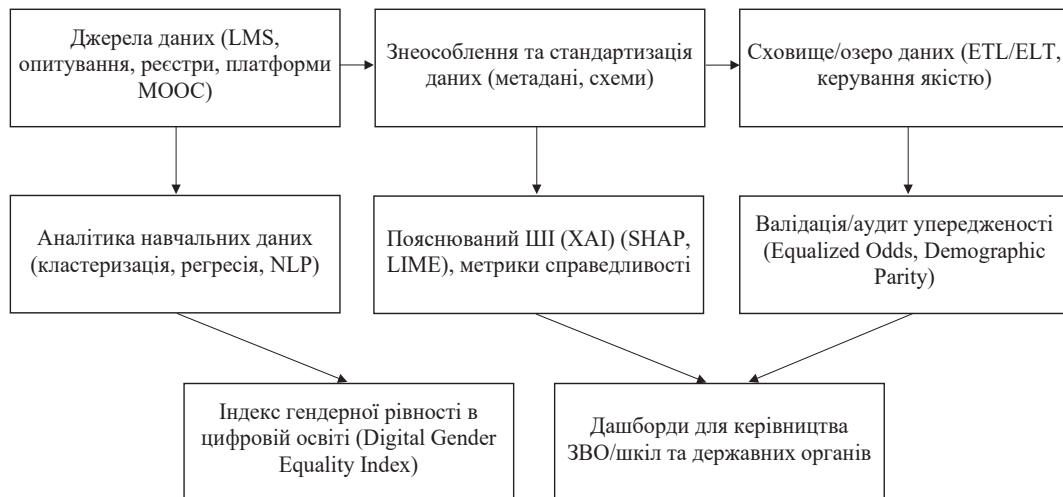


Рис. 1. Концептуальна схема Gender-Aware Learning Analytics (GALAF)

Таблиця 1
Гендерний розподіл випусників STEM-дисциплін за регіонами

Регіон / Країна	Частка жінок серед STEM-випусників, %	Частка жінок у STEM-робочій силі, %	Джерело даних
Глобальний показник	35	28–33	UNESCO, 2023–2024
Європейський Союз	24–28	52	OECD, 2024
США	36	28	NSF, 2023
Індія	45 (випусники)	27	World Bank, 2024
Південна Африка	47 (випусники)	н/д	UNESCO, 2024
Японія	16	н/д	OECD, 2024
Франція	28	н/д	OECD, 2024
Саудівська Аравія	9,2 → 16,1*	н/д	Coursera, 2021–2022

* Зростання частки жінок в онлайн-навчанні ІКТ (2019–2021)

У контексті гендерної рівності аналітика освітніх даних виконує функцію об'єктивного інструменту для виявлення структурних асиметрій у цифровому освітньому просторі. Її потенціал полягає у можливості аналізу великих вибірок освітніх даних із метою визначення гендерних відмінностей у доступі до технологій, академічній результативності, цифровій залученості та соціальній взаємодії. За даними звіту UNESCO “Gender Equality in the Digital Learning Era” (2023), жінки становлять менше третини учасників технічних дисциплін і лише близько 28 % користувачів онлайн-курсів із програмування чи аналізу даних [8]. Подібні тенденції підтверджує OECD Education at a Glance (2024), що фіксує суттєву нерівномірність у STEM-освіті на глобальному рівні. Водночас у гуманітарних напрямках спостерігається протилежна ситуація – понад 60 % слухачів є жінками, що свідчить про стійку професійну сегрегацію [9].

Для кращого розуміння масштабів гендерної нерівності у сфері STEM-освіти варто звернути увагу на регіональні особливості участі жінок. Як показують дані UNESCO та OECD, рівень представленості жінок серед випусників технічних спеціальностей суттєво відрізняється залежно від регіону (таблиця 1). Ці відмінності дуже показові і допомагають побачити загальну картину проблеми. В одних регіонах ситуація трохи краща, а в інших – надзвичайно критична.

Однак гендерна нерівність – це не тільки питання технічних дисциплін, вона набагато ширша і має дві сторони. Це явище виходить далеко за рамки STEM-освіти і торкається різних аспектів суспільства. Аналіз освітніх напрямів демонструє стійку професійну сегрегацію: жінки недопредставлені у STEM, тоді як чоловіки – у гуманітарних науках, освіті та сфері догляду. Ця асиметрія обмежує можливості розвитку для обох гендерних груп і потребує комплексного підходу до забезпечення рівності (рис. 2).

Важливо зазначити, що гендерний розрив у цифрових навичках має складну вікову динаміку. Дані Eurostat демонструють, що у молодших вікових групах жінки навіть випереджають чоловіків за рівнем цифрових компетентностей, проте ситуація кардинально змінюється після 45 років (рис. 3). Це свідчить про важливість цілеспрямованих освітніх інтервенцій для різних демографічних груп.

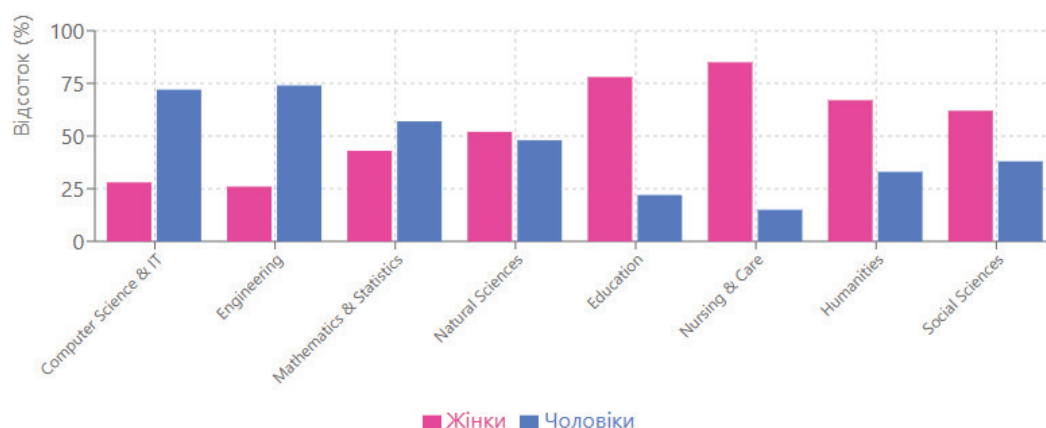


Рис. 2. Гендерна сегрегація за освітніми напрямками

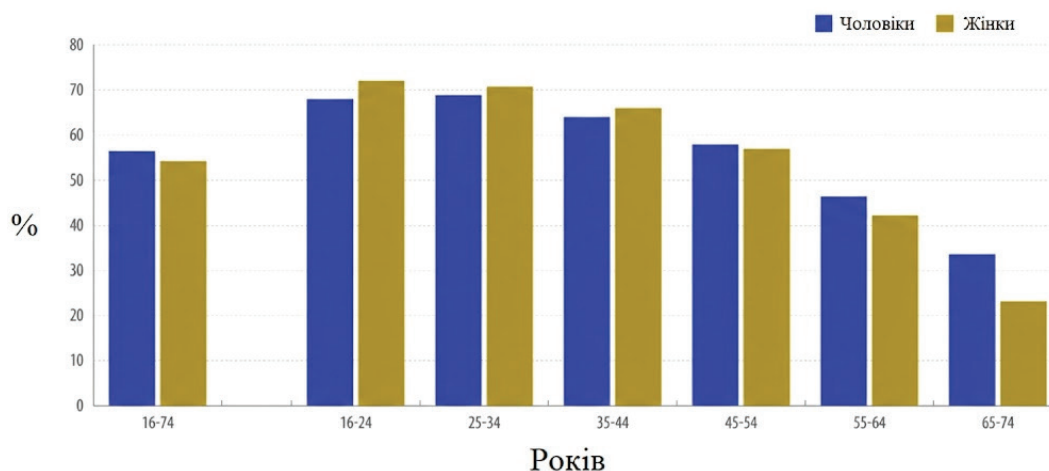


Рис. 3. Гендерний розподіл цифрових компетентностей за віковими групами

Гендерна сегрегація не обмежується лише традиційними формами освіти – вона активно проникає і в цифровий освітній простір. Аналіз поведінкових патернів на платформах масових відкритих онлайн-курсів (МООК) виявляє стійкі відмінності у виборі освітніх напрямів між гендерними групами. Дані платформи Coursera за 2024 рік демонструють, що жінки становлять лише 24–32 % слухачів технічних дисциплін (Data Science, Programming, Cybersecurity), тоді як у гуманітарних напрямках, психології та вивченні мов їх частка сягає 65–71 % (рис. 4). Ця поляризація – це не просто статистика, а віддзеркалення глибших соціокультурних тенденцій, які формуються роками. Вона чітко показує, що гендерні стереотипи залишаються потужним фактором, який впливає на те, які освітні шляхи обирають хлопці та дівчата. І тут замах самостійно подолати ці бар'єри недостатньо – потрібні цілеспрямовані дії.

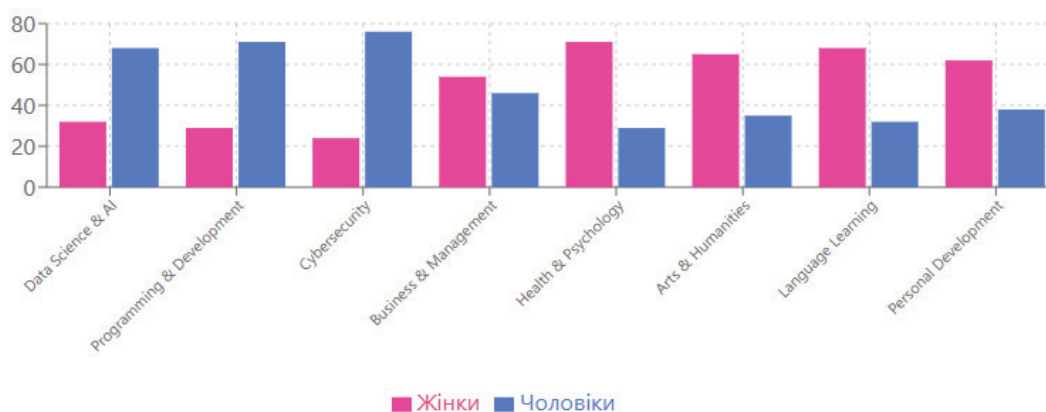


Рис. 4. Гендерний розподіл у онлайн-освіті за категоріями курсів

Аналіз цих даних демонструє, що причини гендерних дисбалансів у цифровій освіті мають складний, багатошаровий характер. Крім культурних стереотипів і соціальних очікувань, важливу роль відіграють технологічні фактори. Зокрема, алгоритмічні системи, які пропонують користувачам персоналізований контент, іноді посилюють існуючі упередження. Наприклад, жінкам частіше рекомендують курси з гуманітарних дисциплін, тоді як чоловікам – з програмування чи аналітики.

Це підкреслює необхідність перевірки алгоритмів на наявність дискримінаційних патернів і впровадження процедур Fairness Metrics, таких як Demographic Parity, Equalized Odds та Counterfactual Fairness [16].

Поглиблений аналіз освітніх даних допомагає не лише виявляти проблеми, але й розуміти їх причини завдяки методам Explainable AI, таким як SHAP і LIME. Ці інструменти дозволяють оцінювати, як окремі параметри – наприклад, активність на форумах, час взаємодії з контентом або використання аналітичних інструментів – впливають на результати навчання [15; 16].

На основі таких підходів можна створити Індекс гендерної рівності у цифровій освіті (Digital Gender Equality Index), який враховує академічні, поведінкові, соціально-економічні та когнітивні показники. Він дозволяє кількісно оцінити рівень інклюзивності освітніх середовищ і виявити приховані бар'єри для доступу до цифрових можливостей [14].

Дослідження на базі великих масивів даних LMS свідчать, що гендерний дисбаланс проявляється не лише в кількості учасників, а й у специфічних паттернах взаємодії з цифровим контентом. Порівняння рівнів залученості (Engagement Rate) студентів у залежності від типу освітніх ресурсів наведено в таблиці 2.

Таблиця 2
Паттерни залученості в цифрове навчання за гендерною ознакою

Тип активності	Жінки (сер. показник)	Чоловіки (сер. показник)	Різниця (%)
Участь у дискусійних форумах	12,4 хв/тиждень	8,2 хв/тиждень	+51 % (Ж)
Завершення відеолекцій (STEM)	78 %	85 %	+9 % (Ч)
Виконання практичних ІТ-зادань	72 бали	74 балла	+2,7 % (Ч)
Використання адаптивних підказок	45 разів/курс	30 разів/курс	+50 % (Ж)

Проблема етичності використання освітньої аналітики залишається однією з ключових у сучасних дослідженнях. Регламент Європейського Союзу EU Artificial Intelligence Act (2024) прямо вимагає забезпечення прозорості, пояснюваності та підзвітності алгоритмів у сферах із високим соціальним впливом, зокрема в освіті [17]. Ці принципи відповідають міжнародним стандартам ISO/IEC 42001:2023 "Artificial Intelligence Management System" та ISO/IEC 23894:2023 "AI Risk Management", які регламентують створення систем управління ризиками, аудиту та контролю даних [18]. В українському нормативному полі поки відсутні прямі механізми сертифікації або незалежного моніторингу освітніх аналітичних систем, що створює правову прогалину та загрозу несанкціонованого використання студентських даних.

В умовах сучасної цифровізації освіти постає важлива проблема – як зробити роботу з даними етичною та без дискримінації? Ось тут-то дослідники і пропонують цікавий підхід – Gender-Aware Learning Analytics Framework (GALAF). Це своєрідна «карта» для створення таких систем аналітики освітніх даних, де людина залишається в центрі уваги.

Давайте розберемося, чому це важливо. Спочатку треба навчитися правильно збирати дані, враховуючи різні соціальні та демографічні фактори. Потім – стежити, щоб алгоритми не мали упереджень. І, звичайно, потрібні спеціальні інструменти для перевірки результатів на об'єктивність. Приємно, що такий підхід добре вписується в європейські стандарти цифрової освіти, яких Україна намагається дотримуватися.

Застосування подібних етичних підходів до аналізу освітніх даних має кілька важливих переваг. По-перше, користувачі починають довіряти цифровим платформам. По-друге, стає зрозумілішим, як саме приймаються рішення. І найголовніше – зменшується ризик дискримінації.

З практики досліджень видно, що аналітика освітніх даних – це не просто набір графіків і цифр. Це потужний інструмент, який допомагає не лише оцінювати успішність студентів, але й розуміти ширші суспільні процеси. Особливо це важливо для вирішення проблем гендерної нерівності в ІТ-сфері.

На жаль, все не так просто. Щоб такі системи працювали ефективно, потрібно враховувати кілька важливих моментів. По-перше, законодавство має чітко регулювати обробку даних. По-друге, якість самих даних має бути на висоті – адже будь-які перекося в них можуть призвести до упереджених результатів. І, звичайно, важливо, щоб ці інструменти гармонійно вписувалися в існуючу систему освіти – як на рівні окремих закладів, так і в масштабах всієї країни.

Встановлено, що на глобальному рівні (згідно з дослідженнями UNESCO, OECD, Coursera Global Skills Report 2024) зберігаються значні гендерні відмінності у виборі освітніх напрямів: частка жінок у STEM-дисциплінах становить у середньому 30 %, тоді як у гуманітарних – понад 60 %. Такі диспропорції відображають не лише соціальні чинники, але й технологічні, зокрема – алгоритмічну упередженість рекомендаційних систем, що пропонують користувачам навчальні курси на основі історичних даних, які вже містять сліди нерівності. У цьому контексті аналітика освітніх даних постає не просто як технічний засіб, а як науковий інструмент системного спостереження за соціальними процесами у цифровому середовищі.

Висновки

Рівень етичної зрілості державної та інституційної політики у сфері цифрової освіти прямо впливає на об'єктивність і справедливість освітньої аналітики. Етичні принципи – прозорість, пояснюваність, недискримінаційність і підзвітність – мають бути інтегровані в архітектуру аналітичних систем як базові параметри. Без дотримання цих принципів аналітика освітніх даних може перетворитися з інструменту розвитку на механізм відтворення нерівності, що суперечить як міжнародним нормам (EU AI Act, ISO/IEC 42001:2023, ISO/IEC 23894:2023), так і гуманістичній місії освіти.

Дослідження має важливе значення як з наукової, так і з практичної точки зору. Воно показує, що аналітика освітніх даних – це не просто технічний інструмент для оцінки успішності навчання. Це ще й чутливий соціальний «індикатор», який допомагає виявляти системні проблеми, що перешкоджають рівному доступу до освітніх можливостей у цифрову епоху.

Отримані результати можуть стати фундаментом для нової освітньої політики держави. Такої, що враховує не лише технологічний прогрес, але й принципи гендерної справедливості та соціальної включеності. Аналітика освітніх даних посідає особливе місце – вона одночасно є і науковим інструментом, і механізмом суспільного регулювання.

Саме тому вона є невід'ємною частиною цифрової трансформації освіти. Вона допомагає будувати нові моделі навчання, де людина залишається в центрі уваги, а всі рішення базуються на достовірних даних і ґрунтуються на довірі.

Подальші дослідження в цій галузі мають кілька перспективних напрямків. По-перше, треба покращувати методи аналізу даних, враховуючи культурні особливості. По-друге, потрібно робити штучний інтелект більш прозорим і зрозумілим. І, звичайно, важливо продовжувати інтеграцію принципів рівності у цифрові освітні платформи.

Література

1. Holmes W., Bialik M., Fadel C. Штучний інтелект в освіті: обіцянки та наслідки для навчання та викладання. Бостон : Center for Curriculum Redesign, 2019. 212 с.
2. Dwivedi Y. K., Hughes L., Rana N. P., et al. Generative AI for Education: Opportunities and Challenges. *Computers in Human Behavior*. 2024. Vol. 150. P. 108512. DOI: 10.1016/j.chb.2024.108512.
3. NAFSA. Generative AI and Global Education. *NAFSA: Association of International Educators*. Вашингтон : NAFSA, 2024. 42 с.
4. European Commission. Ethics Guidelines for Trustworthy AI / European Commission. Брюссель : EC, 2019. 35 с.
5. Floridi L., Cowls J. AI4People – An Ethical Framework for a Good AI Society. *Minds and Machines*. 2018. Vol. 28. No. 4. P. 689–707.
6. Crawford K. Атлас штучного інтелекту: влада, політика та планетарна ціна технологій. Нью-Гейвен : Yale University Press, 2021. 327 с.
7. OECD. Recommendation of the Council on Artificial Intelligence / Organisation for Economic Co-operation and Development. Париж : OECD Publishing, 2019. 24 с.
8. UNESCO. Gender Equality in the Digital Learning Era. *UNESCO Institute for Information Technologies in Education*. Париж : UNESCO Publishing, 2023. 96 с.
9. OECD. Education at a Glance 2024: OECD Indicators / Organisation for Economic Co-operation and Development. Париж : OECD Publishing, 2024. DOI: 10.1787/eag-2024-en.
10. Lund B. D., Wang T. The Impact of Artificial Intelligence on Higher Education. Берлін : Springer, 2023. 243 с.
11. Floridi L. The Ethics of Information. Оксфорд : Oxford University Press, 2019. 272 с.
12. Crawford K., Paglen T. Excavating AI: The Politics of Images in Machine Learning Training Sets. *AI Now Institute*. 2020. URL: <https://ainowinstitute.org> (дата звернення 22.10.2025).
13. Міністерство цифрової трансформації України. Концепція розвитку штучного інтелекту в Україні [Електронний ресурс]. Київ : Міністра, 2023. URL: <https://thedigital.gov.ua> (дата звернення 22.10.2025).
14. EDUCAUSE. AI in Higher Education: 2024 Horizon Report / EDUCAUSE. США, Боулдер : EDUCAUSE, 2024. 118 с.
15. Siemens G., Baker R. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration Proc. 2nd Int. Conf. on Learning Analytics and Knowledge. Ванкувер : ACM, 2022. P. 252–260.
16. DreamBox Learning. Adaptive Learning System Overview [Електронний ресурс]. Сіетл : DreamBox Learning, 2023. URL: <https://www.dreambox.com> (дата звернення 22.10.2025).
17. Європейський Парламент. Регламент (ЄС) 2024/1689 – Регламент про штучний інтелект (EU AI Act) [Електронний ресурс]. Офіційний журнал ЄС, 2024. URL: <https://eur-lex.europa.eu> (дата звернення 22.10.2025).
18. Штучний інтелект. Система управління (ISO/IEC 42001:2023, IDT) : ДСТУ ISO/IEC 42001:2023 / Нац. орган стандартизації України. Київ : ДП «УкрНДНЦ», 2024. 230 с. (Національний стандарт України).

References

1. Holmes W., Bialik M., Fadel C. *Shtuchnyi intelekt v osviti: obitsianky ta naslidky dlia navchannia ta vykladannia*. Boston : Center for Curriculum Redesign, 2019. 212 s.
2. Dwivedi Y. K., Hughes L., Rana N. P., et al. Generative AI for Education: Opportunities and Challenges. *Computers in Human Behavior*. 2024. Vol. 150. P. 108512. DOI: 10.1016/j.chb.2024.108512.
3. NAFSA. *Generative AI and Global Education* / NAFSA: Association of International Educators. Vashynhton : NAFSA, 2024. 42 s.
4. European Commission. *Ethics Guidelines for Trustworthy AI* / European Commission. Briussel : EC, 2019. 35 s.
5. Floridi L., Cowls J. AI4People – An Ethical Framework for a Good AI Society. *Minds and Machines*. 2018. Vol. 28. No. 4. P. 689–707.
6. Crawford K. *Atlas shtuchnoho intelektu: vlada, polityka ta planetarna tsina tekhnolohii* / K. Crawford. Niu-Heiven : Yale University Press, 2021. 327 s.
7. OECD. *Recommendation of the Council on Artificial Intelligence* / Organisation for Economic Co-operation and Development. Paryzh : OECD Publishing, 2019. 24 s.
8. UNESCO. *Gender Equality in the Digital Learning Era* / UNESCO Institute for Information Technologies in Education. Paryzh : UNESCO Publishing, 2023. 96 s.
9. OECD. *Education at a Glance 2024: OECD Indicators* / Organisation for Economic Co-operation and Development. Paryzh : OECD Publishing, 2024. DOI: 10.1787/eag-2024-en.
10. Lund B. D., Wang T. *The Impact of Artificial Intelligence on Higher Education*. Berlin : Springer, 2023. 243 s.
11. Floridi L. *The Ethics of Information*. Oksford : Oxford University Press, 2019. 272 s.
12. Crawford K., Paglen T. *Excavating AI: The Politics of Images in Machine Learning Training Sets*. *AI Now Institute*. 2020. URL: <https://ainowinstitute.org> (data zvernennia 22.10.2025).
13. Ministerstvo tsyvrovoi transformatsii Ukrainy. *Kontseptsiia rozvytku shtuchnoho intelektu v Ukraini* [Elektronnyi resurs]. Kyiv : Mintsyfra, 2023. URL: <https://thedigital.gov.ua> (data zvernennia 22.10.2025).
14. EDUCAUSE. *AI in Higher Education: 2024 Horizon Report* / EDUCAUSE. Boulder (SShA) : EDUCAUSE, 2024. 118 s.
15. Siemens G., Baker R. *Learning Analytics and Educational Data Mining: Towards Communication and Collaboration*. Proc. 2nd Int. Conf. on Learning Analytics and Knowledge. Vankuver : ACM, 2022. P. 252–260.
16. DreamBox Learning. *Adaptive Learning System Overview* [Elektronnyi resurs]. Sietl : DreamBox Learning, 2023. URL: <https://www.dreambox.com> (data zvernennia 22.10.2025).
17. Ievropeiskyi Parlament. *Rehlament (leS) 2024/1689 – Rehlament pro shtuchnyi intelekt (EU AI Act)* [Elektronnyi resurs]. Ofitsiinyi zhurnal YeS, 2024. URL: <https://eur-lex.europa.eu> (data zvernennia 22.10.2025).
18. *Shtuchnyi intelekt. Systema upravlinnia (ISO/IEC 42001:2023, IDT)* : DSTU ISO/IEC 42001:2023 / Nats. orhan standartyzatsii Ukrainy. Kyiv : DP "UkrNDNTs", 2024. 230 s. (Natsionalnyi standart Ukrainy).

Дата першого надходження статті до видання: 22.10.2025
 Дата прийняття статті до друку після рецензування: 25.11.2025
 Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.8:519.246

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-17>

КЛАСИЧНІ ТА ГІБРИДНІ МЕТОДИ ПРОГНОЗУВАННЯ ДАНИХ ЛІЧИЛЬНИКІВ ЗІ СПОЖИВАННЯ РЕСУРСІВ

Фуріхата Д. В., аспірант кафедри інженерії програмного забезпечення факультету, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0002-6093-664X>, fyrihata.denus@gmail.com

Вакалюк Т. А., д. пед. н., проф., завідувач кафедри інженерії програмного забезпечення, Державний університет «Житомирська політехніка», Житомир, Україна. <https://orcid.org/0000-0001-6825-4697>, tetianavakaliuk@gmail.com

Анотація. У статті проведено комплексний аналіз сучасних методів прогнозування споживання ресурсів на основі даних з лічильників, що набуває критичного значення для ефективного управління енергетичними та комунальними системами, особливо в умовах пошкодженої інфраструктури України. Досліджено три основні класи методів: традиційні статистичні моделі (ARIMA, SARIMA), евристичну модель Prophet, методи глибокого навчання (LSTM, RNN), а також гібридні підходи, що поєднують переваги різних методологій.

Виявлено, що класичні моделі ARIMA/SARIMA ефективні для короткострокового прогнозування стаціонарних рядів з лінійною структурою та чіткою сезонністю, забезпечуючи високу інтерпретованість результатів та можливість роботи з обмеженими обсягами даних. Проте ці моделі демонструють суттєві обмеження при роботі з нелінійними залежностями та динамічними змінами в структурі даних, що знижує точність прогнозів у складних реальних сценаріях. Модель Prophet демонструє переваги в автоматизації процесу моделювання, врахуванні календарних ефектів і свят, однак залишає невикористаною інформацію в автокорельованих залишках, що обмежує її застосування в завданнях, де критична максимальна точність.

Моделі LSTM показують найвищу точність при прогнозуванні складних нелінійних патернів завдяки механізму довгострокової пам'яті та здатності автоматично виявляти приховані залежності у даних. Експериментальні дослідження підтверджують суттєве підвищення точності порівняно з традиційними методами, особливо за наявності великих обсягів даних. Особливу увагу приділено гібридним підходам (ARIMA-LSTM, Prophet-LSTM), що поєднують інтерпретованість статистичних методів з потужністю нейромереж, забезпечуючи найвищу точність прогнозування. Результати дослідження мають практичне значення для оптимізації управління енергетичними ресурсами в умовах пошкодженої інфраструктури та можуть бути використані для розробки адаптивних систем прогнозування.

Ключові слова: часові ряди, прогнозування споживання, ARIMA, SARIMA, Prophet, LSTM, рекурентні нейронні мережі, гібридні моделі, глибоке навчання, лічильники ресурсів.

CLASSICAL AND HYBRID METHODS FOR FORECASTING RESOURCE CONSUMPTION METER DATA

Denys Furikhata, postgraduate student of the Department of Software Engineering, State University "Zhytomyr Polytechnic", Zhytomyr, Ukraine. <https://orcid.org/0000-0002-6093-664X>, fyrihata.denus@gmail.com

Tetiana Vakaliuk, Dr. Ped. Sc., Prof., Head of the Department of Software Engineering, State University "Zhytomyr Polytechnic", Zhytomyr, Ukraine. <https://orcid.org/0000-0001-6825-4697>, tetianavakaliuk@gmail.com

Abstract. The article provides a comprehensive analysis of modern methods for forecasting resource consumption based on meter data, which is critical for the effective management of energy and utility systems, especially in the context of Ukraine's damaged infrastructure. Three main classes of methods are investigated: traditional statistical models (ARIMA, SARIMA), the Prophet heuristic model, deep learning methods (LSTM, RNN), as well as hybrid approaches that combine the advantages of different methodologies.

It was found that classical ARIMA/SARIMA models are effective for short-term forecasting of stationary series with a linear structure and clear seasonality, providing high interpretability of results and the ability to work with limited amounts of data. However, these models have significant limitations when working with nonlinear dependencies and dynamic changes in data structure, which reduces the accuracy of forecasts in complex real-world scenarios. The Prophet model demonstrates advantages in automating the modeling

process and accounting for calendar effects and holidays, but leaves unused information in autocorrelated residuals, which limits its application in tasks where maximum accuracy is critical.

LSTM models show the highest accuracy in predicting complex nonlinear patterns thanks to their long-term memory mechanism and ability to automatically detect hidden dependencies in data. Experimental studies confirm a significant increase in accuracy compared to traditional methods, especially when large amounts of data are available. Particular attention is paid to hybrid approaches (ARIMA-LSTM, Prophet-LSTM), which combine the interpretability of statistical methods with the power of neural networks, ensuring the highest prediction accuracy. The results of the study are of practical importance for optimizing energy resource management in conditions of damaged infrastructure and can be used to develop adaptive forecasting systems.

Key words: time series, consumption forecasting, ARIMA, SARIMA, Prophet, LSTM, recurrent neural networks, hybrid models, deep learning, resource meters.

Вступ

Прогнозування споживання ресурсів на основі даних з лічильників набуває критичного значення для забезпечення ефективного функціонування енергетичних, водопостачальних та комунальних систем. Точне передбачення майбутнього попиту дозволяє оптимізувати розподіл ресурсів, знижувати операційні витрати, попереджувати аварійні ситуації та підвищувати загальну надійність інфраструктурних мереж. Особливої актуальності ця проблема набула в контексті глобальних викликів, пов'язаних зі зміною клімату, зростанням енергоспоживання та необхідністю переходу до сталого розвитку. В Україні, особливо в умовах воєнного стану та пошкоджені енергетичної інфраструктури, питання точного прогнозування споживання стає питанням національної безпеки та економічної стабільності.

Традиційні статистичні методи прогнозування, такі як ARIMA та SARIMA, хоча і залишаються популярними завдяки своїй інтерпретованості, виявляють суттєві обмеження при роботі зі складними нелінійними залежностями та багатофакторними впливами. Водночас, стрімкий розвиток технологій глибокого навчання відкриває нові можливості для підвищення точності прогнозів через застосування рекурентних нейронних мереж, зокрема LSTM. Проте питання вибору оптимального методу для конкретних умов застосування залишається відкритим і потребує комплексного наукового аналізу. Крім того, зростання обсягів даних, що генеруються сучасними системами обліку (розумні лічильники, IoT-пристрої), вимагає розробки ефективних методів їх обробки та аналізу.

Аналіз останніх джерел та публікацій

Останніми роками класичні та гібридні методи прогнозування часових розглядаються як вітчизняними, так і закордонними вченими. Так, у статті [4] розглянуто три основні підходи до прогнозування часових рядів: ARIMA, SARIMA та Prophet. Модель ARIMA добре працює зі стаціонарними даними, враховує автокореляції та лінійні залежності, однак виявляє обмеження у випадках, коли дані мають виражену сезонність. Розширенням цього підходу є SARIMA, яка завдяки додатковим параметрам дозволяє ефективніше моделювати циклічні та сезонні компоненти, проте вимагає більшої складності у налаштуванні. На відміну від них, Prophet, розроблений Facebook, ґрунтується на адитивній моделі з трендом, сезонністю, впливом свят та випадковим шумом [1]. Він автоматично враховує кілька сезонних компонентів, дозволяє швидко отримати прогнози без глибокого налаштування параметрів та ефективно опрацьовує складні часові ряди, що містять як тренди, так і нерегулярні події. Автори вказали, що Prophet продемонстрував найкращі результати серед розглянутих моделей завдяки здатності автоматично враховувати нелінійні тренди, багатосезонність і святкові ефекти [4].

В схожому дослідженні [7], але у сфері електричних показів, класична модель ARIMA виявилася менш ефективною для короткострокового прогнозування електричних навантажень, оскільки вона чутлива до аномалій і не здатна якісно відображати різкі зміни в даних. Натомість модель LSTM продемонструвала значно вищу точність завдяки здатності враховувати нелінійні залежності та виявляти складні патерни, проте її результативність суттєво залежить від обсягу та якості вхідних даних. Найкращі результати показала гібридна модель ARIMA-LSTM, яка об'єднала сильні сторони обох підходів, оскільки ARIMA забезпечила коректне відображення загального тренду, тоді як LSTM поліпшила прогнозування пікових і нестабільних значень. Це дозволило зменшити похибки та підвищити стійкість прогнозу до шуму, що робить ARIMA-LSTM найбільш придатним методом для практичного застосування в енергосистемах.

Дослідження виключно по нейронних мережах [2] показало, що використання глибоких рекурентних моделей, зокрема LSTM, забезпечує істотно кращі результати в короткостроковому прогнозуванні навантажень у порівнянні з класичними нейронними мережами. LSTM завдяки механізму пам'яті ефективно відображає як короткострокові, так і довгострокові залежності, що робить її оптимальною для прогнозування в умовах сезонних і добових коливань попиту. Водночас RNN також показує добрі результати, хоча й поступається LSTM за точністю. Включення додаткових факторів (погодних умов, типу дня та часу доби) підвищує якість прогнозів. Загалом дослідження [2] підкреслює, що LSTM є найефективнішим інструментом для задач короткострокового прогнозування енергоспоживання,

тоді як класичні RNN можуть бути використані як базовий орієнтир, але не відповідають сучасним вимогам до точності.

Отже, прогнозування майбутніх значень на основі часових рядів є критично важливим для прийняття рішень у різних галузях. Цей процес вимагає глибокого розуміння природи даних, що характеризуються послідовним часовим порядком, наявністю трендів (довгострокові зміни), сезонності (повторювані патерни), циклічності (нерегулярні коливання) та випадкового шуму (залишки).

Тому, метою статті є огляд класичних статистичних моделей (ARIMA/SARIMA), евристичних процедур (Prophet) та сучасних моделей глибокого навчання (LSTM, RNN, CNN) для прогнозування даних з часовими рядами.

Дослідження

Існує широкий спектр підходів до прогнозування споживання в лічильниках. Аналіз джерел дав змогу узагальнити їх у вигляді моделі наявних методів та алгоритмів для прогнозування даних [2; 4; 7] (див. рис. 1). Розглянемо їх детально.

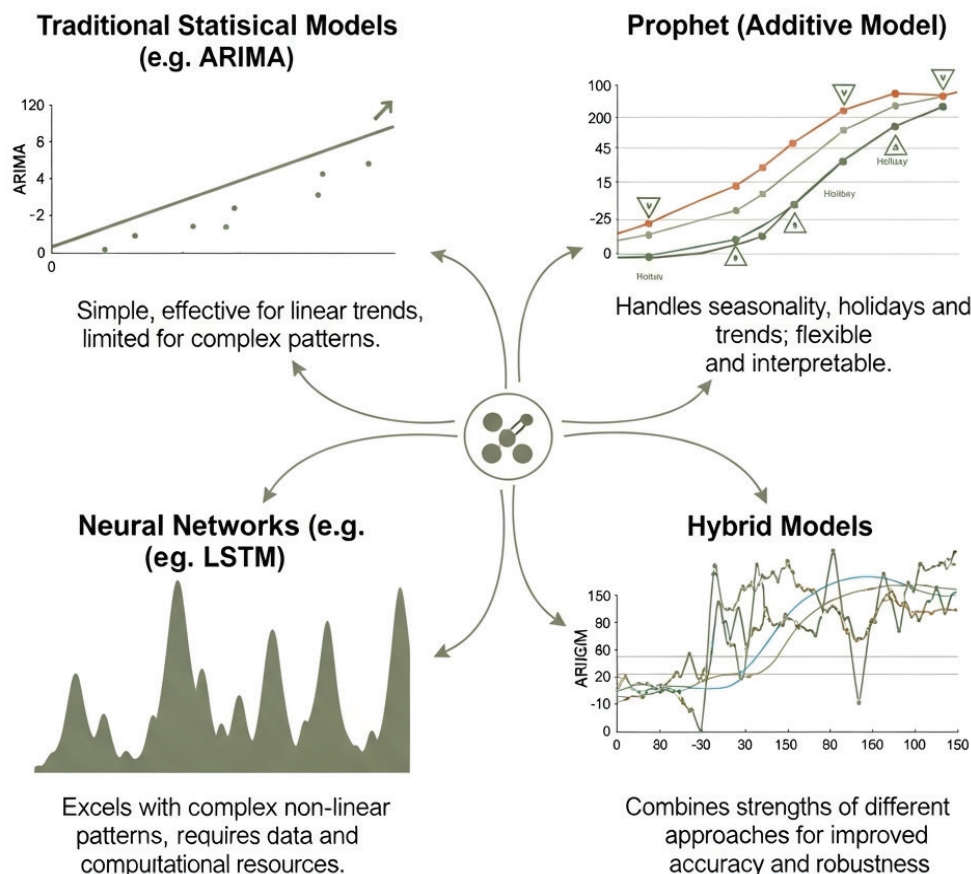


Рис. 1. Узагальнена класифікація методів прогнозування даних.

Традиційні статистичні моделі (ARIMA, SARIMA). Моделі ARIMA (Autoregressive Integrated Moving Average) та SARIMA (Seasonal ARIMA) є класичними лінійними статистичними підходами, що десятиліттями використовуються для аналізу та прогнозування часових рядів. Модель ARIMA ґрунтується на трьох основних компонентах: авторегресії (AR), що використовує минулі значення для прогнозування, інтеграції (I), що застосовує диференціювання для досягнення стаціонарності даних та ковзного середнього (MA), що враховує минулі помилки прогнозування [8]. SARIMA є розширенням ARIMA, що додає сезонні компоненти для моделювання періодичних патернів у даних [4]. До основних переваг цих моделей відносяться їхня інтерпретовність та ефективність.

Традиційні статистичні моделі ARIMA та SARIMA є класичними інструментами прогнозування часових рядів. ARIMA-модель (авторегресійна інтегрована модель ковзного середнього) застосовується для аналізу даних з інтегрованою різницею та ковзним середнім.

Модель характеризується трьома основними параметрами: оператором зсуву, параметрами моделі та випадковим шумом. Ці параметри можна представити у канонічній формі авторегресійно-ковзного середнього [8].

Для врахування сезонності застосовується модель SARIMA, яка є розширенням базової моделі ARIMA. До неї додаються сезонні параметри з періодом для сезонної авторегресії, різниці та ковзного

середнього відповідно. Математично модель записується через поліноми сезонних параметрів. Ідентифікація моделі здійснюється за допомогою аналізу графіків автокореляційної та часткової автокореляційної функцій або шляхом застосування інформаційних критеріїв. Оцінка параметрів проводиться методом максимальної правдоподібності, а перевірка адекватності моделі – через аналіз автокореляції залишків [4].

Узагальнюючи математичну модель, автори [7] наводять схему алгоритму роботи моделі ARIMA (див. рис. 2). Згідно з цим алгоритмом спостерігається покроковий процес створення прогнозової моделі ARIMA. На першому етапі за допомогою статистичного тесту перевіряється стаціонарність даних, тобто встановлюється наявність стабільних статистичних характеристик у часі. У разі не-стаціонарності даних застосовується метод диференціювання для приведення їх до стаціонарного вигляду. На наступному етапі здійснюється визначення параметрів моделі. Для цього аналізуються два графіки: діаграма автокореляційної функції та діаграма часткової автокореляційної функції. Ці візуалізації сприяють виявленню внутрішньої структури даних і дозволяють підібрати оптимальні параметри для моделі. На завершальному етапі для вибору найкращої з альтернативних моделей застосовується інформаційний критерій Байєса. Модель, що характеризується найменшим значенням цього критерію, визнається найефективнішою і обирається як остаточна для побудови прогнозів.

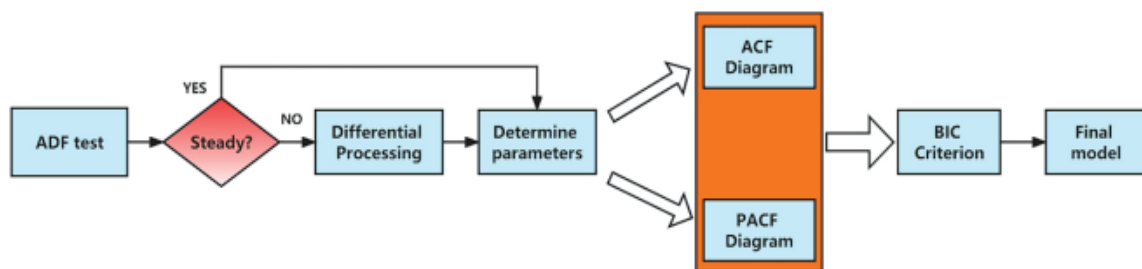


Рис. 2. Схему алгоритму роботи моделі ARIMA [7]

Попри свою популярність, моделі ARIMA/SARIMA мають значні наукові обмеження. Головний недолік полягає в їхньому припущенні про лінійні взаємозв'язки між змінними [4]. Це робить їх неефективними для моделювання складних нелінійних патернів, що є поширеними у реальних даних, наприклад, у прогнозуванні попиту на електроенергію, що залежить від динамічних та нелінійних факторів, таких як погодні умови чи економічна активність.

Обмеження ARIMA/SARIMA стали рушійною силою для розвитку гібридних моделей. Адже більшість реальних часових рядів містять як лінійні, так і нелінійні патерни. Один з відомих підходів полягає у моделюванні лінійної частини даних (тренди та сезонність), з подальшою передачею залишків, що містять нелінійні патерни, нелінійній моделі для подальшого аналізу та прогнозування [7]. Таким чином, нелінійна модель фокусується виключно на патернах, пропущених лінійним підходом, що призводить до значного підвищення загальної точності.

Прогностична модель Prophet – розроблена Facebook, є прогностичною моделлю, що ґрунтується на регресійній моделі та розкладає часовий ряд на три основні компоненти: тренд, сезонність (річну, тижневу, денну) та свята. Тренд може бути лінійним або логістичним, а сезонність моделюється за допомогою рядів Фур'є [4].

Модель була розроблена для бізнес-користувачів і вимагає мінімального досвіду в аналізі часових рядів, що робить її надзвичайно зручним інструментом [1]. Prophet автоматично обробляє пропущені значення, викиди та зміни трендів, що значно спрощує етап попередньої обробки даних [6]. Крім того, модель легко дозволяє включати зовнішні фактори, такі як свята та спеціальні події, через «додаткові регресори» [1].

Prophet фокусує увагу на декомпозиції ряду. Спочатку він підлаштовує глобальний тренд $g(t)$ (з можливими логістичними обмеженнями), відокремлює компонент сезонності $s(t)$ за допомогою Фур'є-базисів та додає фіксовані ефекти $h(t)$. Така структура робить модель частково інтерпретованою, оскільки можна окремо розглядати тренд та внесок сезонностей/свят [1; 4]. Після оцінки ці компоненти екстраполюються на майбутнє, після чого сумуються для отримання прогнозу. Модель Prophet можна записати як:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t,$$

де $g(t)$ – тренд (тут рісесевіс-лінійний або логістичний), $s(t)$ – сумарна сезонна компонента (різні періодичності), $h(t)$ – ефект свят / спеціальних подій, а ε_t – шум. Точки змінного тренду підбираються автоматично, або задаються вручну, а за потреби $g(t)$ може бути логістичним зі строгим насичен-

ням. Сезонність $s(t)$ задається набором тригонометричних базисів для річних, тижневих та інших циклів (за замовчуванням порядок 10 для року, 3 – для тижня). Оцінка моделі проводиться байєсівським методом, що автоматично налаштовує чутливість до змін тренду та інших параметрів [1]. Дослідження показують, що, попри ефективне моделювання трендів та сезонності, залишки моделі Prophet часто містять патерни, що свідчать про наявність невивчених залежностей між послідовними помилками [7].

Це означає, що модель може ігнорувати інформацію, що міститься у нелінійних та автокорельованих помилках [7]. Як наслідок, для досягнення найвищої точності її часто потрібно комбінувати з іншими, більш технічно складними методами.

Моделі глибокого навчання, такі як LSTM (Long Short-Term Memory), RNN (Recurrent Neural Networks) та CNN (Convolutional Neural Networks), розвинули прогнозування часових рядів, оскільки вони здатні автоматично вивчати складні, нелінійні залежності без ручного відбору ознак. RNN є фундаментальною архітектурою для послідовних даних, але залежить від проблеми зникаючого градієнта, що перешкоджає моделюванню довгострокових залежностей. LSTM – це різновид рекурентних нейронних мереж, створений для подолання проблеми затухання градієнта та забезпечення довготривалої пам'яті [2].

LSTM – це глибока нейронна мережа, що обробляє часову послідовність по кроках. На відміну від традиційного ARIMA, LSTM «вчиться» прямо на вхідних даних, без явного розділення на тренд чи сезонність. Зазвичай мережа складається з одного або кількох шарів LSTM, за якими слідує щільний шар для прогнозу. Вхід може містити кілька вхідних ознак (наприклад, попередні витрати, температуру тощо). Після навчання мережа прогнозує наступні значення послідовності, використовуючи свої збережені ваги. LSTM відрізняються великою гнучкістю, адже вони можуть захоплювати складні нелінійні залежності у даних і «пам'ятати» тривалі часові зв'язки [3]. Також LSTM підтримують онлайн-оновлення, що полягає у тому, що ваги мережі можна підлаштовувати по мірі надходження нових даних, а також використовувати перенавчання (transfer learning) на схожих задачах [3]. Дослідження показують, що LSTM можуть добре навчатися навіть на відносно невеликому обсязі даних (кілька днів чи тижнів спостережень достатньо для «розігріву» моделі), а також зазвичай забезпечують вищу точність прогнозів порівняно з простими моделями чи однорівневими RNN [5]. Зокрема, у задачі короткострокового прогнозу енергоспоживання LSTM давали менші помилки, ніж звичайні RNN [5], а у прогнозуванні водоспоживання LSTM показали найкращі результати серед традиційних методів [3]. LSTM можуть вивчати нелінійні ефекти, включно з багаторазовими сезонностями чи впливом додаткових ознак.

Гібридні моделі поєднують сильні сторони різних підходів, зазвичай – статистичних і нейромережних. Зокрема, дієвим поєднанням ARIMA (SARIMA) з LSTM чи Prophet з LSTM. У відповідності до даного алгоритму спочатку окреслюється лінійна складова ряду за допомогою ARIMA, після чого залишки моделі (невраховані нерегулярні коливання) прогнозуються LSTM. Сукупний прогноз отримується як сума прогнозу ARIMA та поправок LSTM. Такий підхід дозволяє ARIMA захопити тренд та сезонність, а LSTM – нелінійну флуктуацію. У дослідженні [6] <https://www.mdpi.com/1996-1073/16/12/4739-#:~:text=combined%20to%20provide%20the%20final,90%2C%20respectively> на прикладі пікового споживання електроенергії гібридна модель ARIMA-LSTM дала кращі результати (RMSE гібриду становив 7.35, тоді як ARIMA – 49.90, LSTM – 18.74) [6].

Інша гібридна модель Prophet+LSTM працює аналогічно, спочатку Prophet моделює тренд, сезонність і ефекти свят, потім LSTM навчається на залишках після віднімання прогнозу Prophet [1]. В гібридній моделі Prophet+LSTM тренд і сезонність моделі стають прозорими (як у Prophet), а їх результати доповнюються «інтелектом» нейромережі для нестандартних коливань [1].

Гібридна модель може забезпечити вищу точність прогнозу, комбінуючи переваги обох методів. За рахунок ARIMA/Prophet вона не втрачає інтерпретованості щодо основних компонент, а LSTM дає змогу моделювати складні залишкові патерни. Науковці довели, що гібридні моделі часто перевершують одиночні моделі (згаданий ARIMA-LSTM дав помітний приріст точності) [6]. Однак гібридні моделі значно складніші і ресурсомісткіші у реалізації, оскільки потрібно тренувати дві моделі, що вимагає більше обчислень. Вибір структури гібриду (послідовний чи паралельний) та баланс ваг може додатково ускладнити налаштування.

Гібридні моделі часто застосовують у задачах прогнозування енергоспоживання, де потрібна висока точність. Так, ARIMA+LSTM успішно використано для прогнозу пікових навантажень електромереж [6]. Пропозиції Prophet+LSTM з'являються у задачах, де важливі як календарні ефекти, так і складні залишки (наприклад, у прогнозуванні попиту на пропускну спроможність у веб-сервісах) [1]. Гібридні моделі доцільно застосовувати, коли є помітна лінійна складова (яку варто виділити) та одночасно необхідно моделювати складніші залежності. При цьому слід пам'ятати про збільшені обчислювальні витрати і необхідність узгодження двох частин моделі.

Узагальнена класифікація гібридних моделей на основі Prophet представлена на рис. 3.

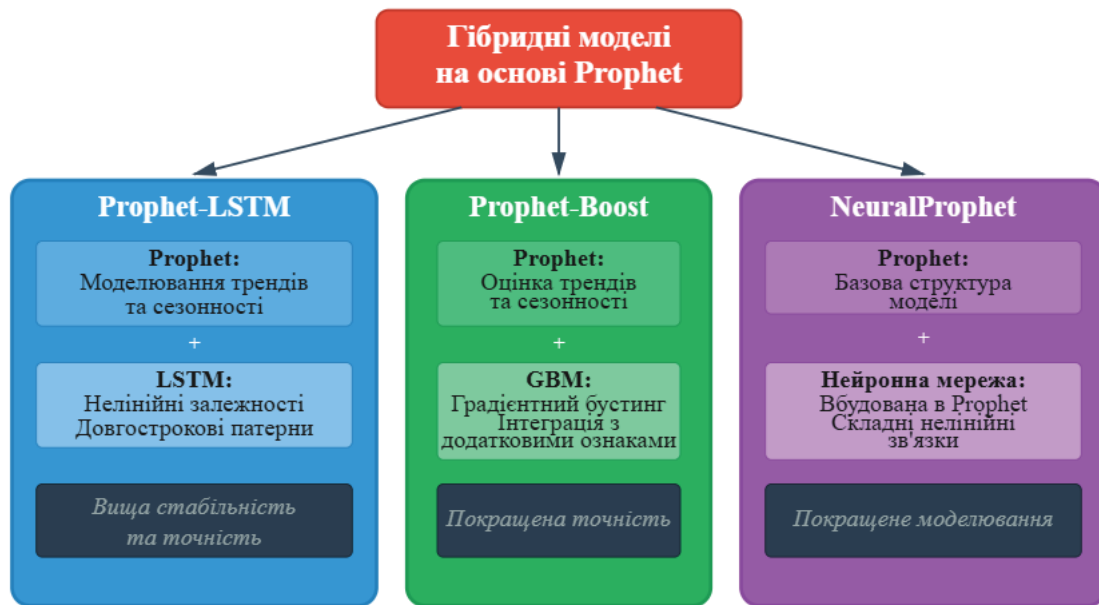


Рис. 3. Класифікація гібридних моделей на основі Prophet

Висновки. Проведене дослідження методів прогнозування даних з лічильників дозволяє стверджувати, що вибір оптимального підходу залежить від специфіки конкретного завдання, характеристик вхідних даних та вимог до точності та інтерпретованості результатів. Класичні статистичні моделі ARIMA та SARIMA залишаються ефективним інструментом для короткострокового прогнозування стаціонарних часових рядів з чіткою лінійною структурою та сезонністю. Їхня головна перевага полягає у високому рівні інтерпретованості та можливості роботи з відносно невеликими обсягами даних, що робить їх придатними для застосування в умовах обмежених обчислювальних ресурсів. Модель Prophet демонструє значні переваги як практичний інструмент для бізнес-застосувань завдяки автоматизації процесу моделювання та можливості врахування впливу календарних подій і свят. Моделі глибокого навчання, зокрема LSTM, відкривають якісно нові можливості для прогнозування складних нелінійних часових рядів завдяки здатності автоматично виявляти приховані патерни та враховувати довгострокові залежності.

Особливої уваги заслуговують гібридні підходи, що поєднують сильні сторони різних методологій. Комбінація ARIMA та LSTM або Prophet і LSTM дозволяє використовувати переваги статистичних моделей для виявлення глобальних трендів і сезонності з одночасним застосуванням потужності нейронних мереж для моделювання складних нелінійних залишків. Дослідження демонструють, що гібридні моделі забезпечують найвищу точність прогнозування, хоча й ціною підвищеної обчислювальної складності та ускладненого процесу розробки й підтримки моделей.

В контексті українських реалій, зокрема необхідності ефективного управління енергетичними ресурсами в умовах пошкодженої інфраструктури, питання вибору методу прогнозування набуває критичного значення.

Подальші дослідження можуть бути спрямовані на розробку адаптивних систем прогнозування, здатних автоматично обирати оптимальний метод залежно від характеристик вхідних даних та поточних умов функціонування системи.

Література

1. Guruge Pasan Bhanu, Priyadarshana Y. H. P. P. Time series forecasting-based Kubernetes autoscaling using Facebook Prophet and Long Short-Term Memory. *Frontiers in Computer Science*. 2025. Vol. 7. DOI: 10.3389/fcomp.2025.1509165.
2. Islam, Badar ul, Ahmed, Shams Forruque. Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks. *Mathematical Problems in Engineering*. 2022. Vol. 2022. Article ID 2316474. DOI: 10.1155/2022/2316474.
3. Kühnert C., Gonuguntla N. M., Krieg H., Nowak D., Thomas J. A. (2021). Application of LSTM Networks for Water Demand Prediction in Optimal Pump Control. *Water*. 2021. Vol. 13. No. 5. Article ID 644. DOI: 10.3390/w13050644.
4. Kwarteng S. B., Andreevich P. A. Comparative Analysis of ARIMA, SARIMA and Prophet Model in Forecasting. *Research & Development*. 2024. Vol. 5, No. 4. P. 110–120. DOI: 10.11648/j.rd.20240504.13.
5. Mahjoub S., Chrifi-Alaoui L., Marhic B., Delahoche L. (2022). Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks *Sensors*. 2022. Vol. 22. No. 11. Article ID. 4062. DOI: 10.3390/s22114062.

6. Pierre A. A., Akim S. A., Semeny A. K., Babiga B. (2023). Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies*. 2023. Vol. 16. No. 12. Article ID 4739. DOI: 10.3390/en16124739.
7. Ruiyan Zhou, Xingchen Zhang. Short-term power load forecasting based on ARIMA-LSTM. *Journal of Physics: Conference Series*. 2024. Vol. 2803. Article ID 012002. DOI: 10.1088/1742-6596/2803/1/012002.
8. Samuel John Parreño. Forecasting Electricity Consumption in the Philippines Using ARIMA Models. *International Journal of Machine Learning and Computing*. 2022. Vol. 12. No. 6. P. 279–285. DOI: 10.18178/ijmlc.2022.12.6.1112.
9. Wang W., Shao J., Jumahong H. Fuzzy inference-based LSTM for long-term time series prediction. *Scientific Reports*. 2023. Vol. 13. Article ID 20359. DOI: 10.1038/s41598-023-47812-3.

References

1. Guruge P. B., Priyadarshana Y. H. P. P. (2025). Time series forecasting-based Kubernetes autoscaling using Facebook Prophet and Long Short-Term Memory. *Frontiers in Computer Science*. 7. DOI: <https://doi.org/10.3389/fcomp.2025.1509165>.
2. Islam B. U., Ahmed S. F. (2022). Short-term electrical load demand forecasting based on LSTM and RNN deep neural networks. *Mathematical Problems in Engineering*. 2022. Article 2316474. DOI: <https://doi.org/10.1155/2022/2316474>.
3. Kühnert C., Gonuguntla N. M., Krieg H., Nowak D., Thomas J. A. (2021). Application of LSTM networks for water demand prediction in optimal pump control. *Water*. 13 (5). 644. DOI: <https://doi.org/10.3390/w13050644>.
4. Kwarteng S. B., Andreevich P. A. (2024). Comparative analysis of ARIMA, SARIMA and Prophet model in forecasting. *Research & Development*. 5 (4). 110–120. DOI: <https://doi.org/10.11648/j.rd.20240504.13>.
5. Mahjoub S., Chrifi-Alaoui L., Marhic B., Delahoche L. (2022). Predicting energy consumption using LSTM, multi-layer GRU and Drop-GRU neural networks. *Sensors*. 22 (11). 4062. DOI: <https://doi.org/10.3390/s22114062>.
6. Pierre A. A., Akim S. A., Semeny A. K., Babiga B. (2023). Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU approaches. *Energies*. 16 (12). 4739. DOI: <https://doi.org/10.3390/en16124739>.
7. Zhou R., Zhang X. (2024). Short-term power load forecasting based on ARIMA-LSTM. *Journal of Physics: Conference Series*. 2803, 012002. DOI: <https://doi.org/10.1088/1742-6596/2803/1/012002>.
8. Parreño S. J. (2022). Forecasting electricity consumption in the Philippines using ARIMA models. *International Journal of Machine Learning and Computing*. 12 (6). 279–285. DOI: <https://doi.org/10.18178/ijmlc.2022.12.6.1112>.
9. Wang W., Shao J., Jumahong H. (2023). Fuzzy inference-based LSTM for long-term time series prediction. *Scientific Reports*. 13. Article 20359. DOI: <https://doi.org/10.1038/s41598-023-47812-3>.

Дата першого надходження статті до видання: 20.10.2025

Дата прийняття статті до друку після рецензування: 24.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

УДК 004.94:004.42

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-18>

АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ВЕРИФІКАЦІЇ ТА ВИКОНАННЯ КООПЕРАТИВНИХ СЦЕНАРІЇВ У РОЗПОДІЛЕНИХ ТРЕНАЖЕРНИХ СИСТЕМАХ

Шевченко С. С., аспірант, Інститут проблем моделювання в енергетиці ім. Г. Є. Пухова НАН України, Київ, Україна. <https://orcid.org/0009-0000-3146-3591>, st.shevchenko@united productions

Анотація. У статті розв'язується актуальна науково-прикладна задача програмної реалізації та автоматизованої верифікації сценаріїв для багатокористувацьких 3D-тренажерів в енергетиці. В умовах зростання складності технологічних процесів та вимог до резильєнтності критичної інфраструктури, ефективність навчання персоналу залежить не лише від індивідуальних навичок, але й від здатності до скоординованої командної взаємодії. Існуючі підходи до створення контенту часто покладаються на ненадійну ручну перевірку або використовують великозагові формальні методи, складні для інтеграції у веб-середовища. На основі формальної моделі ролевої взаємодії (MPB) запропоновано комплексне алгоритмічне забезпечення, що включає структуру даних на базі JSON, алгоритм статичного аналізу та ядро динамічного виконання. Розроблений метод статичної верифікації, що базується на модифікованому пошуку в глибину (DFS), дозволяє виявляти специфічні логічні колізії («синхронізаційні тупики» (Sync Deadlocks) та недосяжні стани) безпосередньо на етапі проектування. Для забезпечення коректного виконання сценаріїв розроблено алгоритм роботи серверного ядра (Runtime Engine), що реалізує архітектуру «авторитетного сервера», гарантуючи узгодженість станів учасників навіть при стохастичних системних подіях. Ефективність підходів підтверджено експериментально. Навантажувальне тестування на синтетичних графах продемонструвало лінійну залежність часу верифікації від розмірності сценарію, що підтверджує можливість перевірки складних структур у реальному часі. Порівняльний аналіз показав значну перевагу автоматизованого методу над експертним за швидкістю та точністю виявлення помилок. Також розглянуто використання верифікатора як «фільтра безпеки» при генерації сценаріїв за допомогою великих мовних моделей (LLM), що дозволяє нівелювати ризики логічних галюцинацій штучного інтелекту.

Ключові слова: кооперативне навчання, 3D-тренажер, верифікація сценаріїв, алгоритм синхронізації, клієнт-серверна архітектура, перевірка моделей.

ALGORITHMIC SUPPORT FOR VERIFICATION AND EXECUTION OF COOPERATIVE SCENARIOS IN DISTRIBUTED TRAINING SYSTEMS

Shevchenko S. S., PhD Student, G. E. Pukhov Institute for Modelling in Energy Engineering of the National Academy of Sciences of Ukraine, Kyiv, Ukraine. <https://orcid.org/0009-0000-3146-3591>, st.shevchenko@united productions

Abstract. The article addresses the urgent scientific and applied problem of software implementation and automated verification of scenarios for multi-user 3D simulators in the energy sector. Given the increasing complexity of technological processes and requirements for critical infrastructure resilience, personnel training efficiency depends not only on individual skills but also on the ability for coordinated teamwork. Existing approaches to content creation often rely on unreliable manual checking or use heavyweight formal methods that are difficult to integrate into web environments. Based on the formal Role Interaction Model (RIM), the paper proposes comprehensive algorithmic support, including a JSON-based data structure, a static analysis algorithm, and a dynamic execution engine. The developed static verification method, based on a modified Depth-First Search (DFS), allows for the detection of specific logical collisions ("Sync Deadlocks" and unreachable states) directly at the design stage. To ensure correct scenario execution, a server kernel algorithm (Runtime Engine) has been developed, implementing an "authoritative server" architecture that guarantees the consistency of participant states even during stochastic system events. The effectiveness of the proposed approaches is confirmed experimentally. Stress testing on synthetic graphs demonstrated a linear dependence of verification time on scenario dimension, confirming the feasibility of verifying complex structures in real-time. A comparative analysis showed a significant advantage of the automated method over expert analysis in terms of speed and accuracy of error detection. The paper also discusses the potential of using the verifier as a "safety filter" when generating scenarios using Large Language Models (LLM), which helps mitigate the risks of logical hallucinations in artificial intelligence.

Keywords: cooperative learning, 3D simulator, scenario verification, synchronization algorithm, client-server architecture, model checking.

Вступ

Забезпечення стійкості (резильєнтності) об'єктів критичної інфраструктури в умовах сучасних викликів вимагає від персоналу не лише індивідуальних технічних навичок, але й здатності до ефективної командної взаємодії [1]. Традиційні методи навчання часто не дозволяють відпрацювати сценарії складної координації дій (наприклад, спільний ремонт агрегату або ліквідація аварії) без ризику для обладнання. Вирішенням цієї проблеми є впровадження інтерактивних автоматизованих навчально-тренувальних платформ (АНТП) з підтримкою багатокористувацького режиму у 3D-середовищі [2].

У попередніх дослідженнях автора [3; 13] було запропоновано комплексний підхід до проектування таких систем та розроблено формальну модель ролевої взаємодії (MPB), яка математично описує стани об'єктів, ролі учасників та точки їх синхронізації. Проте наявність формальної моделі є необхідною, але недостатньою умовою для створення працездатного програмного продукту.

Практична реалізація MPB у програмному коді стикається з низкою інженерних викликів: необхідністю обробки паралельних подій від різних клієнтів, забезпеченням цілісності даних на сервері та, що найважливіше, гарантуванням логічної коректності сценаріїв. Складні сценарії з розгалуженою логікою та множинними точками синхронізації схильні до помилок проектування, таких як "дедлоки" (deadlocks) – стани взаємного блокування, коли учасники очікують дій один від одного, які ніколи не відбудуться. Ручне тестування таких сценаріїв є трудомістким та ненадійним.

Аналіз останніх досліджень і публікацій

Питання створення ефективних тренажерних комплексів для енергетики розглядаються у багатьох сучасних дослідженнях [1; 4; 18]. Особлива увага приділяється забезпеченню резильєнтності енергосистем [2] та використанню VR-технологій для навчання персоналу в умовах підвищеного ризику [1]. Разом з тим, сучасні дослідження [15] підтверджують, що інтеграція генеративного штучного інтелекту у VR-середовища значно підвищує адаптивність навчання, проте вимагає суворого контролю якості контенту.

Однак, із зростанням складності сценаріїв навчання критично важливим стає питання їхньої логічної коректності. Класичним підходом до моделювання та верифікації паралельних процесів є апарат мереж Петрі [6, 9], який і сьогодні активно розвивається для аналізу програмного коду [7]. Також застосовуються методи на основі темпоральної логіки [8] та дерев поведінки (Behavior Trees). Фундаментальні принципи виявлення взаємних блокувань (deadlocks) були закладені ще у класичних працях [10] і зараз адаптуються для розподілених систем [11].

Проте, існуючі формальні методи часто вимагають високої математичної кваліфікації від методистів, що ускладнює їх інтеграцію у сучасні Web/VR-редактори сценаріїв. У роботах, присвячених мультикористувацькій взаємодії у VR [5], акцент робиться на UX та мережевій синхронізації, але недостатньо уваги приділяється автоматизованій верифікації логіки сценарію «на льоту». Це обумовлює необхідність розробки спеціалізованого легкого алгоритмічного забезпечення.

Постановка задачі

Метою роботи є розробка алгоритмічного забезпечення для автоматизації процесів верифікації та виконання кооперативних сценаріїв у середовищі розподілених тренажерних систем.

Для досягнення мети необхідно вирішити такі завдання:

1. Розробити структуру даних для програмного представлення формальної моделі MPB, придатну для обробки в сучасних веб-орієнтованих архітектурах.
2. Створити алгоритм статичного аналізу (формальної верифікації) сценарію для виявлення логічних колізій до етапу компіляції або запуску ("Model Checking").
3. Розробити алгоритм роботи серверного ядра (Engine), що забезпечує динамічну обробку подій та синхронізацію клієнтів згідно з логікою MPB.

Об'єктом дослідження є процеси управління даними в інтерактивних багатокористувацьких системах. Робота базується на архітектурі системи, описаній у патенті на корисну модель № 149786 [4], зокрема, деталізує логіку роботи блоків «Обробка сценарію» та «Центральна логіка і перевірка результатів».

Математична формалізація та структура даних

Для забезпечення строгості верифікації перейдемо від описової моделі до формальної. Сценарій кооперативного навчання можна представити як орієнтований граф станів.

Визначення 1. Граф сценарію G – це кортеж $G = \langle R, S, T, s_0 \rangle$, де:

- $R = \{r_1, r_2, \dots, r_k\}$ – скінченна множина ролей учасників (наприклад, «Майстер», «Оператор»);
- S – скінченна множина станів системи, яка є об'єднанням локальних станів усіх ролей $S = \bigcup_{r \in R} S_r$;
- $s_0 \in S$ – початковий стан;
- T – множина переходів між станами.

Визначення 1.1 (Глобальний стан). Глобальний стан системи S_{global} визначається як вектор локальних станів усіх активних ролей у момент часу τ :

$$S_{global}^{\tau} = \langle s_{curr}^{r_1}, s_{curr}^{r_2}, \dots, s_{curr}^{r_k} \rangle$$

Перехід системи з одного глобального стану в інший відбувається асинхронно при зміні стану будь-якої окремої ролі.

Визначення 1.2 (Умова завершення). Сценарій вважається успішно завершеним, якщо вектор глобального стану досягає термінальної конфігурації, де кожна роль r_i знаходиться у своєму цільовому стані $s_{end} \in \Omega_{r_i}$.

Визначення 2. Перехід $t \in T$ визначається як кортеж:

$$t = \langle s_{src}, s_{dst}, r, E_{trig}, C_{guard} \rangle$$

де:

- $s_{src}, s_{dst} \in S_r$ – вихідний та цільовий стани для ролі r ;
- E_{trig} – подія-тригер (дія користувача або системний таймер);
- C_{guard} – умова переходу (guard condition).

Саме компонент C_{guard} дозволяє реалізувати механізм синхронізації. У контексті запропонованого підходу виділимо специфічний тип переходу – **синхронізаційний перехід**.

Визначення 3 (Семантика синхронізації). Перехід t для ролі r_i є синхронізаційним, якщо його умова C_{guard} залежить від стану іншої ролі r_j :

$$C_{sync}(r_i) = true \Leftrightarrow CurrentState(r_j) == s_{req}$$

Це реалізує механізм «бар'єру» або «рукоштовування» (handshake), коли виконання сценарію для r_i блокується до моменту, поки r_j не досягне необхідного стану s_{req} .

Визначення 4 (Умова Sync Deadlock). Логічна колізія типу «Синхронізаційний тупик» (Sync Deadlock) виникає, якщо для активного стану $s_{curr} \in S_{r_i}$ існує перехід із умовою синхронізації з r_j , але у графі ролі r_j відсутній шлях від його поточного стану до необхідного стану s_{req} :

$$\sigma path(s_{curr}^j \rightarrow s_{req})$$

Наведена математична модель трансформується у структуру даних JSON для обробки програмними засобами. Структура об'єкта *Scenario* включає:

- *roles* – масив ідентифікаторів ролей (наприклад, ['master', 'mechanic']);
- *states* – словник станів, де ключем є ID стану (s_i), а значенням – об'єкт з описом властивостей та доступних переходів;
- *transitions* – масив правил переходу, що відповідає множині T у формальній моделі.

Особливістю запропонованої структури є явне виділення умов синхронізації у полі *condition* об'єкта переходу. Нижче наведено фрагмент JSON-структури, що описує точку синхронізації, де «Механік» (*mechanic*) очікує виконання дії «Майстром» (*master*):

```
"s1": {
  "description": "Очікування фіксації деталі",
  "transitions": [
    {
      "action_id": "sync_action_01",
      "type": "synchronization",
      "role": "mechanic",
      "condition": {
        "partner_role": "master",
        "partner_state_required": "part_fixed"
      },
      "target_state": "s2"
    }
  ]
}
```

Така структура дозволяє інтерпретатору сценарію (Блок 6 на схемі архітектури [4]) однозначно ідентифікувати поточний стан системи та перевіряти умови переходів. Загальна логіка взаємодії розроблених програмних модулів у процесі проектування та виконання сценарію наведена на рис. 1.

Алгоритм формальної верифікації сценарію (Static Analysis)

Однією з ключових проблем проектування нелінійних сценаріїв є ризик виникнення логічних колізій, таких як тупикові стани (deadlocks), коли учасники взаємно блокують один одного, або недосяжні стани (unreachable states). Для вирішення цієї проблеми розроблено алгоритм статичного аналізу графа станів сценарію, який виконується перед завантаженням сценарію на сервер.

Алгоритм базується на модифікованому методі пошуку в глибину (Depth-First Search, DFS) [12] з емуляцією станів партнерів.

Формальний опис алгоритму верифікації:

1. **Ініціалізація:** Створення стеку для обходу *Stack* та множини відвіданих станів *Visited*. Поміщення початкового стану s_0 у стек.

2. **Обхід:** Поки *Stack* не порожній:
 - о Вилучити поточний стан s_{curr} зі стеку.
 - о Якщо $s_{curr} \in Visited$, перейти до наступної ітерації (запобігання зацикленню).
 - о Додати s_{curr} до множини *Visited*.
 - о Визначити множину доступних переходів T_{avail} із s_{curr} .
 - о Якщо $T_{avail} = \emptyset$ і $s_{curr} \neq s_{final}$ (цільовий стан), зафіксувати помилку типу *Deadlock*.
3. **Перевірка синхронізації:** Для кожного переходу $t \in T_{avail}$:
 - о Якщо t є точкою синхронізації (тип *synchronization*), перевірити досяжність необхідного стану партнера ($s_{partner_req}$) у паралельній гілці графа. Для умовних розгалужень (OR-splits) алгоритм використовує стратегію екзистенційної досяжності: стан вважається досяжним, якщо існує хоча б один шлях виконання, що веде до нього, що є достатнім для сценаріїв навчання з варіативним проходженням.
 - о Якщо $s_{partner_req}$ недосяжний, зафіксувати помилку типу *Sync Deadlock*.
4. **Завершення:** Якщо після обходу всього графа цільовий стан s_{final} відсутній у множині *Visited*, зафіксувати помилку *Goal Unreachable*.

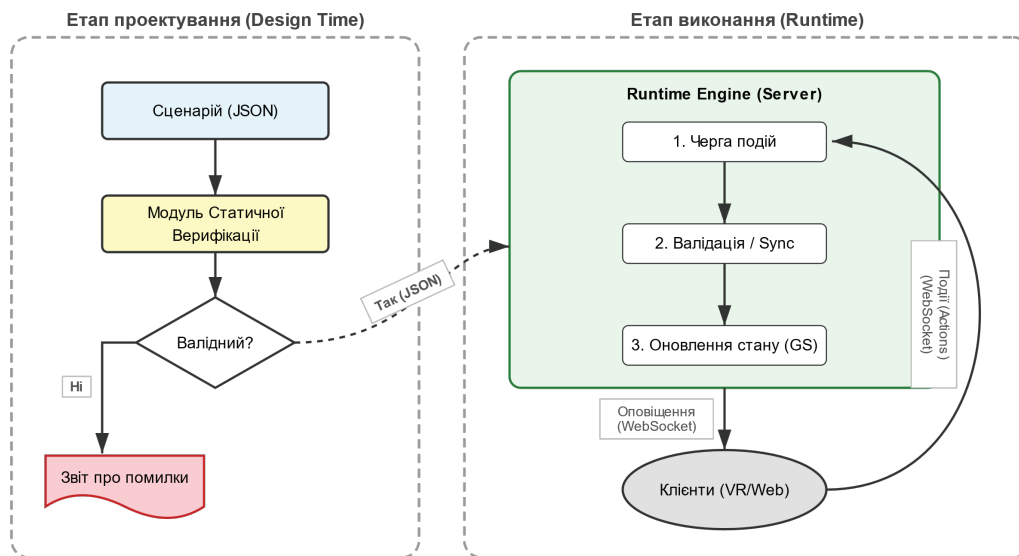


Рис. 1. Структурно-функціональна схема алгоритмічного забезпечення системи

Застосування цього алгоритму дозволяє скоротити час відлагодження сценаріїв, автоматично вказуючи розробнику на проблемні вузли логіки.

Алгоритм динамічної синхронізації та виконання (Runtime Engine)

Безпосереднє виконання сценарію забезпечується серверним ядром ("двигуном"), яке реалізує функції блоку "Центральна логіка" [4]. Ядро працює за принципом розширеної подійно-орієнтованої машини станів (Event-driven State Machine), яка обробляє гетерогенний потік подій.

Алгоритм обробки подій у реальному часі виглядає наступним чином:

1. Реєстрація подій:

У єдину чергу обробки надходять події двох типів:

- о Користувачські (e_{user}): дії учасників (натискання, переміщення об'єктів), отримані через мережевий інтерфейс.

- о Системні (e_{sys}): події, генеровані внутрішнім планувальником (спрацювання таймерів, випадкові відмови обладнання, зміна параметрів середовища).

2. Валідація:

- о Для e_{user} : перевірка, чи дозволена дія у поточному стані s_{curr} для ролі r .
- о Для e_{sys} : перевірка умов активації (наприклад, чи не завершився сценарій, перевірка таймерів).

3. **Обробка синхронізації:** Якщо подія вимагає синхронізації, сервер перевіряє стан інших учасників у глобальному об'єкті сесії *GS*. Якщо умова не виконана, клієнту надсилається статус *WAIT*. Якщо умова не виконана, сервер додає ідентифікатор сесії та події у чергу очікування *PendingEvents*, а клієнту надсилається статус *WAIT* (блокування інтерфейсу). При кожній наступній зміні глобального стану (*GS*) сервер перевіряє чергу *PendingEvents*: якщо умови для відкладених подій стають істинними, вони автоматично виконуються, а клієнтам надсилається *RESUME*.

4. **Оновлення стану:** $GS' = update(GS, e)$. При цьому системні події можуть змінювати стан примусово (наприклад, перехід у стан "Аварія" при вичерпанні часу).

5. Оповіщення: Генерація та розсилка пакету оновлень (*State Update*) усім клієнтам (включно з повідомленнями про системні події, наприклад, аварійну зупинку за таймером) для синхронізації візуального відображення на клієнтських пристроях.

Використання асинхронної моделі обробки подій (наприклад, на базі Node.js) дозволяє обслуговувати сотні паралельних сесій з мінімальною затримкою, що є критичним для комфортного сприйняття VR-середовища.

Експериментальні дослідження та результати

Для всебічної оцінки запропонованого алгоритмічного забезпечення було проведено два етапи експериментів: оцінка обчислювальної ефективності (масштабованості) на синтетичних даних та валідація практичної користі на реальному сценарії.

1. Аналіз масштабованості (Scalability Analysis).

Метою цього етапу було визначення залежності часу верифікації від складності сценарію. Для цього було розроблено генератор синтетичних графів сценаріїв, який створює випадкові структури з заданою кількістю станів (N) та переходів (M), імітуючи розгалужену логіку з циклами та точками синхронізації.

Тестування проводилося на стенді з процесором Intel Core i5 (3.2 GHz) та 16 GB RAM у середовищі Node.js runtime. Результати вимірювань (усереднені за 10 запусками для кожної розмірності) наведено у Таблиці 1.

Таблиця 1
Залежність часу верифікації від розмірності графа сценарію

Кількість станів (N)	Кількість переходів (M)	Час верифікації (мс)	Споживання пам'яті (MB)
50	120	15	4,2
100	280	32	6,5
500	1450	145	18,1
1000	3100	290	34,4
5000	15500	1480	128,0

Як видно з таблиці, алгоритм демонструє лінійну залежність часу виконання $T \approx k \cdot (N + M)$, що узгоджується з теоретичною складністю алгоритму пошуку в глибину ($O(N + M)$). Це означає, що навіть для надзвичайно складних сценаріїв (5000 станів, що відповідає багатогодинному тренінгу на АЕС), час автоматичної перевірки не перевищує 1.5 секунди. Це дозволяє інтегрувати верифікатор безпосередньо у веб-редактор сценаріїв для перевірки «на льоту» (real-time validation).

2. Порівняльний аналіз на реальному сценарії.

Другий етап експерименту мав на меті підтвердити практичну цінність методу у порівнянні з традиційним підходом. В якості об'єкта використано реальний сценарій «Спільне обслуговування на-сосного агрегату» (45 станів, 3 ролі, 12 точок синхронізації).

Для порівняння застосовано метод експертної оцінки («настільної перевірки» діаграми). Результати наведено у Таблиці 2.

Таблиця 2
Порівняння ефективності методів виявлення логічних помилок

Критерій	Експертний аналіз	Автоматизована верифікація
Час виконання	~90 хвилин	< 15 мс
Повнота покриття	~80 % (вибіркова перевірка гілок)	100% (повний перебір всіх досяжних станів)
Виявлені помилки	2 з 3 (одну пропущено)	3 з 3 (100 %)
Людський фактор	Високий (втома, втрата контексту)	Відсутній

Експеримент показав, що автоматизований алгоритм не лише радикально перевершує ручну перевірку за швидкістю, але й, що важливіше, гарантує виявлення помилок синхронізації у глибоких гілках сценарію, які важко відстежити людині через когнітивне перевантаження при аналізі паралельних процесів.

Обговорення результатів

Аналіз отриманих даних дозволяє позиціонувати запропонований підхід у контексті існуючих рішень. На відміну від класичних формальних методів, таких як мережі Петрі [6; 7] або використання інструментів Model Checking (наприклад, SPIN або UPPAAL) [8], запропонований алгоритм не вимагає трансляції сценарію у проміжні математичні нотації (Promela, PNML). Це забезпечує можливість інтеграції верифікатора безпосередньо у веб-редактори навчальних курсів (LMS), дозволяючи методистам без математичної підготовки перевіряти логіку «на льоту».

1. Порівняння з аналогами.

Як показав експеримент (Таблиця 1), час верифікації для типових сценаріїв вимірюється мілісекундами. Для порівняння, налаштування та запуск зовнішнього верифікатора типу SPIN для аналогічної задачі займає хвилини (експорт моделі, компіляція верифікатора, прогонка). Отже, ми свідомо йдемо на компроміс: ми жертвуємо універсальністю (наш алгоритм шукає лише специфічні помилки синхронізації ролей) заради швидкодії та юзабіліті. Це робить метод «легковою» альтернативою важким академічним інструментам.

Хоча універсальні інструменти Model Checking (наприклад, SPIN або UPPAAL) забезпечують вищу математичну строгість та здатні виявляти складні колізії (circular waits), вони вимагають конвертації сценарію у специфічну мову (Promela/TLA+) та мають експоненційну складність верифікації ($O(e^N)$). Натомість, запропонований метод працює напряду з JSON-графом за лінійний час $O(N + M)$, що робить його придатним для вбудовування безпосередньо у редактор для перевірки в режимі реального часу (editor-time validation), де миттєвий зворотний зв'язок важливіший за академічну повноту перевірки.

Також варто розглянути запропонований підхід у контексті генеративного штучного інтелекту. Хоча використання великих мовних моделей (LLM) дозволяє автоматизувати створення сценаріїв, імовірна природа цих моделей призводить до ризику виникнення «галюцинацій логіки» [16]. Як показано у роботі [17], застосування формальної верифікації (Model Checking) до згенерованого коду дозволяє виявити значну кількість функціональних помилок ще до етапу запуску. Саме тому запропонований нами алгоритм верифікації може слугувати необхідним «фільтром безпеки» (safety filter) у ланцюжку генерації контенту штучним інтелектом, гарантуючи структурну цілісність сценаріїв, створених неймережами.

2. Обмеження методу та динамічні колізії.

Варто зазначити, що запропонований статичний аналіз (розділ 4) гарантує структурну досяжність станів синхронізації (*pairwise synchronization reachability*). Важливо підкреслити, що поточна реалізація алгоритму фокусується на виявленні локальних помилок синхронізації, коли партнер технічно не може дійти до точки зустрічі. Виявлення складних багатосторонніх циклічних блокувань (*circular waits*, де $r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_1$) вимагає побудови глобального графа очікування (*Wait-For Graph*), що підвищує обчислювальну складність і є завданням наступних етапів дослідження. Однак, як показує практика, більшість помилок у навчальних сценаріях мають саме локальний характер.

Щодо виконання сценарію в реальному часі, система не може гарантувати повну відсутність мережових затримок. Це питання вирішується на рівні архітектури Engine (розділ 5). Використання однопотокової моделі обробки подій (*Event Loop* у Node.js) дозволяє лінеаризувати потік запитів від клієнтів. Архітектура типу "Authoritative Server" виступає єдиним джерелом істини (*Single Source of Truth*), що дозволяє уникнути конфліктів стану (*race conditions*) шляхом послідовної обробки подій: якщо два користувачі одночасно надсилають несумісні команди, сервер обробить першу, а другу відхилить валідатором як неактуальну для нового стану.

При цьому модель runtime-виконання покладається на гарантовану доставку та впорядкування повідомлень (*ordered delivery*), що забезпечується транспортним рівнем (протоколи TCP/WebSocket). Питання обробки розривів з'єднання (*network partitions*) або критичних мережових затримок, які можуть вплинути на узгодженість глобального стану, виходять за межі цієї роботи і вирішуються стандартними механізмами реконекту та синхронізації стану (*state reconciliation*) на рівні мережевого клієнта.

3. Практичне значення.

Впровадження таких алгоритмів є критичним для забезпечення резильєнтності тренажерних систем. Усунення логічних помилок на етапі проектування запобігає ситуаціям, коли група з 5 осіб витрачає годину на проходження сценарію, щоб застрягти в кінці через помилку дизайнера. Для енергетичної галузі, де ціна помилки висока, така надійність навчального ПЗ є обов'язковою вимогою.

Висновки

У роботі вирішено задачу інженерної реалізації формальної моделі ролевої взаємодії (MPV) для систем кооперативного навчання.

1. Розроблено JSON-орієнтовану структуру даних, яка дозволяє уніфікувати опис сценаріїв та інтегрувати їх у веб-середовище.

2. Запропоновано та реалізовано алгоритм статичного аналізу, який, на відміну від існуючих підходів, враховує специфіку точок синхронізації ролей, що дозволяє виявляти помилки типу *Sync Deadlock* на етапі проектування.

3. Алгоритм динамічної синхронізації розширено обробкою системних подій, що дозволяє моделювати не лише дії користувачів, а й стохастичні процеси (аварії, таймери), підвищуючи реалістичність навчання.

Подальші дослідження будуть спрямовані на впровадження адаптивних механізмів, які дозволять ядру автоматично спрощувати або ускладнювати сценарій в залежності від успішності дій команди.

Література

1. Shehadeh A., Alshboul O. Enhancing Engineering and Architectural Design Through Virtual Reality and Machine Learning Integration. *Buildings*. 2025. № 3. С. Т. 15. 328.
2. Саух С., Борисенко А. 2025. Моделивання електроенергетичної системи України та оцінювання її резильєнтності в умовах систематичних терористичних атак. *Технічна електродинаміка*. 2025. № 057. Т. 2.
3. Shevchenko S., Shevchenko S. Increasing the Operational Safety of NPP Pumping Equipment by Using Interactive Automated Remote Educational and Training Systems. *Nuclear and Radiation Safety*. 2024. № 1 (101). С. 19–27.
4. Radianti J. et al. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*. 2020. Т. 147. Ст. 103778.
5. Rasch J. et al. Going, Going, Gone: Exploring Intention Communication for Multi-User Locomotion in Virtual Reality. *CHI Conference on Human Factors in Computing Systems*. 2023.
6. Murata T. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*. 1989. № 4. Т. 77. С. 541–580.
7. Zhang K., Liu G. TRustPN: Transforming Rust Source Code to Petri Nets for Checking Deadlocks. *10th International Conference on Control, Decision and Information Technologies (CoDIT)*. 2024.
8. Biggar O., Zamani M. A Framework for Formal Verification of Behavior Trees With Linear Temporal Logic. *IEEE Robotics and Automation Letters*. 2020.
9. Jensen K., Kristensen L. M. *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Springer, 2009.
10. Coffman E. G., Elphick M., Shoshani A. System Deadlocks. *ACM Computing Surveys*. 1971. № 2. Т. 3. С. 67–78.
11. Wijethunga W. A. T. G. R., Kumara B. T. G. S. Systematic Review of Graph Neural Network and Consensus Algorithm-Based Approaches for Proactive Deadlock Detection in Distributed Systems. *Preprints*. 2025.
12. Tarjan R. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*. 1972. № 2. Т. 1. С. 146–160.
13. Шевченко С. С. Комплексний підхід та модель ролевої взаємодії при проектуванні кооперативних сценаріїв. *Електронне моделювання*. 2025. № 5. Т. 47. С. 75–86.
14. Інтерактивна автоматизована дистанційна навчально-тренувальна система : пат. 149786 Україна / С. С. Шевченко. Опубл. 01.12.2021, Бюл. № 48.
15. Hemminki-Reijonen U. et al. Design of generative AI-powered pedagogy for virtual reality environments in higher education. *npj Science of Learning*. 2025. № 1. Т. 10. Ст. 31.
16. Gao L. et al. Constructing a Dialogue-Level Hallucination Evaluation Benchmark for Large Language Models. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 2024.
17. Gadde D. N. et al. All Artificial, Less Intelligence: GenAI through the Lens of Formal Verification. *arXiv preprint arXiv:2403.16750*. 2024.
18. Самойлов В. Д. та ін. Комп'ютерні технології розробки тренажерних систем для енергетичної галузі. *Електронне моделювання*. 2020. № 3. Т. 42. С. 89–98.

References

1. Shehadeh, A., Alshboul, O. (2025). Enhancing Engineering and Architectural Design Through Virtual Reality and Machine Learning Integration. *Buildings*. 15 (3). 328.
2. Saukh, S., Borysenko, A. (2025). Modeling the electric power system of Ukraine and assessing its resilience under conditions of systematic terrorist attacks. *Tekhnichna Elektrodynamika*. 2025. (2) [in Ukrainian].
3. Shevchenko, S., & Shevchenko, S. (2024). Increasing the Operational Safety of NPP Pumping Equipment by Using Interactive Automated Remote Educational and Training Systems. *Nuclear and Radiation Safety*. 1 (101). 19–27.
4. Radianti, J., et al. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*. 147. 103778.
5. Rasch, J., et al. (2023). Going, Going, Gone: Exploring Intention Communication for Multi-User Locomotion in Virtual Reality. *CHI Conference on Human Factors in Computing Systems*.
6. Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*. 77 (4). 541–580.
7. Zhang, K., Liu, G. (2024). TRustPN: Transforming Rust Source Code to Petri Nets for Checking Deadlocks. *10th International Conference on Control, Decision and Information Technologies (CoDIT)*.
8. Biggar, O., Zamani, M. (2020). A Framework for Formal Verification of Behavior Trees With Linear Temporal Logic. *IEEE Robotics and Automation Letters*.
9. Jensen, K., Kristensen, L. M. (2009). *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Springer.
10. Coffman, E. G., Elphick, M., Shoshani, A. (1971). System Deadlocks. *ACM Computing Surveys*. 3 (2). 67–78.
11. Wijethunga W. A. T. G. R., Kumara B. T. G. S. (2025). Systematic Review of Graph Neural Network and Consensus Algorithm-Based Approaches for Proactive Deadlock Detection. *Preprints*.
12. Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM journal on computing*. 1 (2). 146–160.
13. Shevchenko, S. (2025). A comprehensive approach and a role-based interaction model for designing cooperative scenarios in interactive training systems. *Electronic Modeling*. 47 (5). 75–86 [in Ukrainian].

14. Shevchenko, S. (2021). *Interactive automated remote educational and training system*. (Patent of Ukraine No. 149786). Ukrainian Institute of Intellectual Property [in Ukrainian].
15. Hemminki-Reijonen, U., Hassan, N. M. A. M., Huotilainen, M., Koivisto, J.-M., & Cowley, B. U. (2025). Design of generative AI-powered pedagogy for virtual reality environments in higher education. *npj Science of Learning*. 10 (1). 31.
16. Gao, L., et al. (2024). Constructing a Dialogue-Level Hallucination Evaluation Benchmark for Large Language Models. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.
17. Gadde, D. N., Kumar, A., Nalapat, T., Rezunov, E., & Cappellini, F. (2024). All Artificial, Less Intelligence: GenAI through the Lens of Formal Verification. *arXiv preprint arXiv:2403.16750*.
18. Samoilov, V., et al. (2020). Computer technologies for the development of training systems for the energy industry. *Electronic Modeling*. 42 (3). 89–98 [in Ukrainian].

Дата першого надходження статті до видання: 20.10.2025

Дата прийняття статті до друку після рецензування: 25.11.2025

Дата публікації (оприлюднення) статті: 26.12.2025

НАУКОВЕ ВИДАННЯ

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ
INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 2 (10) 2025

Підписано до друку: 26.12.2025.

Формат 60x84/8. Arial.

Папір офсет. Цифровий друк. Ум. друк. арк. 17,44. Замов. № 0226/208. Наклад 300 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»

65101, Україна, м. Одеса, вул. Інглєзі, 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.