

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)

Кафедра _____ Інформаційних управляючих систем _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

Дослідження методів управління та оцінки якості ІТ-проектів щодо
створення інтелектуальних медичних інформаційних систем
(тема)

Виконав:
здобувач _____ 2 _____ року навчання
групи _____ УПГІТм-24-1 _____

Сергій БОНДАРЕВ
(власне ім'я, ПРІЗВИЩЕ)

Спеціальність _____ 122 – Комп'ютерні науки _____
(повна назва освітньої програми)

Тип програми _____ освітньо-наукова _____

Освітня програма _____ Управління проектами в галузі
інформаційних технологій _____

Керівник _____ проф. каф. ІУС Віктор ЛЕВИКІН _____
(посада, власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри _____ ІУС _____


(підпис)

Костянтин ПЕТРОВ
(власне ім'я, ПРІЗВИЩЕ)

2026 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
Кафедра _____ Інформаційних управляючих систем _____
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 – Комп'ютерні науки _____
Тип програми _____ освітньо-наукова програма _____
Освітня програма Управління проектами в галузі інформаційних технологій
(повна назва)

ЗАТВЕРДЖУЮ: _____
Зав. кафедри _____
(підпис)
« 06 » квітня 2026 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студентові _____ Бондареву Сергію Геннадійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ «Дослідження методів управління та оцінки якості IT-проектів щодо створення інтелектуальних медичних інформаційних систем» _____

Затверджена наказом по університету від _____ 23 березня. 2026 р. № 231 Ст _____

2. Термін подання студентом роботи до екзаменаційної комісії _____ 20.05.2026 _____

3. Вихідні дані до роботи Приклади використання штучного інтелекту в медичній галузі, приклади IT-проектів по створенню медичних систем, фреймворки управління проектами, набір підходів розробки та управління MLOps, фреймворк NASSS, математичний апарат нечіткої логіки, бібліотеки нечіткої логіки scikit-fuzzy та fuzzy logic на мові програмування Python _____

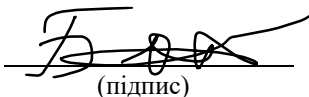
4. Перелік питань, що потрібно опрацювати в роботі Огляд практики впровадження технологій штучного інтелекту в сфері охорони здоров'я, аналіз особливостей управління та контролю якості медичних продуктів IT-проектів, формування вимог до методів оцінки якості, розробка теоретичної складової методів оцінки якості, апробація розроблених методів на даних досліджень та на результатах власноруч проведених опитувань _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	06.04.2026	виконано
2	Аналіз літератури щодо розробки медичних інформаційних систем	07.04.2026 – 12.04.2026	виконано
2	Розробка вимог до методів оцінки якості на основі попереднього аналізу літератури	13.04.2026 – 16.04.2026	виконано
3	Розробка теоретичної складової методів оцінки якості	17.04.2026 – 20.04.2026	виконано
4	Апробація розроблених методів	21.04.2026 – 02.05.2026	виконано
5	Підготовка пояснювальної записки	03.05.2026 – 07.05.2026	виконано
6	Підготовка презентації та доповіді	08.05.2026 – 14.05.2026	виконано
7	Перевірка на плагіат	15.05.2026	виконано
8	Норм. контроль	18.05.2026	виконано
9	Рецензування	19.05.2026 – 20.05.2026	виконано
10	Захист роботи	21.05.2026	виконано

Дата видачі завдання 06 квітня 2026 р.

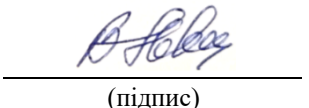
Здобувач (ка)


(підпис)

Сергій БОНДАРЕВ

(власне ім'я, ПРІЗВИЩЕ)

Керівник роботи


(підпис)

проф. каф. ІУС Віктор ЛЕВИКІН

(посада, власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Пояснювальна записка: 94 с., 3 рис., 6 табл., 4 дод., 70 джерел.

МАШИННЕ НАВЧАННЯ, МЕДИЧНІ ІНФОРМАЦІЙНІ СИСТЕМИ,
ОЦІНКА ЯКОСТІ, ОХОРОНА ЗДОРОВ'Я, УПРАВЛІННЯ ІТ-ПРОЄКТАМИ,
СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ.

Об'єктом дослідження є процес управління ІТ-проєктом, направленою на інтеграцію системи штучного інтелекту в роботу медичного закладу, через оцінку якості медичної інформаційної системи як програмного продукту.

Предметом дослідження виступає дослідження методів управління та оцінки якості проєктів по розробці систем штучного інтелекту в сфері охорони здоров'я, за рахунок розгляду та аналізу переваг і недоліків використання розповсюджених методологій в контексті виконання медичних проєктів.

Метою даного дослідження є всебічний аналіз управлінських підходів до розробки та впровадження інтелектуальних систем у сферу охорони здоров'я з акцентом на зміну підходів до управління якістю як самого програмного продукту, так і проєкту в цілому. Особливу увагу приділено визначенню ключових факторів успішної реалізації та основних перешкод, що супроводжують інтеграцію штучного інтелекту в медичні інформаційні системи. Результатом, роботи є специфікація вимог до методів оцінки якості медичних інформаційних систем та розробка таких методів.

Методи дослідження включають в себе аналіз наукової та науково-популярної літератури, інтернет джерел, технічної документації, відеоматеріалів, вибір необхідних програмних середовищ для вирішення та візуалізації практичної сторони планування.

ABSTRACT

Masters's thesis contains: 94 pp., 3 fig., 6 tabl., 4 ann., 70 references.

ARTIFICIAL INTELLIGENCE SYSTEMS, HEALTHCARE, IT PROJECT
MANAGMENT, MACHINE LEARNING, MEDICAL INFORMATION
SYSTEMS, QUALITY ASSESSMENT.

The object of the study is the project management process in the field of information technology, with a focus on software product quality control, directed toward the development and integration of artificial intelligence systems into the operations of medical institutions.

The subject of the research is the development of a methodology for assessing quality of the projects (which are focused on the integration of artificial intelligence systems in healthcare) by analyzing the advantages and disadvantages of widely used methodologies within the context of medical project execution.

The purpose of this study is a comprehensive analysis of management approaches to the development and implementation of intelligent systems in healthcare, with an emphasis on transforming quality management strategies for both the software product and the project. Special attention is given to identifying key success factors and the primary barriers associated with the integration of AI into medical information systems. The results of the work include a specification of requirements for quality assessment methods and the subsequent development and practical demonstration of such methods.

Research methods include the analysis of scientific and popular-scientific literature, online resources, technical documentation, and video materials, as well as the selection of appropriate software environments for addressing and visualizing the practical aspects of project planning.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	9
Вступ.....	10
1 Аналіз предметної галузі та постановка задачі.....	12
1.1 Аналіз ІТ-проектів штучного інтелекту.....	12
1.1.1 Характеристика ІТ-проектів штучного інтелекту.....	12
1.1.2 Аналіз особливостей управління проектами по створенню інтелектуальних інформаційних систем.....	13
1.2 Аналіз ІТ-проектів в сфері охорони здоров'я	16
1.2.1 Характеристика ІТ-проектів в сфері охорони здоров'я	16
1.2.2 Аналіз літератури щодо управління ІТ-проектами в медицині	18
1.2.3 Технічні особливості імплементації штучного інтелекту в медичних інформаційних системах	19
1.2.4 Характеристика проектів штучного інтелекту в сфері охорони здоров'я	21
1.3 Актуальність і постановка задачі дослідження.....	23
1.3.1 Актуальність роботи	23
1.3.2 Мета роботи та постановка задачі.....	24
2 Аналіз методів оцінювання ІТ-проектів штучного інтелекту в медицині ...	27
2.1 Методи оцінки якості ІТ-проектів штучного інтелекту.....	27
2.1.1 Метрики оцінки моделей штучного інтелекту.....	27
2.1.1.1 Формальні показники оцінки точності роботи штучного інтелекту.....	27
2.1.1.2 Вимоги до вхідних даних моделей штучного інтелекту.....	30
2.1.2 Методи управління та розробки систем штучного інтелекту MLOps	31
2.1.3 Оцінка сприйняття медичних інформаційних систем з боку користувачів.....	32

2.2	Управління якістю IT-проектів щодо розробки медичних інформаційних систем	33
2.2.1	Основні методи управління якістю IT-проектів в медичній галузі	33
2.2.2	Проблеми залучення користувачів до роботи над IT-проектами в сфері охорони здоров'я.....	34
2.2.3	Фреймворк NASSS як інструмент попереднього аналізу перспектив виконання IT-проектів в сфері охорони здоров'я.....	35
2.3	Напрями розвитку методів оцінки якості медичних інформаційних систем	36
2.4	Вимоги до показників оцінки якості медичних інформаційних систем штучного інтелекту	38
2.4.1	Вимоги до інтегрального показника точності моделі штучного інтелекту.....	38
2.4.2	Вимоги до показника оцінки якості медичної інформаційної системи користувачами	39
3	Опис запропонованих методів оцінки якості медичних інформаційних систем	41
3.1	Розроблений метод інтегральної оцінки технічної якості моделі штучного інтелекту	41
3.2	Розроблений метод оцінки медичної інформаційної системи користувачами	44
4	Апробація розроблених методів оцінки якості медичних інформаційних систем	48
4.1	Приклад використання інтегрального показника точності моделі штучного інтелекту	48
4.2	Приклад використання індексу сприйняття медичної інформаційної системи системи користувачами	50
4.2.1	Характеристика проведеного опитування користувачів медичних інформаційних систем	50
4.2.2	Характеристика респондентів проведеного опитування	54

4.2.3 Розподіл вагових коефіцієнтів та значень функцій приналежності	55
4.2.4 Розрахунок індексу оцінки медичної інформаційної системи користувачами	57
Висновки	60
Перелік джерел посилання	62
Додаток А Питання форми опитування	71
Додаток Б Графіки функцій приналежності	77
Додаток В Результати опитування	84
Додаток Г Графічний матеріал кваліфікаційної роботи	85

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ГНМ – глибокі нейронні мережі

ІС – інтелектуальна інформаційна система

ІС – інформаційна система

МІС – медична інформаційна система

МН – машинне навчання

МРТ – магнітно-резонансна томографія

СШІ – система штучного інтелекту

ШІ – штучний інтелект

CAI – Composite Accuracy Index – інтегральний показник точності моделі штучного інтелекту

CNN – Convolutional neural network – згорткова нейронна мережа

FHIR – Fast Healthcare Interoperability Resources – протокол обміну медичними даними

HL7 – Health Level 7 – стандарт обміну, управління та інтеграції електронної медичної інформації

USPI – User System Perception Index – індекс сприйняття системи користувачами

XAI – Explainable Artificial Intelligence – пояснювальний штучний інтелект

ВСТУП

Сучасний етап розвитку глобальної цифрової екосистеми характеризується глибокою трансформацією підходів до управління ІТ-проєктами, де на зміну стандартній автоматизації бізнес-процесів приходить створення складних інтелектуальних інформаційних систем (ІС). Сьогоднішні технологічні ініціативи стають дедалі масштабнішими, вимагаючи не лише високої швидкості розробки, а й здатності програмних продуктів адаптуватися до мінливих умов експлуатації. У цьому контексті особливого значення набувають проєкти, що фокусуються саме на розробці систем штучного інтелекту як самостійних інтелектуальних одиниць або модулів вже існуючих інформаційних систем (ІС). Такі проєкти кардинально відрізняються від традиційної розробки програмного забезпечення, оскільки вони базуються на безперервній роботі з великими масивами даних та забезпеченні здатності моделей до генерації нових знань, що висуває унікальні вимоги до методів управління та забезпечення надійності результатів.

Одним із перспективних та водночас відповідальних векторів розвитку є створення інтелектуальних систем для медичної галузі. В сучасному світі технології штучного інтелекту з кожним роком все активніше інтегруються у процеси сфери охорони здоров'я, перетворюючись із експериментальних розробок на повсякденні інструменти підтримки лікарської діяльності. Методи машинного навчання (МН) знаходять широке застосування фактично на всіх етапах взаємодії з пацієнтом, починаючи від первинного діагностування захворювань і закінчуючи вибором оптимального курсу лікування чи прогнозуванням можливих ризиків для здоров'я у довгостроковій перспективі. Така багатофункціональність стає можливою завдяки здатності алгоритмів оперувати різнорідними типами даних, що включають структуровані цифрові медичні картки, складні візуальні образи рентгенографії, магнітно-резонансної томографії (МРТ), кардіограм, а також

текстові описи симптомів і скарг пацієнтів. Останнім часом до цього додається аналіз відеозаписів хірургічних втручань та використання елементів доповненої реальності, що виводить точність медичних маніпуляцій на принципово новий рівень.

Специфіка проєктів зі створення медичних інформаційних систем полягає у необхідності поєднання технологічних інновацій із суворими вимогами до безпеки, оскільки помилка в таких системах має високу ціну. Саме тому ключовою науковою та практичною проблемою залишається вдосконалення методик оцінки якості таких проєктів та кінцевих програмних продуктів. Традиційні підходи до тестування часто виявляються недостатніми для оцінки інтелектуальних систем через їхню динамічну природу та складність логічних зв'язків. Сучасні методи оцінки якості зміщує акцент у бік комплексного аналізу функціональної придатності та надійності рішень у реальних клінічних умовах. Важливу роль відіграє розробка критеріїв, які б дозволяли системно оцінити ефективність інтелектуального продукту, враховуючи специфіку обробки конфіденційних даних та потребу у стабільності результатів незалежно від обсягу вхідної інформації. Формування прозорих підходів до валідації таких систем є стратегічним завданням, що дозволить забезпечити довіру професійної спільноти та успішне масштабування штучного інтелекту в системі охорони здоров'я.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз ІТ-проєктів штучного інтелекту

1.1.1 Характеристика ІТ-проєктів штучного інтелекту

Проєкти зі створення систем штучного інтелекту суттєво відрізняються від традиційної розробки програмного забезпечення в першу чергу через свою внутрішню недетермінованість та впливову частку дослідницького ризику. На відміну від розробки стандартних інформаційних систем, де функціональність можна заздалегідь специфікувати та протестувати на відповідність логічним вимогам, створення інтелектуальних модулів представляє собою процес ітеративної розробки «чорної скриньки». У науковій літературі такі проєкти часто класифікують як гібрид інженерії та експериментальної науки, де кожна ітерація розробки є фактично перевіркою наукової гіпотези про здатність моделі до апроксимації певної залежності. Таким чином на початку проєкта, а іноді і в середині циклу виконання, ніхто не може гарантувати, що інтелектуальну модель необхідної якості в принципі можливо створити.

Ключовою характеристикою процесу створення систем штучного інтелекту є «проблема холодного старту» та критична залежність від середовища даних. Виконання технічної сторони проєкта починається не з написання функціоналу, а з формування репрезентативного набору даних, що в контексті складних систем часто вимагає залучення вузькопрофільних експертів для розмітки та верифікації [1].

Дослідники з Microsoft Research вказують на те, що робота над подібними продуктами вимагає набагато глибшої інтеграції між фахівцями з обробки даних та інженерами програмного забезпечення, оскільки будь-яка зміна в інфраструктурі збору даних може призвести до катастрофічного падіння показників якості всієї системи [2]. Це породжує особливий тип ризику, який називають ризиком моделі: навіть при ідеально написаному коді

система може бути нефункціональною через невідповідність математичного апарату реальним умовам експлуатації.

Іншою важливою особливістю створення СШІ (системи штучного інтелекту) є складність визначення критеріїв «завершеності» проєкту. У класичному проєктному менеджменті виконання всіх пунктів технічного завдання означає успіх, проте в даному випадку досягнення цільової метрики (наприклад, точності розпізнавання) на тестових даних не гарантує коректного функціонування в реальній роботі [3]. В якості відповіді, розробники змушені впроваджувати складні багаторівневі методи оцінювання, що включають тестування на стійкість (robustness), перевірку на відсутність зміщень (bias testing) та оцінку інтерпретованості результатів. Отже, проєкт зі створення СШІ розглядається не як лінійна розробка продукту, а як створення складного продукту, що потребує безперервної валідації/налаштування/доброби або навіть кардинальних змін протягом усього життєвого циклу.

Невдачі в таких проєктах найчастіше пов'язані з розривом між математичною точністю моделі та прикладною цінністю, яка очікується на виході. Часто системи, що демонструють високі показники в лабораторних умовах, виявляються непридатними для використання через надмірну складність інтеграції або неможливість пояснити логіку прийняття рішень кінцевому користувачу. Саме тому сучасні методи оцінки таких проєктів дедалі більше схиляються до системного підходу, де якість моделі є лише одним із багатьох факторів успіху.

1.1.2 Аналіз особливостей управління проєктами по створенню інтелектуальних інформаційних систем

У сучасній науковій літературі, ІТ-проєкти зі створення систем штучного інтелекту розглядаються як окрема категорія інноваційних проєктів,

що характеризуються високим рівнем невизначеності, ітеративністю процесів розроблення та значною залежністю від якості даних. На відміну від класичних інформаційних систем, де функціональні вимоги можуть бути детально формалізовані на початкових етапах, у проєктах зі створення ІС вимоги часто уточнюються поступово, за результатами експериментальних перевірок моделей та оцінювання продуктивності. З цієї причини управління такими проєктами потребує адаптації традиційних підходів до специфіки машинного навчання, включаючи етапи збирання, підготовки й валідації даних, навчання моделей та подальшого моніторингу в наближеному до реального стану роботи середовищі.

З позиції класичного процесного підходу, закріпленого у стандартах PMBOK (A Guide to the Project Management Body of Knowledge), проєкти зі ІІІ можуть бути структуровані відповідно до груп процесів ініціації, планування, виконання, моніторингу й контролю та завершення [4]. Водночас зміст окремих процесів набуває специфічного характеру: управління обсягом проєкту ускладнюється за рахунок формалізації очікуваної точності моделі, управління ризиками включає технологічні та етичні аспекти (які будуть описані згодом), а до управління якістю додається система метрик ефективності алгоритмів. Таким чином, класичні області знань проєктного менеджменту залишаються релевантними, однак застосування вимагає розширеного трактування та інтеграції з практиками data science.

Каскадні (водоспадні) моделі розробки демонструють обмежену ефективність у чистому вигляді для проєктів такого роду, через необхідність багаторазових експериментів і повернення до попередніх етапів. Натомість гнучкі підходи, представлені у Agile Manifesto [5] та подальших фреймворках реалізації та доповнення базових принципів (Scrum, Kanban, Extreme programming та інші), забезпечують кращу адаптивність до змін вимог і результатів проміжних ітерацій [6], [7]. У таких проєктах спринти можуть бути спрямовані не лише на реалізацію функціоналу, а й на дослідження гіпотез, підбір архітектури моделі або оптимізацію параметрів. Водночас застосування

Agile потребує чіткого визначення критеріїв готовності інкременту, що в контексті ШІ часто пов'язано з досягненням заданого порогу якості моделі, а не лише з реалізацією програмного коду.

Структуровані методології аналізу та проєктування інформаційних систем, такі як SSADM (Structured Systems Analysis and Design Method) [8], можуть бути використані на етапах формалізації вимог і моделювання бізнес-процесів, особливо в проєктах інтеграції інтелектуальних програмних рішень у вже існуючі корпоративні системи. Однак жорстка послідовність фаз часто комбінується з ітеративними підходами до розроблення моделей, що формує скоріше гібридні моделі управління. У наукових дослідженнях підкреслюється, що саме виправдане комбінування методологій, тобто поєднання формалізованого управління ризиками з гнучкою організацією дослідницької роботи є найбільш доцільною практикою для проєктів щодо створення СШІ незалежно від галузі застосування.

Окремої уваги заслуговує методологія PRINCE2 [9], яка розглядає управління проєктами через призму чітко визначених принципів, тем і процесів, орієнтованих на бізнес-обґрунтування та керованість. У контексті побудови ІС особливо релевантним є принцип безперервного бізнес-обґрунтування, оскільки доцільність інтелектуального продукту часто залежить від досягнення прогнозованих показників ефективності та економічної віддачі [10]. Поетапне управління дозволяє структурувати проєкт на керовані фази, в межах яких можуть проводитися експериментальні ітерації з розроблення та тестування моделей. Водночас механізм управління за винятками сприяє своєчасному реагуванню на відхилення показників якості моделі або перевищення ресурсних обмежень. Таким чином, PRINCE2 може бути адаптована до проєктів зі ШІ шляхом поєднання формалізованої системи контролю з гнучкими підходами до реалізації технічних завдань, забезпечуючи баланс між інноваційністю та керованістю проєкту. Зокрема, інтеграція теми «Якість» (Quality theme), що є фундаментальною для цієї методології, передбачає створення деталізованих описів продуктів (Product

Descriptions) із чіткими критеріями приймання ще до початку їхньої розробки. Для медичних інтелектуальних систем це означає необхідність попередньої фіксації цільових порогових значень метрик оцінок якості ШІ, вимог до алгоритмічної безпеки та рівня інтерпретабельності рішень штучного інтелекту. Такий формалізований підхід дозволяє менеджеру проєкту не лише контролювати математичну валідність моделей на кожному етапі, а й гарантувати їхню клінічну відповідність суворим регуляторним стандартам охорони здоров'я, мінімізуючи ризик відторгнення системи кінцевими користувачами.

1.2 Аналіз ІТ-проєктів в сфері охорони здоров'я

1.2.1 Характеристика ІТ-проєктів в сфері охорони здоров'я

Цифровізація охорони здоров'я пройшла шлях від простої автоматизації документообігу до створення складних екосистем підтримки клінічних рішень. Проте, аналіз літератури показує, що медичні ІТ-проєкти мають значно вищий поріг входу та ризик неприйняття користувачами, ніж в інших комерційних сферах діяльності.

В якості головної особливості медичних ІТ-проєктів можна відділити дуалістичну природу вимог: мають одночасно задовольнятися адміністративно-господарські потреби та клінічні вимоги. Такі обставини змушують пропонувати особливі, галузеві фреймворки розробки, як наприклад автори статті [11] у своїх дослідженнях про впровадження технологій в медицині пропонують фреймворк NASSS (детальніше в п. 2.2.3), який пояснює, чому такі проєкти часто зазнають невдачі і надає потенційні рішення. Згідно з цією моделлю, складність медицини полягає не лише у складності програмного забезпечення, а й у непередбачуваності стану

пацієнта, застарілих методах роботи персоналу та складності міжвідомчої взаємодії.

До важливих технічних особливостей можна віднести фрагментарність даних. Історично медичні установи використовували (та продовжують використовувати) закриті локальні системи, що призвело до створення «інформаційних силосів». Сучасні проєкти змушені вирішувати проблему сумісності через впровадження міжнародних стандартів, таких як HL7 (Health Level 7) [12] та FHIR (Fast Healthcare Interoperability Resources) [13]. Проте, навіть за наявності стандартів, інтеграція нової системи в існуючу інфраструктуру лікарні часто займає більше часу, ніж розробка самого програмного продукту. Як результат спрацьовує закономірність «інтеграційного парадоксу»: чим складніша система, тим менша ймовірність «безшовного» поєднання з функціоналом інших модулів інформаційної системи медичного закладу. Важливо відмітити, що медичні IT-проєкти перебувають під жорстким контролем законодавства щодо захисту персональних даних (GDPR у ЄС, HIPAA у США, Закон України «Про захист персональних даних») та ряду інших [14], [15], [16], що накладає на проєкт додатковий рівень витрат: архітектура системи має передбачати складні механізми шифрування, аудиту доступу та анонімізації даних (актуально для розробок моделей ІІІ). Будь-яка вразливість у системі може призвести не лише до фінансових втрат, а й до юридичної відповідальності розробника за витік чутливої медичної інформації.

Специфічною рисою галузі є надзвичайна завантаженість медичного персоналу. Дослідження [17] показують, що погано спроектовані інтерфейси медичних систем стають причиною «професійного вигорання» лікарів. Проєкти часто провалюються на етапі впровадження, оскільки вони додають лікарям бюрократичного навантаження замість того, щоб його зменшувати. Можна стверджувати, що успішні медичні IT-проєкти характеризуються в тому числі й підходом human-centered design [18], в рамках якого медики залучаються до проектування інтерфейсів на ранніх стадіях.

1.2.2 Аналіз літератури щодо управління ІТ-проектами в медицині

Автор роботи [19] зосереджує увагу на гібридних методологіях управління проектами в галузі охорони здоров'я, в тому числі на перспективам активного залучення стейкхолдерів при виконанні, а також налагодженню комунікації з останніми. В продовження тематики управління людськими ресурсами, можна згадати дослідження [20], що розкриває методи організації праці в розподілених командах (результатом появи яких може бути активне залучення стейкхолдерів до роботи), та управління змінами в такому контексті.

Робота [21] частково висвітлює проблеми використання методу Kanban в медичній галузі, а також запропоновано ряд модифікацій, які на конкретному прикладі показали деякі позитивні результати, однак за рахунок локальності експерименту, користь від нововведень є сумнівною і потребує більшої доказової бази.

Незважаючи, що дослідження [22] не має прямого відношення до медицини, однак в роботі пропонується вирішення спільних до різних галузей проблем в управлінні за рахунок використання методів машинного навчання, при цьому вибір самих алгоритмів, таких як карти Конохена та генетичних рядів є доволі оригінальним. Отримані результати можуть бути розглянуті в контексті нашої роботи.

Автор [23] фокусується на методах оцінки ефективності проєктів направлених на автоматизацію роботи медичних лабораторій, що є важливим аспектом аналізу потенціалу проєкта з точки зору основних та другорядних економічних витрат.

Робота [24] у свою чергу перелічує деякі особливості сфери охорони здоров'я з саме точки зору управління проектами, однак враховуючи формат, основні етапи та можливі рішення були лише коротко окреслені і потребують більш глибокого аналізу.

1.2.3 Технічні особливості імплементації штучного інтелекту в медичних інформаційних системах

На сучасному етапі технологічного розвитку штучний інтелект успішно вирішує завдання з ранньої діагностики онкологічних та кардіологічних захворювань, прогнозує ризики розвитку ускладнень на основі аналізу життєвих показників пацієнта у реальному часі, допомагає у розробці персоналізованих стратегій лікування, враховуючи індивідуальні особливості організму, використовується механізми доповненої реальності для допомоги при підготовці та проведенні хірургічних втручань та багато іншого. Одночасно з цим галузь охорони здоров'я характеризується низкою унікальних особливостей, які зумовлюють підвищену складність розробки інтелектуальних систем порівняно з іншими сферами [25]. Ключовим бар'єром є проблема доступу до даних. Медична інформація часто зберігається у різноманітних форматах та розподілена між локальними базами даних окремих установ, що перешкоджає створенню централізованих репозиторіїв для навчання моделей ШІ, які стануть основою відповідних систем. Навіть у межах однієї державної інфраструктури може спостерігатись брак стандартизації. Крім того, жорсткі вимоги до конфіденційності потребують складних процедур анонімізації даних при навчанні інтелектуальної моделі [26].

Суттєвою перешкодою є також незбалансованість та малий обсяг вибірок. У медичній статистиці закономірно переважають дані здорових пацієнтів. Для машинного навчання це створює ризик зниження прогнозної точності, оскільки модель отримує недостатньо прикладів патологічних станів, тобто позитивних результатів класифікації (при розв'язанні відповідного типу задач машинного навчання). При цьому наявні набори даних часто не відповідають вимогам глибинного навчання, яке для

ефективної генералізації потребує сотень тисяч або мільйонів записів в залежності від класу задач.

Додатковим дестабілізуючим фактором є варіативність якості даних. Нейронні мережі, оптимізовані для аналізу результатів МРТ-діагностики в одній країні, можуть демонструвати низьку ефективність в іншій через відмінності в технічних характеристиках медичного обладнання та специфіці отриманих зображень, тож використання вже створених технологій може бути тією чи іншою мірою обмежене.

Різноманітність діагностичних методів зумовлює використання широкого спектра алгоритмів машинного навчання. Для опрацювання неструктурованих даних (рентгенограм, томограм і т.д.) домінуючим підходом є використання глибоких нейронних мереж (ГНМ). В якості прикладів можна навести наступні програмні продукти: Google DeepMind Health застосовує ГНМ для аналізу ретинових та церебральних знімків з метою раннього виявлення патологій [27]; EchoPixel використовує технології доповненої реальності для візуалізації інтерактивних тривимірних анатомічних моделей у реальному часі під час хірургічних втручань [28] та інші [29]. Технологічний базис таких систем найчастіше складають згорткові нейронні мережі (CNN), що виконують завдання сегментації та детекції об'єктів. Репрезентативним прикладом є архітектура U-Net, яка стала своєрідним стандартом аналізу медичних зображень [30], [31].

Для аналізу структурованих (табличних) даних також залучаються ГНМ, як-от у проєкті Deep Genomics, де вони використовуються для прогнозування впливу генетичних мутацій на розвиток рідкісних хвороб та формування персоналізованої терапії [32]. Проте в цьому сегменті ГНМ не є безальтернативними. Часто застосовуються класичні методи: логістична регресія, k-найближчих сусідів (k-NN), дерева рішень, метод опорних векторів, а також ансамблі моделей [33], [34]. Вибір на користь класичного МН часто зумовлений обмеженістю обчислювальних ресурсів або малим обсягом навчальної вибірки.

Впровадження СШІ в критичні сфери, зокрема в медицину, актуалізує проблему інтерпретованості рішень особливо при використанні згаданих вище складних алгоритмів глибоких нейронних мереж. Оскільки помилка моделі може мати критичні наслідки, виникає потреба не лише у високій точності, а й в обґрунтованості прогнозів. Таке завдання частково можна вирішити завдяки методам пояснювального штучного інтелекту (Explainable AI, XAI).

Пояснювальний штучний інтелект – це сукупність методів і алгоритмів, що дозволяють людині зрозуміти логіку формування вихідного результату на основі вхідних параметрів, що може бути корисним як розробникам при тестуванні, так і кінцевим користувачам. На практиці XAI виглядає як набір візуалізацій (графіки, додаткові позначки на вхідних зображеннях, тощо) або текстові зауваження, що описують «мотивацію» прийняття системою конкретного рішення.

1.2.4 Характеристика проєктів штучного інтелекту в сфері охорони здоров'я

Розвиток інтелектуальних інформаційних систем (ІС) у медицині для підтримки прийняття клінічних рішень, автоматизації діагностики та прогнозування ризиків супроводжується зростаючою увагою як у науковій спільноті [35], так і в суспільному дискурсі [36], [37], [38]. Проте перехід від теоретичних розробок до практичної імплементації виявляє низку специфічних бар'єрів (деякі вже згадані особливості зберігаються, однак з'являються і нові на рівні проєкта, а не продукту), які потребують адаптації класичних методологій управління проєктами. Систематизація цих проблем дозволяє виділити три ключові вектори, що продовжують та розширюють особливості, які описані в попередньому розділі:

- технічні складнощі підготовки та реалізації;

- відповідність регуляторним та етичним нормам;
- організаційні складнощі.

Окрім вже описаних технічних особливостей, постає питання релевантності метрик оцінювання, адже традиційні для машинного навчання показники, такі як точність або f1-score, не завжди здатні адекватно відобразити реальну клінічну ефективність системи. Це зумовлює необхідність впровадження багатофакторних процедур валідації, що враховують специфічні медичні сценарії та вимагають залучення профільних експертів для аналізу логіки прийняття рішень (більш детально розглянуто в наступному розділі).

На відміну від систем загального призначення, медичні ІС безпосередньо впливають на життя людей, а такий факт диктує потребу в особливому правовому регулюванні. Наразі спостерігається певний правовий вакуум та брак стандартизації, оскільки законодавство більшості країн лише адаптується до стрімкого технологічного прогресу. Важливим вектором у цьому напрямі для України стало підписання міжнародної рамкової конвенції з регулювання ІІ, що створює підґрунтя для майбутньої сертифікації медичного програмного забезпечення та гармонізації норм [39].

Окрім суто юридичних аспектів, надважливою є етична імперативність розробки. Фахівці мають системно нівелювати ризики алгоритмічних упереджень (bias), які можуть виникати через незбалансованість навчальних вибірок і призводити до дискримінації певних груп пацієнтів. Питання моральної та юридичної відповідальності за помилковий діагноз, сформований за участі інтелектуальних систем, залишається однією з найбільш дискусійних тем у сучасній науковій спільноті, вимагаючи чіткого розподілу ролей між алгоритмом та клініцистом [40].

Специфіка медичних ІС змушує переглянути консервативні управлінські підходи, виявляючи низку організаційних бар'єрів. Першочерговим є міждисциплінарний бар'єр: формування команд, куди одночасно входять лікарі, розробник ІІ/data science спеціаліст, юристи та

адміністратори, часто супроводжується термінологічними та концептуальними конфліктами. Відсутність єдиного семантичного поля між технічними фахівцями та медичним персоналом уповільнює комунікацію та ускладнює прийняття стратегічних рішень.

Ситуація ускладнюється високим рівнем невизначеності при плануванні. Проєкти з впровадження ІІІ за своєю природою є нелінійними, тому традиційні каскадні моделі управління виявляються малоефективними. Труднощі з прогнозуванням ресурсів на ітеративне донавчання та валідацію моделей змушують команди обирати гнучкі методології, що, своєю чергою, підвищує фінансові та часові ризики.

До того ж, дефіцит історичних аналогів та унікальність кожного конкретного впровадження обмежують можливості використання стандартних методів оцінки ризиків, роблячи кожен такий проєкт певною мірою експериментальним на етапі його ініціації.

1.3 Актуальність і постановка задачі дослідження

1.3.1 Актуальність роботи

Згідно з результатами досліджень CHAOS Report, що кожного року проводяться командою The Standish Group [41], менше третини ІТ-проєктів завершуються успіхом, інші або завершуються з частковим виконання початкових вимог та перевершенням витрат, або закриваються. У випадку проєктів по створенню інтелектуальних систем відсоток невдач ще вищий і відповідно до результатів ряду досліджень досягає 80% [42], [43], [44]. Важливо підкреслити, що такі низькі показники стосуються не лише невеликих проєктів/стартапів, що виконуються невеликими компаніями або командами ентузіастів (в такій групі проєктів низькі показники успішності є очікуваним явищем), а і масштабні ініціативи на державному або рівні

транснаціональних корпорацій. Одним із таких прикладів є британська програма цифровізації NHS National Programme for IT, помилки в управлінні та некоректному аналізі потреб користувачів призвели до втрат понад 10 мільярдів фунтів стерлінгів. Даний випадок продемонстрував, що в медицині неможливо оцінювати якість програмного продукту у відриві від реальних клінічних процесів [45].

Схожа ситуація трапилась з IBM Watson for Oncology, в рамках проєкта створення інтелектуального асистента для діагностики онкологічних захворювань завершилося невдачею через низьку якість навчальних даних та невідповідність алгоритмів реальним умовам лікарень. Система, на яку було витрачено десятки мільйонів доларів, часто видавала небезпечні рекомендації, оскільки була навчена на ідеалізованих сценаріях, а не на даних з реального середовища, для якого була розроблена [46].

Таким чином створення систем штучного інтелекту для медицини сьогодні перебуває на перетині трьох зон високого ризику: загальної нестабільності IT-проєктів, дослідницької невизначеності алгоритмів ШІ та жорстких вимог до безпеки в охороні здоров'я. Аналіз особливостей управління такими проєктами стає життєво необхідним для трансформації експериментальних моделей у надійні медичні інструменти. У свою чергу управління та контроль якості програмного продукту та проєкту в цілому є важливим аспектом досягнення успіху в даній сфері діяльності, відповідно можна стверджувати, що розробка нових підходів та корегування вже існуючих методів управління якістю є повністю виправданим.

1.3.2 Мета роботи та постановка задачі

Враховуючи вище описані факти та характеристики предметної галузі, що включають в себе технічні та управлінські особливості реалізації проєктів

по створенню та інтеграції систем штучного інтелекту в сфері охорони здоров'я, можна сформулювати мету даного дослідження.

Метою даного дослідження є аналіз методів та підходів до розробки та впровадження інтелектуальних систем у сфері охорони здоров'я з акцентом на зміну підходів до управління якістю як самого програмного продукту, так і проєкту в цілому. Особливу увагу приділено визначенню ключових факторів успішної реалізації та основних перешкод, що супроводжують інтеграцію штучного інтелекту в медичні інформаційні системи.

Основним науковим результатом дослідження має стати розробка та теоретичне обґрунтування запропонованих методів оцінки якості медичних інформаційних систем, що здатні нівелювати виявлені ризики та забезпечити високу надійність розробки програмних продуктів в даній галузі. У межах роботи передбачається не лише пропозиція конкретних інструментів оцінювання, а й практична апробація на прикладі реальних або модельних сценаріїв розробки медичних інтелектуальних систем, що дозволить підтвердити ефективність запропонованих підходів, продемонструвати здатність до зниження ризиків проєктних невдач та надати інструмент оцінки якості програмних продуктів для менеджерів ІТ-проєктів у високотехнологічному секторі медицини.

Враховуючи мету роботи, можна викласти постановку задачі, що складається з наступних пунктів:

- систематичний аналіз наукових та галузевих джерел щодо технічних та управлінських особливостей проєктів зі створення систем штучного інтелекту в загальному ІТ-контексті;
- дослідження специфіки реалізації ІТ-проєктів щодо розробок медичних інформаційних систем;
- аналіз практичного досвіду виконання проєктів з впровадження інтелектуальних систем для виявлення закономірностей успіху та критичних чинників, що призводять до проєктних невдач;

- визначення особливостей процесів оцінки якості в проєктах із штучним інтелектом в медичній галузі та взагалі;
- формулювання комплексного переліку вимог до нових методів оцінки якості медичних програмних продуктів;
- розробка пропозицій щодо методів оцінювання якості медичних інформаційних систем, згідно з вказаними вимогами;
- практична апробація розроблених методів оцінювання якості медичних інформаційних систем;
- ретроспектива отриманих результатів, з метою формування пропозицій щодо напрямку подальшого розвитку методів.

2 АНАЛІЗ МЕТОДІВ ОЦІНЮВАННЯ ІТ-ПРОЄКТІВ ШТУЧНОГО ІНТЕЛЕКТУ В МЕДИЦИНІ

2.1 Методи оцінки якості ІТ-проєктів штучного інтелекту

2.1.1 Метрики оцінки моделей штучного інтелекту

2.1.1.1 Формальні показники оцінки точності роботи штучного інтелекту

Для кращого розуміння методів оцінки якості моделей штучного інтелекту необхідно почати з фундаментальних метрик класифікації. Оскільки більшість медичних проєктів зі ШІ спрямовані на виявлення патологій, наприклад у вигляді бінарна класифікація: «хворий» або «здоровий», основою для оцінювання в таких випадках є матриця помилок (confusion matrix), яка дозволяє розкласти результати роботи моделі на чотири компоненти: насправді позитивні (true positive, TP), насправді негативні (true negative, TN), хибно позитивні (false positive, FP) та хибно негативні (false negative, FN) результати. На основі вказаних випадків нижче розглянуто основні показники точності.

Найбільш розповсюдженим методом є стандартна точність (accuracy), що визначає частку правильних прогнозів (як позитивних, так і негативних) серед усіх представлених випадків. Формула 2.1 має наступний вигляд.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

Незважаючи на інтуїтивну простоту такої метрики, отримані результати можуть бути оманливими, в першу чергу через так званий «парадокс точності». Як можна бачити з формули 2.1, наприклад, якщо серед 100 пацієнтів є лише 5 хворих, система, яка відповідає на будь-які вхідні дані негативним результатом (здоровий пацієнт) буде мати точність 95%,

незважаючи на те, що працює абсолютно некоректно. Для вирішення такої проблеми підходять метрики, які описані далі.

Метрика *precision* (формула 2.2) відображає здатність моделі ШІ не класифікувати здорового пацієнта як хворого, таким чином визначаючи частку справжніх діагнозів серед усіх випадків, які система позначила як позитивні.

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

Показник *recall* або *sensitivity* (формула 2.3) у свою чергу демонструє частку реальних випадків захворювання система змогла виявити. Якщо низький показник *precision* на практиці призводить до перевантаження медичних ресурсів, через велику кількість хибних тривог, то *sensitivity* відповідає за непомічені захворювання (FN).

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.3)$$

Як доповнення до попередньої метрики, існує *specificity* (формула 2.4), що визначає частку істинно негативних результатів, тобто здатність моделі правильно ідентифікувати відсутність патології (FP).

$$Specificity = \frac{TN}{TN + FP} \quad (2.4)$$

Враховуючи метрики *sensitivity* та *precision*, можна використати показник *f1-score* (F1), який є своєрідним вдосконаленням класичної точності для коректного відображення при незбалансованих вибірках даних. Описана формула 2.5 наведена нижче.

$$F1 = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \quad (2.5)$$

Іншим вирішенням проблеми аналізу незбалансованих наборів даних при бінарній класифікації є Matthews correlation coefficient (MCC), що приймає значення від -1 до 1 (Формула 2.6). Метрика дозволяє більш детально дослідити природу помилок за рахунок, за рахунок використання всіх даних з матриці помилок [47].

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (2.6)$$

Вказані метрики можуть використовуватись в комбінації з підходом перехресної валідації (cross-validation), якщо не вдаватись в деталі, це підхід, при якому тестування моделі проводять декілька разів на різних наборах даних та акумулюють середній результат. З практичними аспектами імплементації різних видів перехресної перевірки можна ознайомитись в роботі [48].

У випадках бінарної класифікації більш-менш універсальним показником є площа під кривою (area under the ROC-curve, AUC-ROC), що використовує залежність між рівнем істинно позитивних та хибно позитивних випадків класифікації у вигляді кривої на графіку. Значення зазвичай нормують від 0.5 до 1 (від випадкового вгадування до абсолютної точності).

При прогнозування ризику розвитку тих чи інших захворювань або результативності використання лікарських засобів за допомогою моделей штучного інтелекту можна використовувати метрику оцінки concordance індекс (C-індекс) [49], [50]. Однак задачі такого прогнозування, а відповідно і використання індексу потребує складного типу даних (наприклад потрібна не лише наявність чи відсутність у пацієнта захворювання, а і ймовірність відсутності/наявності), тому детальніше зупинятись в роботі ми не будемо. Особливості використання C-індексу описані в роботі [51].

Список перелічених метрик не є повним. В залежності від задачі машинного навчання, методи оцінки можуть додаватись або змінюватись. При пошуку об'єктів на зображеннях (object detection) існують специфічні для такої задачі метрики, аналогічна ситуація з обробкою тексту чи відео. Однак, враховуючи специфіку даної наукової роботи, детально розглядати кожний з варіантів немає необхідності.

2.1.1.2 Вимоги до вхідних даних моделей штучного інтелекту

Незважаючи на те, що фокусом нашого дослідження є розробка методів оцінювання проєктів та якості програмних продуктів, неможливо повністю оминати питання якості вхідних даних, оскільки останні виступають фундаментом для побудови будь-якої моделі штучного інтелекту. Ключовими характеристиками виступають повнота та репрезентативність, що гарантують відсутність критичних прогалин у клінічній історії пацієнтів та дозволяють уникнути алгоритмічного зміщення на певних демографічних групах. Не менш важливою складовою є узгодженість даних, яка в медичному IT-проєкті визначається відповідністю міжнародним стандартам обміну інформацією та єдністю термінології між різними клінічними базами. Також враховується точність та своєчасність, де перевіряється коректність внесених діагнозів та актуальність лабораторних показників відносно поточного стану пацієнта.

Варто підкреслити, що в межах нашої наукової роботи ці параметри сприймаються як фіксовані умови середовища, в якому розробляється система. Ми не ставимо за мету глибоке дослідження методів очищення чи розмітки даних, натомість зосереджуємося на тому, як менеджер проєкту може інтегрувати контроль та оцінку таких показників у загальну систему управління якістю.

2.1.2 Методи управління та розробки систем штучного інтелекту MLOps

Основна філософія MLOps полягає в автоматизації та моніторингу всіх етапів створення інтелектуальної системи: від збору даних до розгортання та регулярного оновлення моделей. Управління якістю реалізується через розширену концепцію безперервних процесів. Окрім стандартної інтеграції та доставки коду (CI/CD), ключовим елементом стає безперервне навчання (що критично важливо принаймні для деяких задач та архітектур), тобто система повинна вміти автоматично виявляти зниження точності та за потреби ініціювати процес донавчання на нових даних без втручання розробників у базову архітектуру. Даний функціонал гарантує, що якість діагностики не погіршиться наприклад через появу нового медичного обладнання.

Однією з головних вимог медичних регуляторів є можливість повного аудиту рішення. Методи MLOps забезпечують версіонування даних та моделей. Для кожного рішення прийнятого системою, менеджер проєкту може точно відстежити, на якій саме версії датасету була навчена модель. Така прозорість є основою для управління юридичними ризиками та забезпечення відповідності стандартам якості. До того ж, система може відстежувати ситуації, при яких вхідні дані пацієнтів суттєво відрізняються від тих, на яких модель була навчена, за потреби можлива реалізація автоматичної блокування видачу результатів або створення запит на ручну перевірку спеціалістом [52].

Системне управління через MLOps дозволяє інтегрувати автоматизовані тести на алгоритмічне зміщення у процес розробки. Перед кожним оновленням системи відбувається перевірка, чи не погіршилася якість роботи моделі для окремих демографічних груп (наприклад, за віком чи статтю). Таким чином виконавці проєкту можуть контролювати якість не лише «в середньому», а й гарантувати рівну ефективність системи для всіх пацієнтів, що є фундаментальним етичним аспектом медицини.

2.1.3 Оцінка сприйняття медичних інформаційних систем з боку користувачів

Як вже зазначалось у попередньому розділі, одним із головних бар'єрів на шляху впровадження інтелектуальних систем у клінічну практику є проблема довіри до роботи моделі. Саме тому оцінка сприйняття системи користувачами вимагає використання пояснювального штучного інтелекту. ХАІ не просто покращує інтерфейс, він стає необхідною умовою для того, щоб медичний персонал міг верифікувати логіку системи та інтегрувати її висновки у свій клінічний досвід. Для задоволення потреб медичного персоналу методи пояснення повинні бути наочними та контекстуальними. У проєктах з аналізу медичних зображень (рентгенографія, МРТ) критично важливим є використання теплових карт (heatmaps) або методів активації класів (saliency maps), які підсвічують ділянки знімка, що найбільше вплинули на рішення ІС. Для моделей, що працюють із текстовими даними, ефективними є візуалізації на основі значень вектора Шеплі (SHapley Additive exPlanations, SHAP values) або алгоритму local interpretable model-agnostic explanations (LIME), які демонструють вагу кожного окремого симптому у фінальному прогнозі. З практичним використанням як вказаних, так і ряду інших методів можна ознайомитись в роботах [53], [54], [55].

Іншим аспектом розгляду сприйняття ІС є позиціонування ІІІ не як альтернативи лікарю, а як системи підтримки прийняття клінічних рішень. В управлінні якістю таких проєктів має чітко простежуватися принцип «доповненого інтелекту» (augmented intelligence), де остаточне рішення та відповідальність завжди залишаються за людиною. Таким чином можна частково знизити психологічний опір персоналу, який може побоюватися професійної заміни, вирішити проблему відповідальності при прийнятті рішення за рахунок використання ІІІ в якості інструменту «другої думки». Детальніше про роботу з користувачами в наступному пункті.

2.2 Управління якістю IT-проектів щодо розробки медичних інформаційних систем

2.2.1 Основні методи управління якістю IT-проектів в медичній галузі

Як вже зазначалось ефективне управління якістю в проектах по створенню медичних систем штучного інтелекту вимагає синтезу класичних управлінських інструментів із гнучкими дослідницькими підходами. На етапі формування вимог до продукту менеджер проекту стикається з необхідністю розробки динамічних специфікацій зі складною та відповідальною предметною галуззю, працівники якої здебільшого не мають достатнього досвіду для репрезентативного представлення вимог (в тій чи іншій мірі таке спостерігається в більшості сфер діяльності). Основний акцент зміщується на вимоги до якості та репрезентативності даних, а також на визначення прийнятних порогів похибки (наприклад, мінімально допустимого рівня точності для діагностики). Важливим аспектом є інтеграція регуляторних вимог безпеки ще на рівні проектування архітектури, що дозволяє уникнути критичних переробок на пізніх стадіях. Процес збору вимог перетворюється на ітераційну перевірку гіпотез, де кожна нова порція отриманих даних може коригувати функціональний обсяг проекту.

Управління ризиками у таких проектах набуває специфічного характеру. Основним методом тут є регулярний ризик-аудит, що включає тестування моделей на нетипових вибірках даних та оцінку ймовірності «деградації» алгоритму після впровадження. Робота з ризиками також передбачає створення планів реагування на випадок клінічних помилок системи, де чітко прописано механізм втручання лікаря (наприклад при використанні нетипових даних, про що ми згадували в пункті 2.1.2), що мінімізує юридичні та репутаційні ризики.

Для оцінки фінансової ефективності та контролю стану проекту використовуються EVA (Earned Value Analysis) та ROI (Return on Investment).

Аналіз освоєного обсягу (EVA) дозволяє менеджеру в реальному часі відстежувати відхилення від графіка та бюджету через такі показники, як індекс виконання вартості, що має сенс особливо для проєктів, де витрати на обчислювальні потужності або залучення висококваліфікованих експертів можуть стрімко зростати. Оцінка ROI в медичному контексті часто базується не лише на прямих прибутках, а й на розрахунку зниження витрат: наприклад, в результаті зменшення кількості повторних госпіталізацій або прискорення діагностичного процесу. Такий підхід дозволяє обґрунтувати економічну доцільність інтелектуальної системи для медичного закладу.

Процеси тестування та контролю якості в таких проєктах трансформуються у багаторівневу валідацію. Класичне тестування програмного коду доповнюється статистичною валідацією моделі. Менеджер проєкту використовує такі результати для прийняття рішень про перехід між етапами, спринтами тощо.

Кінцева якість продукту оцінюється через здатність системи підтримувати задані метрики точності протягом тривалого часу, що робить моніторинг після впровадження невід'ємною частиною управлінського циклу.

2.2.2 Проблеми залучення користувачів до роботи над ІТ-проєктами в сфері охорони здоров'я

Безумовно, залучення майбутніх користувачів до процесу створення медичних ІС є критичною умовою забезпечення якості, оскільки саме на перетині клінічної експертизи та технічної розробки формується придатний до експлуатації продукт. У межах підходу до проектування human-centered design участь медичного персоналу має бути наскрізною: від початкового етапу формування візії системи до фінальної верифікації результатів. Проте такий процес супроводжується значними управлінськими викликами, головним з

яких є критичний дефіцит часу у медичних працівників, що у свою чергу є результатом дефіциту кадрів в галузі, особливо у випадку середнього медичного персоналу. Детальніше з цією проблемою можна ознайомитись у звітах Всесвітньої організації охорони здоров'я [56], [57].

Другою системною проблемою є невідповідність медичного персоналу до чіткої формалізації вимог, що особливо гостро проявляється в проєктах зі ІІІ. Якщо вимоги до звичайної інформаційної системи (такі як реєстрації пацієнтів) лікарі можуть описати через звичні бізнес-процеси, то вимоги до інтелектуальної моделі потребують розуміння природи технології. Без спеціальної підготовки користувачі схильні висувати нереалістичні вимоги (наприклад, 100% точність), або, навпаки, не можуть чітко окреслити сценарії, в яких помилка системи є найбільш критичною, що створює ризик розробки системи, яка технічно відповідає документації, але не має клінічної цінності через невідповідність реальним робочим процесам.

2.2.3 Фреймворк NASSS як інструмент попереднього аналізу перспектив виконання ІТ-проєктів в сфері охорони здоров'я

Фреймворк NASSS представляє собою багаторівневу теоретичну модель, яка дозволяє прогнозувати ступінь успішності або ймовірний крах технологічних програм у сфері охорони здоров'я [11]. З точки зору управлінського підходу, методологія виходить за межі простого виявлення бар'єрів і зосереджується на динамічній взаємодії між різними аспектами проєкту. Ключовим елементом управління ризиками в межах цієї моделі є класифікація різних аспектів проєкту (в роботі виділяють сім доменів) – від стану пацієнта та особливостей технології до адаптації працівників під нову робочу середовище та потенціалу розширення системи – за рівнем їхньої складності. Керівник проєкту отримує можливість ідентифікувати прості,

ускладнені або складні компоненти системи, розуміючи, що наявність високої складності одночасно в багатьох доменах є критичним ризиком, який часто призводить до того, що інновація не стає частиною повсякденної практики. Таким чином можна, ще на ранніх етапах вносити конструктивні зміни в план проєкту або взагалі прийняти рішення про недоцільність останнього.

2.3 Напрями розвитку методів оцінки якості медичних інформаційних систем

Враховуючи проведений аналіз літератури в першому та другому розділі роботи можна звузити та специфікувати напрям розвитку методів оцінки проєктів по створенню ІС в рамках даного дослідження. Далі ми зосередимось на особливостях взаємодії виконавців проєкту і користувачів системи з метою розгляду перспектив залучення останніх в процес створення системи з точки зору управління якістю. Такий підхід зумовлений необхідністю подолання когнітивного розриву між технічними розробниками та кінцевими споживачами продукту, що є критичним для успішної інтеграції ІС в клінічне середовище.

Першим напрямом є розробка інтегрального показника формальної якості, що базується на агрегації класичних метрик машинного навчання. Необхідність створення такого методу зумовлено складністю інтерпретації розрізнених статистичних даних нетехнічними стейкхолдерами та замовниками. Об'єднання відповідних параметрів у єдиний, математично обґрунтований коефіцієнт дозволить сформулювати універсальний індикатор стану розробки. Таким чином забезпечуючи прозорість процесу управління якістю та надаючи можливість зацікавленим сторонам здійснювати оперативний моніторинг прогресу системи без необхідності глибокого занурення в математичний апарат моделей. З іншого боку висування вимог на

етапі ініціації проєкту за цим показником неможливо, однак враховуючи дослідження висування вимог на кількісні показники якості моделей ШІ на ранніх етапах проєкту і так неможливе (принаймні сильно ускладнене) з багатьох причин: від нерозуміння користувачами специфіки даних метрик до неможливості передбачити можливість.

Другим напрямом дослідження визначено метод ретроспективного аналізу та акумуляції результатів тестування користувацького досвіду. Враховуючи високу інтенсивність праці медичного персоналу, пропонується підхід, що базується на систематичному зборі даних про взаємодію користувачів із вже розробленими модулями системи, що потребують користувацького тестування. Об'єктом аналізу в даному контексті виступатиме не лише ергономічність інтерфейсу, а й динаміка зміни рівня довіри та інтерпретованості блоків штучного інтелекту в медичній інформаційній системі. Такий підхід дозволить виявити приховані закономірності у сприйнятті системи ще на ранніх етапах тестування, що забезпечить превентивний контроль якості та мінімізує ризики відторгнення фінального продукту.

Синтез вказаних двох методів дозволить збалансувати формальні характеристики системи з суб'єктивними факторами експлуатаційної придатності. Таким чином створивши надійний інструментарій для менеджера проєкту, що поєднує в собі контроль математичної точності та управління очікуваннями користувачів.

Відповідно подальше дослідження буде спрямоване на формалізацію вимог до вказаних інструментів та теоретичне моделювання розроблених методів як ключових елементів підтримки прийняття рішень в процесах управління та оцінки якості ІТ-проєктів штучного інтелекту по розробці медичних інформаційних систем.

2.4 Вимоги до показників оцінки якості медичних інформаційних систем штучного інтелекту

Враховуючи описане вище можна сформулювати мінімально необхідні вимоги та обмеження, що накладаються на метрики оцінки якості інтелектуальних систем в сфері охорони здоров'я з метою максимізації користі для зацікавлених сторін проєктів та направлені на вирішення проблем вказаних в минулому пункті.

2.4.1 Вимоги до інтегрального показника точності моделі штучного інтелекту

В якості вимог до інтегрального показника точності інтелектуальної моделі можна виділити наступні пункти:

- математична незалежність компонентів;
- залежність від задачі, а не архітектури;
- нормування шкали;
- чутливість до незбалансованості даних.

Перейдемо до опису вказаних компонентів. Перша вимога встановлює обмеження на характер набору технічних показників, що одночасно використовуються при акумулюванні інтегрального метрики: для уникнення ефекту подвійного зарахування одного і того ж фактору до складу мають входити показники, що не мають прямої математичної залежності один від одного, наприклад, не можна одночасно включати f1-score та recall, оскільки перший вже містить в розрахунках другий. Зрозуміло, що кореляція показників між собою завжди присутня, однак необхідно уникати хоча б прямих залежностей.

Отримана метрика повинна мати окреслені механізми порівняння декількох моделей (хоча б в рамках одного типу задач машинного навчання), які навчаються на одному типі даних. Відповідно, особливості архітектури, як-от кількість параметрів або показник функції втрат (специфічних для одної моделі), в такому випадку не мають впливати на аналіз (з цією причини при описі метрик в пункті 2.1.1 були проігноровані подібні архітектурні особливості). З метою більш простого розуміння та використання отриманий індекс можна нормувати від нуля до одиниці (де нуль – найнижчий показник точності) або привести у відсотках. Остання вимога вимагає використання показників, що коректно відображають результати отримані на незбалансованих наборах даних (наприклад за рахунок використання f1-score замість accuracy).

2.4.2 Вимоги до показника оцінки якості медичної інформаційної системи користувачами

В якості обмежень та вимог до показника оцінки якості системи користувачами як з точки зору зручності користувацького інтерфейсу, так і довіри до висновків інтелектуальної складової можна перелічити наступне:

- квантифікація якісних відгуків;
- контекстуальна поправка на досвід користувача;
- часова поправка при використанні системи;
- профілізація за посадою користувача;
- динаміка змін оцінок в залежності від часу роботи та рівня завершеності прототипу.

Перша вимога окреслює необхідність впровадження механізму перетворення отриманих суб'єктивних оцінок у кількісні величини, тим чи іншим способом.

Поправки на досвід користувача вносить коректив на основі терміну роботи у відповідному закладі або на даній посаді (можливе додавання досвіду користування інформаційними системами взагалі). Аналогічні поправки вносяться за посадою та часом використання в рамках робочого дня або зміни. Наприклад, якщо більшість функціоналу системи орієнтована на лікарів, тоді оцінки молодшого медичного персоналу, що користується системою в середньому двадцять хвилин за зміну, має оцінюватись нижче, ніж відгуки самих лікарів (такий підхід застосовується до прототипу системи, що вже не має суттєвих недоліків, які унеможливлюють роботу того ж молодшого медичного персоналу).

Динаміка змін оцінок та застосування поправки на теперішній стан прототипу ставить за мету позбутись впливу ефекту новизни (необґрунтовано високої оцінки на ранньому етапі) або навпаки раннього скептицизму користувачів щодо якості інформаційної системи або доцільності розробки останньої взагалі.

3 ОПИС ЗАПРОПОНОВАНИХ МЕТОДІВ ОЦІНКИ ЯКОСТІ МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

3.1 Розроблений метод інтегральної оцінки технічної якості моделі штучного інтелекту

Індекс узагальненої оцінки формальних показників оцінки точності штучного інтелекту, згідно з вимогами викладеними в попередньому розділі, в розрахунку потребує використання незалежних технічних метрик (інакше внесок того чи іншого фактору приймався би до уваги більше одного разу). Дотримання даної вимоги лежить на виконавцях проєкта: при порівняння моделей між собою необхідно вказувати перелік характеристик, які були використані. У випадку порівняння декількох моделей в рамках одного проєкту (та однієї задачі машинного навчання), дана задача вирішується сама собою, однак у випадку порівняння зовнішніх систем, додавання переліку метрик є необхідним. Схожим чином виконується вимога щодо залежності від задачі, а не від архітектури: виконавці мають брати до уваги спільні для всіх моделей метрики.

Відповідно перший етап методу полягає у виборі (з переліку використаних формальних метрик) незалежної підгрупи показників, згідно з вимогами вказаними вище та особливостями використання конкретної МІС.

На другому етапі проводиться нормування метрик за необхідністю. При згортанні різних технічних параметрів необхідно спочатку привести останні до єдиного діапазону значень. Найбільш зрозумілим та популярним є проміжок від нуля до одиниці. Деякі метрики вже приведені до такого виду, інші приводяться за допомогою мінімаксного нормування, за умови відомих граничних значень. Однак такі значення не завжди існують або мають сенс, наприклад в показниках більшості функції втрат максимальне значення відсутнє (або дорівнює нескінченності) а мінімальне дорівнює нулю, однак на практиці такий результат є недосяжним (навіть при максимальній точності,

функція втрат зазвичай не дорівнює нулю). В статистиці є методи вирішення задачі нормування при таких умовах, однак враховуючи, що такі показники як квартилі чи стандартне відхилення таких величин, в загальному випадку невідомі, арсенал методів нормування є обмеженим. До того ж, значення функції втрат необхідно не тільки нормувати, а й інвертувати (на відміну від точності, менше значення є кращим).

Серед можливих варіантів нормування з інвертуванням значення для необмежених величин можна використовувати стандарту та експоненціальну інверсію (у даних методів можливі різні модифікації, як-то логарифмічне згладжування або інверсія з насиченням), формули яких наведено нижче (як для інвертованих, так і неінвертованих).

$$N_i = \frac{1}{1 + \alpha V}, \quad (3.1)$$

$$N_i = e^{-\alpha V}, \quad (3.2)$$

$$N = \frac{V}{V + \alpha}, \quad (3.3)$$

$$N = 1 - e^{-\alpha V}, \quad (3.4)$$

де N_i – нормоване та інвертоване значення;

α – коефіцієнт розмірності;

V – початкове значення величини;

N – нормоване значення.

Основною проблемою таких підходів до нормалізації та інвертування значень (формули 3.1 – 3.4) є необхідність розрахунку коефіцієнта альфа, що характеризує допустиме значення діапазону вибірки та дозволяє коректно

відображати співвідношення всередині діапазону. На практиці такий коефіцієнт можна розрахувати на основі мінімально допустимого значення початкові величини (наприклад функції втрат) встановлене через експертну оцінку або наявність базової моделі (існує набір методів, які розвивають цю ідею, однак не розглянуті, в першу чергу через обмежений потенціал їх використання при порівняння різних проєктів), від показників якого можна розвивати якість системи. Використання альфа, що дорівнює одиниці в формулі 3.1 може бути виправданим в деяких випадках, однак приводить до спотворень: гірші результати можуть виглядати кращими, ніж вони є, а логарифмічні згладжування лише частково вирішують проблему.

На третьому етапі, після приведення до єдиної шкали, значення параметрів можна комбінувати за допомогою середнього геометричного. Такий спосіб більш підходить для ситуацій, при яких низький показник хоча б з однієї метрики, має більший вплив на результат, ніж при середньому арифметичному. Адаптовану формулу наведено нижче.

$$CAI = \left(\prod_{i=1}^n k_i * x_i \right)^{\frac{1}{n}}, \quad (3.5)$$

де CAI – інтегральний показник точності моделі;

n – кількість параметрів;

k_i – ваговий коефіцієнт метрики;

x_i – нормалізоване значення метрики.

Кожне значення метрики в розрахунках супроводжується ваговим коефіцієнтом за необхідністю, під контекст конкретної задачі: особливості галузі, структура набору даних та інші. За замовчуванням такі коефіцієнти можна залишити рівними одиниці.

Отриманий інтегральний показник точності (CAI) є корисною абстракцією для нетехнічних виконавців проєкта та інших зацікавлених сторін для відстежування прогресу розробки та встановлення подальших вимог після перших тестувань.

3.2 Розроблений метод оцінки медичної інформаційної системи користувачами

Показник оцінки якості системи на основі відгуків користувачів доцільно проводити за рахунок усного або письмового/електронного анкетування (відповідно можна використовувати анонімний або відкритий збір інформації). Як вже зазначалось, дана методика підходить для тестування системи або її окремого блоку, що вже має закінчений (згідно з попередніми вимогами) функціонал, який охоплює один бізнес-процес або більше. В таких умовах окрім чітко виражених проблем в користувацькому інтерфейсі виникає необхідність оцінити загальне враження від провадженої системи. Іншим аспектом може бути оцінка якості елементів пояснювального штучного інтелекту, перевірка на доступність автоматично створених візуалізацій та звітів. Схожим чином можна оцінювати вже попередні системи (або версії даної системи) для визначення поточних переваг та недоліків.

Відповідно до описаних у минулому розділі (п. 2.4.2) вимог запропоновано метод аналізу за рахунок квантування результатів анкетування опитаних користувачів методами нечіткої логіки на основі шкал Лікерта. Нечітка логіка (fuzzy logic) дозволяє математично формалізувати розмиті поняття людської мови та мислення, як-от «майже», «досить» або «переважно», через так звані функції належності, що перетворюють суб'єктивні оцінки у конкретні числові значення. В опитуваннях нечітка логіка застосовується шляхом використання лінгвістичних шкал (замість

«так/ні» – варіанти «часто», «іноді», «рідше»), що дозволяють респондентам вказати ступінь своєї згоди.

Такий підхід є ефективним, оскільки значно точніше відображає реальний спектр людських думок, мінімізує втрату нюансів при виборі між жорстко обмеженими варіантами та дає змогу отримувати математично обґрунтовані результати на основі якісних, а не лише кількісних даних. Таким чином досягається виконання вимог щодо квантифікації відгуків користувачів. Детальніше з використанням нечіткої логіки в проведенні опитувань та аналізі результатів можна ознайомитись в перелічених книзі та статтях [58], [59], [60].

Виконання вимог щодо врахування часу роботи, досвіду та кваліфікації користувача, відбувається, через додавання до розрахунків вагових коефіцієнтів, які задаються анкетером на основі особливостей системи, що розробляється. Відстежування динаміки змін досягається при аналізі результатів декількох опитувань (можливо відслідковувати зміни у кожного користувача при відкритій формі опитування). Основні питання можна формувати у вигляді тестування з вибором п'яти можливих варіантів відповіді (від суто негативного до позитивного), що описують ставлення людини, до аспекту взаємодії зі системою. Для шкали кожного питання (або для спільної шкали питань) надається на кожний варіант відповіді нечітке число (задане у вигляді функції приналежності). Для спрощення діапазон можна задати від нуля до одиниці або у відсотках. У випадку функції приналежності трикутного виду для негативної відповіді значення може виглядати наступним чином (0, 0, 0,25), для середнього – (0,4, 0,55, 0,7), для позитивного – (0,8, 1,0, 1,0).

На першому етапі методу формуються запитання в анкетуванні з урахуванням задач та цільової групи, варіанти відповіді, обирається та використовується тип нечіткої функції приналежності та значення для альтернатив. За потреби додаються вагові коефіцієнти по кожному питанню.

Другий етап починається після проведення анкетування (безпосередньо проведення анкетування не є частиною методу, а виступає скоріше вхідними

даними, незважаючи на те, що структура опитування потребує підготовки для подальшої обробки методом оцінки медичних систем користувачами, отримані результати анкетування можуть бути використані і за межами запропонованого методу) та представляє з себе первинну обробку для розрахування загального вагового коефіцієнту кожного опитаного на основі відповідей про кваліфікацію, час роботи з системою за зміну, стаж та за іншими факторами у разі потреби. Перелічені категорії також можуть мати вагові коефіцієнти. Формулу розрахунку наведено далі.

$$C_r = w_p * P + w_s * S + w_t * T + w_l * L, \quad (3.5)$$

де C_r – загальний ваговий коефіцієнт респондента;

w_p, w_s, w_t, w_l – вагові коефіцієнти категорій кваліфікації, загального стажу роботи, часу використання системи за зміну та стажу роботи в даному закладі;

P, S, T, L – значення відповідей по вказаним вище категоріям.

Для простоти подальших обчислень, на даному етапі також необхідно привести отримані значення до такого вигляду, щоб сума вагових коефіцієнтів дорівнювала одиниці, зі збереженням статистичних відношень. Досягти такого ефекту можна шляхом поділу кожного отриманого значення на суму усіх коефіцієнтів респондентів (у тому числі й враховуючи ваговий коефіцієнт респондента, значення якого перетворюються)

На третьому етапі проводиться агрегація значень за кожним питанням. Результат розраховується як добуток вагових коефіцієнтів респондентів на значення функцій приналежності. Конкретна формула може суттєво відрізнятись в залежності від обраної функції приналежності. За необхідністю можна проводити додаткову аналітику: виявити яка цільова група задоволена чи незадоволена тим чи іншим аспектом системи. Такий аналіз хоча і не частиною методу, однак враховуючи структур опитування, може бути

корисним для поглибленого розуміння переваг та недоліків медичної інформаційної системи, що розглядається. Одночасно, таким чином можна викрити аномальні зміщення різного роду у відповідях опитаних і відслідковувати зміну ставлення працівників до МІС впродовж деякого часу.

На четвертому етапі методу необхідно розрахувати значення нечіткої функції вже для самого індексу сприйняття системи користувачами (USPI) як зваженої суми по кожному питанню, як показано у формулі 3.7.

$$USPI_{fuzzy} = \sum_{i=1}^n v_i * Q, \quad (3.6)$$

де $USPI_{fuzzy}$ – значення індексу у вигляді нечіткої функції;

n – загальна кількість питань, які задіяні в методі;

v_i – ваговий коефіцієнт питання;

Q – агреговане значення нечіткої функції окремого питання.

На останньому етапі виконується процес дефазифікація отриманого на минулому кроці значення тим чи іншим методом, наприклад центром площі. Як і на третьому етапі, формула може суттєво відрізнятись від обраної функції приналежності. Приклад використання методу наведено у наступному розділі.

4 АПРОБАЦІЯ РОЗРОБЛЕНИХ МЕТОДІВ ОЦІНКИ ЯКОСТІ МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

4.1 Приклад використання інтегрального показника точності моделі штучного інтелекту

Для демонстрації на практиці описаної у минулому розділі теоретичної бази, в якості джерела даних було обрано роботу [61], що відображає результати використання декількох моделей, побудованих на архітектурі автокодувальників для вирішення задачі бінарної класифікації МРТ знімків опорно-рухового апарату. В роботі розглядається вирішення за рахунок методів *unsupervised learning*: в умовах незбалансованих вибірок моделі навчаються на даних здорових пацієнтів для згортання зображень у вектор латентного простору і розгортання назад у зображення. Таким чином у випадку подачі на вхід даних хворих пацієнтів, система не може коректно провести цикл кодування (низький рівень співпадіння початкового і отриманого зображень), відповідно можна провести класифікацію і відділити дані здорових та хворих пацієнтів [62].

У вказаній роботі проаналізовано три моделі автокодувальників за наступними п'ятьма метриками: пікове відношення сигнал-шум (PSNR), середня абсолютна помилка (MAE), середньоквадратична помилка (MSE), індекс структурної подібності (SSIM) та площа під кривою (AUC). Нижче, в таблиці 4.1 наведено результатами оцінки моделей [61].

Таблиця 4.1 Ефективність роботи методів III з дослідження [61]

Model / Metric	AUC	MSE	PNSR [dB]	SSIM	MAE
Vanilla Autoencoder	0.737	0.057	18.65	0.409	0.173
Autoencoder with Perceptual loss	0.857	0.019	23.59	0.640	0.098
Variational autoencoder	0.839	0.025	22.35	0.570	0.114

Метрика AUC приймає значення від нуля до одиниці, не залежить від інших метрик та не потребує нормалізації або інвертування (на відмінну від решти). Аналогічна ситуація з метрикою SSIM, що описує структурну подібність на рівні фрагментів зображення, а не окремих пікселів. Незважаючи на те, що діапазон значень останньої від -1 до 1, в додатковій обробці немає потреби, бо значення менше нуля неможливе в контексті задачі (детальніше в роботі [63]). Інші три метрики входять в категорію залежних між собою.

Нами було обрано метрику функції середньої квадратичної помилки в якості третього параметру, через зручність, простоту оптимізації та інтерпретації (різниця між двома пікселями в другому степні). Значення було нормовано, за допомогою формули 3.1 зі значенням коефіцієнту поправки альфа, що дорівнює 3,84. Таке значення отримане за рахунок припущення, що граничне допустиме нормоване та інвертоване значення метрики становить 0,8 і відповідає ненормованому значенню втрат 0,065. Вагові коефіцієнти кожної метрики взяті за одиницю. Отримані результати розрахунків за формулами 3.1 та 3.5 занесені в таблицю 4.2.

Таблиця 4.2 Результати використання методу оцінки за інтегральним показником точності

Модель / Метрика	MSE норм.	CAI
Vanilla Autoencoder	0,82	0,63
Autoencoder with Perceptual loss	0,93	0,80
Variational autoencoder	0,91	0,76

Як можна бачити отримані результати розрахунків за індексом CAI відповідають співвідношенню між моделями по іншим метрикам. Архітектура кодувальника з перцепційною функцією втрат показала найкращий результат, на другому місці варіаційний автокодувальник. Таким чином інтегральний показник точності виступає в ролі високорівневої абстракції над стандартними

метриками за рахунок інкапсуляції формальних деталей, відповідно полегшуючи розуміння та аналіз рівня якості моделі ШІ зацікавленими сторонами проєкта. В даному випадку для простоти демонстрації було обрано дослідження з невеликою кількістю метрик, однак в залежності від задачі ця кількість може досягати десятків, а в такому випадку користь від використання інтегрального індексу суттєво зростає.

4.2 Приклад використання індексу сприйняття медичної інформаційної системи користувачами

В умовах відсутності у відкритому доступі необхідних результатів інтерв'ю або опитувань медичних працівників в необхідному форматі, нами було проведено власне анкетування серед персоналу різних закладів охорони здоров'я міста Харків, Україна. У зв'язку з цим, в опитуванні відсутні питання, що стосуються досвіду використання саме інтелектуальної інформаційної системи (так як в Україні, такі технології ще не впроваджені, принаймні масово), до того ж питання мають загальний характер, що також пояснюється особливостями збирання даних. Однак, незважаючи на це, як демонстрація ходу обчислень за методом оцінки якості сприйняття системи користувачами такий варіант є допустимим.

4.2.1 Характеристика проведеного опитування користувачів медичних інформаційних систем

Анкетування було проведено в форматі анонімного онлайн опитування через платформу Google Forms. Запитання розділені на два блоки: перший

представляє з себе загальні питання, необхідні для статистичної звітності та розрахунку вагових коефіцієнтів респондентів, другий блок складається з питань, що відносяться безпосередньої до вражень та досвіду користування медичними інформаційними системами (МІС). Напочатку є преамбула з коротким описом анкети, приблизного часу проходження, мотивацією такого дослідження та контактами з автором.

Перший блок складається з п'яти обов'язкових питань (шосте питання стосується назви МІС та не є обов'язковим), відповіді на які відображають кваліфікацію опитаного (рис 4.1), стаж роботи взагалі та на останньому робочому місці, середній час використання МІС за робочу зміну, місце роботи та назву медичної системи.

Вкажіть вашу кваліфікаційну категорію / посаду. *

Выбрать

! Это обязательный вопрос.

Выбрать

Лікар

Середній медичний персонал

Молодший медичний персонал

Завідувач відділення / адміністративний медичний персонал

Працівник реєстратури / інших служб закладу

Рисунок – 4.1 Питання про кваліфікацію респондента та можливі варіанти відповіді

Повний перелік питань першого блоку наведено нижче:

- 1) «впишіть назву медичного закладу в якому працюєте зараз (або заклад останнього місця роботи, про комп'ютерну систему якого будете проходити опитування)»;
- 2) «вказіть вашу кваліфікаційну категорію / посаду»;
- 3) «вказіть ваш загальний стаж професійної діяльності»;
- 4) «вказіть час, який ви вже пропрацювали на нинішньому (або останньому) місці роботи»;
- 5) «вказіть який час ви зазвичай витрачаєте на роботу за комп'ютером за зміну»;
- 6) «вказіть (необов'язкове питання), назву медичної комп'ютерної системи, що використовується на вашому робочому місці (наприклад МедІнфоСервіс, eHealth, тощо)».

У другому блоці сім питань покривають основні концепти роботи з МІС в форматі шкали Лікерта з п'ятьма варіантами відповіді, від найгіршого до найкращого враження (усі питання мають пряму шкалу). Варіанти відповідей надано у вигляді словосполучень або коротких речень, які зрозуміло описують опцію в залежності від контексту. Питання другого блоку поставлені наступним чином (нумерацію продовжено з першого блоку):

- 7) «наскільки легко вам орієнтуватися в розділах/сторінках комп'ютерної системи без звернення до колег або інструкцій»;
- 8) «оцініть рівень оптимізації навігації в системі під час заповнення медичної документації (кількість необхідних кліків миші, перемикань між сторінками та вкладками екрану)»;
- 9) «наскільки оперативно та зручно вам вдається знаходити необхідні відомості (наприклад результати досліджень, висновки профільних спеціалістів, дані анамнезу) у вашій електронній системі»;
- 10) «оцініть зручність сприйняття графіків, таблиць і списків пацієнтів на вашому моніторі»;

11) «наскільки корисними та своєчасними є попередження системи про некоректне введення даних (порожнє ім'я/прізвище, неможливий рік народження, некоректне дозування ліків та інше)»;

12) «як часто за тиждень в середньому ви стикались з технічними збоями або зависаннями системи»;

13) «як швидко технічні спеціалісти/колеги здебільшого вирішують проблеми, що виникають у вас у комп'ютерній системі».

Нижче, на рисунку 4.2 наведено приклад питання, про оптимізації навігації в системі при заповненні даних пацієнта. Описані вище блоки запитань та варіанти відповідей занесено в додаток А.

Оцініть рівень оптимізації навігації в системі під час заповнення медичної документації (кількість необхідних кліків миші, перемикань між сторінками та вкладками екрану). *

- ☐ Вкрай низький: надмірна кількість переходів і маніпуляцій, що суттєво уповільнює робочий процес
- ☐ Низький: структура інтерфейсу потребує виконання зайвих дій для завершення стандартних операцій
- ☐ Середній: рівень навігаційного навантаження прийнятний, хоча потребує концентрації на технічних діях
- ☐ Високий: навігація логічна, кількість кліків мінімізована для більшості завдань
- ☐ Оптимальний: введення даних здійснюється "безшовно" з мінімальними витратами зусиль

Рисунок – 4.2 Питання про оптимізацію навігації в системі

4.2.2 Характеристика респондентів проведеного опитування

За десять днів в опитуванні прийняли участь 17 медичних працівників, серед яких чотири лікаря, одинадцять представників середнього мед. персоналу, по одному молодшому та адміністративному персоналу. Як медичні інформаційні системи було обрано програмне забезпечення МедІнфоСервіс [64] та eHealth (підключаються через першу). Такий набір є стандартним для закладів охорони здоров'я в Україні.

Серед харківських мед. закладів працівники яких, прийняли участь в опитуванні, є КНП ХОР ОКПЛ №3, ТОВ Слобожанський центр нефрології, ХОКЛ, ЦКБ №5, КНП «МКЛШНМД ім. проф. О. І. Мещанінова» ХМР, КНП «Міська поліклініка №21» ХМР. Розподіл у відсотках наведено на діаграмі 4.3.

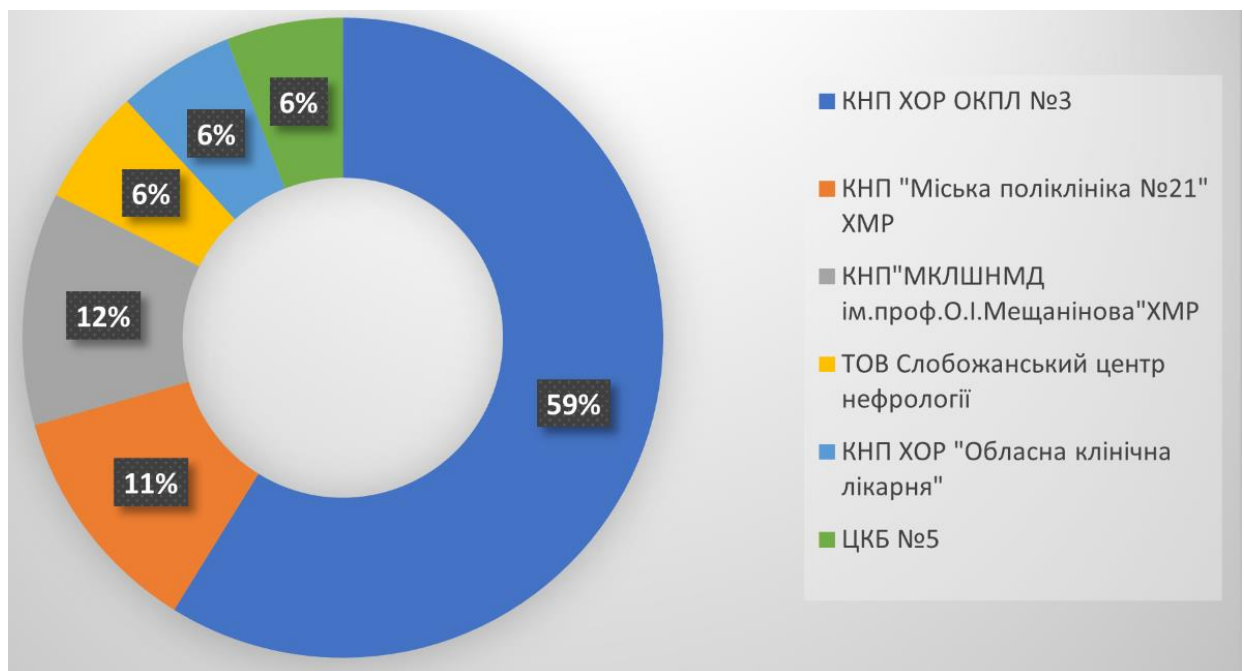


Рисунок – 4.3 Розподіл респондентів по медичним закладам

4.2.3 Розподіл вагових коефіцієнтів та значень функцій приналежності

Насамперед необхідно розподілити внесок категорій та відповідей на питання по кожній категорії. Нижче, в таблиці 4.3, вказані необхідні вагові коефіцієнти: для простоти поряд з коефіцієнтами наведено пояснення у вигляді тексту (у вигляді рисунків наведені у додатку А), останній рядок відноситься до самих категорій. Обидва типи коефіцієнтів мають змінюватись враховуючи специфіку конкретного проєкта, наприклад в даному випадку система орієнтована на лікарів та середній медичний персонал (найбільші коефіцієнти у відповідних категоріях і найбільший коефіцієнт самої категорії). Коефіцієнти стажу та часу використання також можуть змінюватись в залежності від ситуації.

Таблиця 4.3 Вагові коефіцієнти категорій

Відповідь №	Посада	Загальний Стаж	Стаж на останньому у місці	Час в системі за зміну
1	0,05 (персонал реєстрації)	0,1 (менше 2 років)	0,15 (менше 1 року)	0,05 (до 45 хвилин)
2	0,1 (адміністративний персонал)	0,2 (від 2 до 4 років)	0,2 (від 1 до 2 років)	0,15 (до 1 години)
3	0,15 (молодший персонал)	0,3 (від 5 до 10 років)	0,3 (від 2 до 4 років)	0,2 (від 1 до 2 годин)
4	0,25 (середній персонал)	0,4 (більше 10 років)	0,35 (більше 4 років)	0,25 (до 3 годин)
5	0,45 (лікар)	-	-	0,35 (більше 3 годин)
Загальний по категорії	0,3	0,25	0,2	0,25

Для спрощення обчислень та візуалізації в якості нечітких функцій приналежності було обрано трикутні функції (іншим можливим варіантом є

використання подвійної функції Гауса або функцій з тривалим максимум). На кожне питання є 5 можливих варіантів відповідей від негативного до позитивного, в такому порядку значення наведені в таблиці 4.4. Функції приналежності [65] в графічному вигляді наведені в додатках Б.

Таблиця 4.4 Значення функцій приналежності по відповідям на питання другого блоку

Орієнтація	Опт. навігації	Пошук	Зручність	Попередження	Кількість збоїв	Підтримка
1	2	3	4	5	6	7
0	0	0	0	0	0	0
0,05	0,05	0,05	0,1	0,25	0	0
0,12	0,1	0,10	0,25	0,3	0,1	0,1
0,1	0,1	0,08	0,25	0,25	0,05	0,05
0,15	0,25	0,2	0,3	0,3	0,15	0,15
0,35	0,35	0,4	0,4	0,35	0,25	0,25
0,3	0,3	0,3	0,35	0,3	0,2	0,2
0,4	0,4	0,6	0,45	0,45	0,35	0,5
0,55	0,5	0,7	0,55	0,65	0,45	0,65
0,5	0,4	0,65	0,5	0,55	0,45	0,55
0,7	0,65	0,7	0,6	0,75	0,7	0,6
0,8	0,8	0,85	0,75	0,9	0,85	0,85
0,8	0,75	0,8	0,7	0,85	0,8	0,8
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0,1	0,1	0,15	0,2	0,2	0,2	0,05

Окрім функцій приналежності в таблиці 4.4 наведено вагові коефіцієнти кожного питання (в останньому рядку). Як можна бачити, значення відрізняються в залежності від питання, наприклад питання про оперативність роботи технічної підтримки не відноситься напряду до роботи системи (хоча якість та простота системи зменшує час на вирішення різного роду проблем з останньою, однак в більшій мірі відповідальність за такого роду технічну підтримку лежить на відповідних працівниках медичних закладів), тому має найменший вплив на остаточні розрахунки.

4.2.4 Розрахунок індексу оцінки медичної інформаційної системи користувачами

Першим кроком аналізу результатів та попередніх розрахунків є заміна отриманих відповідей на числові аналоги. Додаток В містить таблицю, яка відображає опитування персоналу питаннями другого блоку у вигляді значень трикутної функції приналежності.

Далі були проведено розрахунки вагових коефіцієнтів респондентів за формулою 3.6 та наступне приведення до необхідного діапазону (таблиця 4.5), за рахунок поділу на суму коефіцієнтів усіх опитаних (сумарне значення яких дорівнює 5,3975). Найбільші значення отримали лікарі з найбільшим загальним стажем та стажем роботи на останньому місці, за рахунок максимальних значень вагових коефіцієнтів одразу за декількома категоріями. Для простоти візуалізації в якості індексів респондентів приведені (та далі будуть використовуватись) номери від 1 до 17 (аналогічний підхід в додатку В), в порядку проходження анкетування.

Таблиця 4.5 Розрахунки вагових коефіцієнтів респондентів

№	Посада	Загальний стаж	Стаж на останньому місці	Час в системі за зміну	Коеф. респондента	Нормований коеф.
1	2	3	4	5	6	7
1	0,25	0,4	0,35	0,35	0,3325	0,061602594
2	0,25	0,4	0,35	0,35	0,3325	0,061602594
3	0,45	0,1	0,2	0,35	0,2875	0,053265401
4	0,25	0,4	0,35	0,35	0,3325	0,061602594
5	0,45	0,4	0,35	0,35	0,3925	0,072718851
6	0,25	0,4	0,35	0,15	0,2825	0,052339046
7	0,25	0,4	0,35	0,15	0,2825	0,052339046
8	0,25	0,4	0,35	0,35	0,3325	0,061602594
9	0,25	0,4	0,2	0,35	0,3025	0,056044465
10	0,25	0,4	0,35	0,35	0,3325	0,061602594

Кінець таблиці 4.5

1	2	3	4	5	6	7
11	0,25	0,4	0,35	0,25	0,3075	0,05697082
12	0,25	0,4	0,35	0,25	0,3075	0,05697082
13	0,25	0,3	0,35	0,15	0,2575	0,047707272
14	0,1	0,4	0,35	0,35	0,2875	0,053265401
15	0,15	0,2	0,3	0,35	0,2425	0,044928208
16	0,45	0,4	0,35	0,35	0,3925	0,072718851
17	0,45	0,4	0,35	0,35	0,3925	0,072718851

Далі проводиться розрахунок значень нечіткої функції приналежності для кожного питання як сумарного добутку відповідей респондентів на вагові коефіцієнти останніх. У випадку трикутної функції даний процес зводиться до окремого розрахунку мінімального, середнього та максимального значень. Результати розрахунків наведено в таблиці 4.6. Отримана агрегація відображає нечіткі границі оцінок респондентів по заданим питанням, наприклад найвище опитані оцінили вплив підказок та попереджень системи (заповнення некоректними даними), в той же час найнижчі показники отримали кількість збоїв (шкала Лікерта на всіх питаннях є прямою, тому низькі показники свідчать про високу кількість таких технічних проблем) та оперативність роботи технічного персоналу по виправленню подібних проблем.

Таблиця 4.6 Агрегація значень нечітких функцій для кожного питання з другого блоку

Питання	Функція приналежності			Коеф. Питання
	Min	Mid	Max	
Орієнтація	0,3442334	0,464845	0,602455	0,1
Опт. навігації	0,2599815	0,370218	0,469245	0,1
Пошук	0,3280269	0,499166	0,6456	0,15
Зручність	0,3783001	0,47522	0,586707	0,2
Попередження	0,4197545	0,583279	0,720079	0,2
Кількість збоїв	0,0601204	0,134414	0,237031	0,2
Підтримка	0,1529643	0,312668	0,448263	0,05

Останнім кроком є розрахунок кінцевого показника сприйняття системи користувачами у нечіткому вигляді та подальша дефазифікація. Перше досягається за формулою 3.7 – зважена сума значень функції приналежності відповідей на питання (зважена за коефіцієнтами запитань з таблиці 4.4 або таблиці 4.5). В результаті отримано нечітке число (0,29 0,41 0,54).

У випадку трикутної функції найбільш простий та підходящий для контексту нашої задачі алгоритм дефазифікації є метод центру тяжіння, який з точки зору розрахунків є середнім арифметичним трьох чисел. В результаті маємо показник 0,41. Значення є доволі низьким, особливо враховуючи, що внесок одного з найбільш низьких за показниками питань (про якість роботи технічного персоналу щодо виправлення технічних проблем в системі) має найменший ваговий коефіцієнт – 0.05. Хоча отриманих результатів об'єктивно недостатньо для тих чи інших висновків, стосовно стану програмного забезпечення (або можливо рівня підготовки медичного персоналу для роботи з даними МІС), такі показники є невтішними. Як можна бачити з таблиці В.1 лише два респондента відповіли найбільш позитивною оцінкою по одному з питань, в той же час відповідей з мінімальними оцінками майже у п'ять разів більше і сконцентровані такі відповіді на питаннях про кількість збоїв та рівень технічної підтримки. Частково це можна пояснити проблемами в забезпеченні в одному або декількох закладах (тобто проблема є локальною), до того ж кількість респондентів доволі мала та не дозволяє зробити статистичне узагальнення – вибірка є потенційно нерепрезентативною. Більше того, існує цілий ряд аспектів, що відносяться до психології проведення подібних анкетувань оснований на шкалах Лікерта, наприклад респонденти рідко обирають крайні варіанти відповіді, намагаються усереднювати відповіді та інше.

З іншого боку, робити висновки про якість МІС, що використовуються в Україні не є метою даного дослідження. Проведене опитування та подальші розрахунки продемонстрували на практиці використання методу оцінки медичних інформаційних систем користувачами.

ВИСНОВКИ

В процесі виконання даної науково-дослідницької кваліфікаційної роботи магістрів було розглянуто особливості медичної галузі з точки зору виконання ІТ-проектів щодо створення та інтеграції систем штучного інтелекту. Основним вектором дослідження є аналіз контексту подібних проектів з метою виявлення специфічних переваг та недоліків при плануванні оцінки якості інтелектуальних медичних систем, подальшого формування теоретичної основи та практичної реалізації запропонованих метрик.

На першому етапі, беручи до уваги розглянуті літературні джерела, було описано особливості ІТ-проектів в сфері охорони здоров'я та проектів штучного інтелекту (як окремих, так і в медицині). В результаті було сформовано особливості планування та виконання таких проектів. Отримані результати опубліковані у вигляді тез міжнародної конференції у виданні [66].

Подальший аналіз літератури зосереджено саме на управлінні якістю в умовах особливостей предметної галузі описаних на минулому етапі. Було розглянуто технічні та організаційні сторони питання, серед яких як окремі метрики (точність, f1-score, MCC та інші), так і готові фреймворки (MLOps, NASSS). На основі виявлених закономірностей, спектр подальших досліджень звужено до розробки двох методів оцінки якості ІС, вимоги до яких було сформовано в другому розділі роботи. Отримані специфікації опубліковані у вигляді статті на міжнародній конференції [67].

Третій та четвертий розділи сфокусовані на розробці двох методів оцінки якості інтелектуальних медичних систем з теоретичної та практичної точки зору відповідно до встановлених у другому розділі вимог. Інтегральний показник точності (CAI) представляє з себе математичну композицію формальних метрик оцінки якості моделі штучного інтелекту, таким чином створюючи високорівневу абстракцію, що допомагає зацікавленими сторонам проекту контролювати прогрес розробки системи без необхідності отримання

глибоких знань в галузі data science. Демонстрація методу проводилась на даних з досліджень [61] та [62]. Публікація відбулась у вигляді статті в міжнародній конференції [68].

Індекс сприйняття системи (USPI), у свою чергу, є методом квантифікації результатів опитування користувачів щодо поточного стану останньої (як з точки зору загальних аспектів роботи з системою, так і якості та рівня довіри або інтерпретації штучного інтелекту). В даній реалізації методу конвертація якісних відповідей респондентів у числові значення відбувається за допомогою алгоритмів нечіткої логіки. Для демонстрації ходу обчислень було проведено анкетування медичних працівників міста Харків. Стаття з результатами розробки USPI опублікована в наступному міжнародному виданні [69].

Розроблені методи мають широкий простір для подальшого розвитку як в межах медичної галузі, так і в інших сферах. По-перше необхідно провести ряд досліджень та модернізацій з метрикою CAI, направлених на узагальнення використання при необхідності нормалізації та інвертування внутрішніх показників (детальніше проблема описана в третьому розділі). В той же час, інтегральний показник точності вже можна без змін використовувати при вирішенні задач машинного навчання різних типів в інших галузях. Використання USPI за межами медицини потребує довшого циклу підготовки: визначення категорій респондентів, формування вагових коефіцієнтів і т.і. Однак такі елементи відносяться більше до адаптації анкетування під конкретну групу користувачів, ніж до апробації самого методу. Індекс оцінки сприйняття системи користувачами можна використовувати через інші види функції приналежності, або взагалі перевизначити квантифікацію через апарат, відмінний від нечіткої логіки. Не менш важливим напрямом розвитку є інтеграція розроблених методів у вже існуючі підходи та методології з метою налагодження гармонічної взаємодії з усіма інструментами управління проєктом.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Everyone wants to do the model work, not the data work: Data Cascades in High-Stakes AI / Nithya Sambasivan та ін. New York, NY, USA : *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, 2021, Article 39, doi: 10.1145/3411764.3445518.
2. Software Engineering for Machine Learning: A Case Study / S. Amershi та ін., Montreal, QC, Canada : *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 2019, pp. 291-300, doi: 10.1109/ICSE-SEIP.2019.00042.
3. Ashmore R., Calinescu R., Paterson C. Assuring the Machine Learning Lifecycle. *ACM Computing Surveys*. 2021. Т. 54, № 5. С. 1–39. URL: <https://doi.org/10.1145/3453444> (дата звернення: 21.02.2026).
4. Project Management Institute. A Guide to the Project Management Body of Knowledge – Seventh Edition and The Standard for Project Management. Project Management Institute, 2021. 250 с.
5. Manifesto for Agile Software Development. URL: <https://agilemanifesto.org/iso/en/manifesto.html> (дата звернення: 01.03.2026).
6. Agile Project Management: Explained for Beginners. Kanban Software for Agile Project Management. URL: <https://businessmap.io/agile/project-management> (дата звернення: 25.03.2026).
7. Scrum J. Agile Project Management: The New Step by Step Guide to Learn the Kanban Process, Scrum and Lean Thinking, and Understanding Methodologies for Quality Control. Charlie Creative Lab, 2020.
8. Work B. SSADM Version 4: A User's Guide. *Software Engineering Journal*. 1995. Т. 10, № 4. С. 157. URL: <https://doi.org/10.1049/sej.1995.0020> (дата звернення: 05.02.2026).

9. Managing Successful Projects Prince2. 5-те вид. London: Stationery Office, The, 2009. 342 с.
10. Using PRINCE2 for AI & Data Projects: Governance and Risk Control. TSG Training. URL: <https://tsg-training.co.uk/blog/using-prince2-for-ai-data-projects/> (дата звернення: 20.02.2026).
11. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies / T. Greenhalgh та ін. *Journal of Medical Internet Research*. 2017. Т. 19, № 11. С. e367. URL: <https://doi.org/10.2196/jmir.8775> (дата звернення: 28.02.2026).
12. HL7. Health Level Seven International (HL7). URL: <https://www.hl7.org/about/index.cfm?ref=nav> (дата звернення: 16.01.2026).
13. HL7 FHIR. FHIR CI-Build. URL: <https://build.fhir.org/> (дата звернення: 26.09.2025).
14. General Data Protection Regulation (GDPR) Compliance Guidelines. GDPR.eu. URL: <https://gdpr.eu/> (дата звернення: 16.01.2026).
15. Final Regulations for Health Coverage Portability for Group Health Plans and Group Health Insurance Issuers Under HIPAA Titles I & IV. Federal Register. URL: <https://www.federalregister.gov/documents/2004/12/30/04-28112/final-regulations-for-health-coverage-portability-for-group-health-plans-and-group-health-insurance> (дата звернення: 16.01.2026).
16. Про захист персональних даних. *Офіційний вебпортал парламенту України*. URL: <https://zakon.rada.gov.ua/laws/show/2297-17#Text> (дата звернення: 16.01.2026).
17. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction / T. D. Shanafelt та ін. *Mayo Clinic Proceedings*. 2016. Т. 91, № 7. С. 836–848. URL: <https://doi.org/10.1016/j.mayocp.2016.05.007> (дата звернення: 24.01.2026).
18. The Role of Human-Centered Design in Healthcare Innovation: a Digital Health Equity Case Study / X. A. Levander та ін. *Journal of General Internal*

Medicine. 2023. URL: <https://doi.org/10.1007/s11606-023-08500-0> (дата звернення: 11.01.2026).

19. Брандт Н. М. Дослідження методів управління ІТ-проєктами в сфері надання медичних послуг: пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / Н. М. Брандт; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2024. – 68 с.

20. Задніпровський Д. Б. Дослідження та аналіз підходів до управління змінами на проєктах із розподіленими командами: пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / Д. Б. Задніпровський; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2024. – 102 с.

21. Казимов Латіф Бабір Дослідження методу Kanban для управління ІТ-проєктом медичної системи: пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / Казимов Латіф Бабір огли; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2024. – 90 с.

22. Красенков І. О. Дослідження моделей використання засобів штучного інтелекту для планування ІТ-проєктів: пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / І. О. Красенков; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2024. – 116 с.

23. Веретельніков Д. М. Дослідження методів оцінювання ефективності ІТ-проєктів для медичних лабораторій: пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / Д. М. Веретельніков; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2024. – 96 с.

24. Зінченко О.А, Пономаренко Л.Р. Особливості проектного менеджменту в закладах охорони здоров'я Features of project management in

healthcare institutions. Економіка та управління підприємствами. 2018. № 18. С. 123–126. URL: http://www.market-infr.od.ua/journals/2018/18_2018_ukr/22.pdf.

25. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare / М. І. Ahmed та ін. *Cureus*. 2023. URL: <https://doi.org/10.7759/cureus.46454> (дата звернення: 02.12.2025).

26. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies / Lushun Jiang та ін. *Journal of International Medical Research*. 2021. Т. 49(3). URL: <https://doi.org/10.1177/03000605211000157> (дата звернення: 02.12.2025).

27. Developing reliable AI tools for healthcare. *Google DeepMind*. URL: <https://deepmind.google/discover/blog/codoc-developing-reliable-ai-tools-for-healthcare/> (дата звернення: 01.12.2025).

28. EchoPixel. *EchoPixel Inc.* URL: <https://echopixeltech.com/about-echopixel-1> (дата звернення: 01.12.2025).

29. Healthcare Research Insights. Top 10 Healthcare Predictive Analytics Companies Setting the Standard. *LinkedIn*. URL: <https://www.linkedin.com/pulse/top-10-healthcare-predictive-analytics-5hkoc> (дата звернення: дата звернення: 11.12.2025).

30. Ronneberge O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer*. 2015. Т. 9351, № 234--241.

31. New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation / S. Surono та ін. *Emerging Science Journal*. 2023. Т. 7, № 2. С. 498 – 506. URL: <https://doi.org/10.28991/esj-2023-07-02-014> (дата звернення: 05.01.2026).

32. AI Platform, Deep Genomics. *Deep Genomics*. URL: <https://www.deepgenomics.com/AI-Platform/> (дата звернення: 07.12.2025).

33. Top 10 Real-World Machine Learning Healthcare Projects in 2024 | *SPD Technology. Software Product Development Company*. URL:

<https://spd.tech/machine-learning/machine-learning-in-healthcare/> (дата звернення: 27.12.2025).

34. Dietterich T. G. Ensemble Methods in Machine Learning. *Multiple Classifier Systems*. Berlin, Heidelberg, 2000. С. 1–15. URL: https://doi.org/10.1007/3-540-45014-9_1 (дата звернення: 17.01.2026).

35. Artificial intelligence in healthcare: transforming the practice of medicine / J. Bajwa та ін. *Future Healthcare Journal*. 2021. Т. 8, № 2. С. e188-e194. URL: <https://doi.org/10.7861/fhj.2021-0095> (дата звернення: 21.05.2025).

36. Sinead Butler. China reveals the world's first AI hospital with 42 AI doctors. indy100. URL: <https://www.indy100.com/science-tech/china-first-ai-hospital-42-ai-doctors> (дата звернення: 17.05.2025).

37. Ewen Callaway. Medical AI trained on whopping 57 million health records. *Nature*. URL: <https://www.nature.com/articles/d41586-025-01422-3> (дата звернення: 15.05.2025).

38. Hern A. TechScape: Can AI really help fix a healthcare system in crisis?. the Guardian. URL: <https://www.theguardian.com/technology/article/2024/jul/09/techscape-ai-nhs-healthcare-artificial-intelligence-cancer-care>. (дата звернення: 12.05.2025).

39. The Framework Convention on Artificial Intelligence. Council of Europe. URL: <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence> (дата звернення: 15.05.2025).

40. «The AI Index 2025 Annual Report» AI Index Steering Committee, Nestor Maslej, Loredana Fattorini, Raymond Perrault та ін. Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025.

41. CHAOS Report Beyond Infinity (digital version). *The Standish Group*. URL: <https://www.standishgroup.com/products/copy-of-chaos-report-beyond-infinity-digital-version> (дата звернення: 29.01.2025).

42. Between 70-85% of GenAI deployment efforts are failing to meet their desired ROI. NTT DATA. URL:

<https://www.nttdata.com/global/en/insights/focus/2024/between-70-85p-of-genai-deployment-efforts-are-failing> (дата звернення: 18.02.2026).

43. Why 80 Percent of Artificial Intelligence Projects Fail While Successful Implementations Achieve 383 Percent Return on Investment - SoftwareSeni. *SoftwareSeni*. URL: <https://www.softwareseni.com/why-80-percent-of-artificial-intelligence-projects-fail-while-successful-implementations-achieve-383-percent-return-on-investment/> (дата звернення: 20.02.2026).

44. James Ryseff, Brandon F. De Bruhl, Sydne J. Newberry. The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed. Avoiding the Anti-Patterns of AI. 2024. 20 с.

45. Robertson A., Bates D. W., Sheikh A. The rise and fall of England's National Programme for IT. *Journal of the Royal Society of Medicine*. 2011. Т. 104, № 11. С. 434–435. URL: <https://doi.org/10.1258/jrsm.2011.11k039> (дата звернення: 21.12.2025).

46. Strickland E. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*. 2019. Т. 56, № 4. С. 24–31. URL: <https://doi.org/10.1109/mspec.2019.8678513> (дата звернення: 29.12.2025).

47. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020. Т. 21, № 1.

48. Дерев'янюк І., Бондарев С. Практичне дослідження алгоритмів перехресної перевірки (Cross-validation) з використанням інструментарію Scikit-learn на Python. Scientific Collection «InterConf». 2022. №. 136. С. 382–390.

49. The C-index Multiverse / Begoña B. Sierra та ін. *Arxiv.org*. 2025. URL: <https://doi.org/10.48550/arXiv.2508.14821> (дата звернення: 20.02.2026).

50. The Concordance Index Decomposition – A Measure for a Deeper Understanding of Survival Prediction Models / A. Alabdallah та ін. *SSRN Electronic Journal*. 2022. URL: <https://doi.org/10.2139/ssrn.4024162> (дата звернення: 05.06.2024).

51. Pitfalls of the concordance index for survival outcomes / N. Hartman та ін. *Statistics in Medicine*. 2023. URL: <https://doi.org/10.1002/sim.9717> (дата звернення: 05.06.2024).
52. Moskalenko V., Kharchenko V. Resilience-aware MLOps for AI-based medical diagnostic system. *Frontiers in Public Health*. 2024. Т. 12. URL: <https://doi.org/10.3389/fpubh.2024.1342937> (дата звернення: 15.02.2026).
53. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods / S. S Band та ін. *Informatics in Medicine Unlocked*. 2023. Т. 40. С. 101286. URL: <https://doi.org/10.1016/j.imu.2023.101286> (дата звернення: 21.01.2026).
54. RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records / B. C. Kwon та ін. *IEEE Transactions on Visualization and Computer Graphics*. 2019. Т. 25, № 1. С. 299–309. URL: <https://doi.org/10.1109/tvcg.2018.2865027> (дата звернення: 22.01.2026).
55. NeuroXAI++: An Efficient X-AI Intensive Brain Cancer Detection and Localization / A. Rahman та ін. *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, м. Gazipur, Bangladesh, 16 – 17 черв. 2023 р. 2023. URL: <https://doi.org/10.1109/ncim59001.2023.10212818> (дата звернення: 17.01.2026).
56. Health and care workforce in Europe: time to act. UN City, Marmorvej 51, DK-2100 Copenhagen Ø, Denmark : World Health Organization Regional Office for Europe, 2022. 190 с.
57. State of the world’s nursing 2025: investing in education, jobs, leadership and service delivery. Geneva: World Health Organization; 2025. Licence: CC BY-NC-SA 3.0 IGO.
58. Ross Timothy J. Fuzzy logic with engineering applications. 3-тє вид. University of New Mexico, USA : John Wiley & Sons, Ltd, 2010. 585 с.
59. Zhong Z., Lin C.-M. Large-Scale Fuzzy Interconnected Control Systems Design and Analysis. IGI Global, 2017.

60. Analyzing data from a fuzzy rating scale-based questionnaire. A case study / M. Gil та ін. *Psicothema*. 2015. Т. 2, № 27. С. 182–191. URL: <https://doi.org/10.7334/psicothema2014.268> (дата звернення: 02.04.2026).

61. Дерев'янка І. І. Дослідження методів обчислювального інтелекту для детекції аномалій на медичних зображеннях : пояснювальна записка до кваліфікаційної роботи здобувача вищої освіти на другому (магістерському) рівні, спеціальність 122 Комп'ютерні науки / І. І. Дерев'янка ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків, 2025. – 82 с.

62. Derevianko Iryna, Bondarev Serhii, Zolotukhin Oleg. Research on unsupervised anomaly detection in spinal mri: advancing autoencoder-based reconstruction methods. *IX International Scientific and Practical Conference "Development of innovation systems: trends, challenges, prospects"*. 2025. С. 51–60.

63. Nilsson, J., & Akenine-Möller, T. (2020). Understanding SSIM. ArXiv, abs/2006.13846.

64. Приватне підприємство «МЕДІНФОСЕРВІС». Медична інформаційна система «МедІнфоСервіс». URL: <https://www.infomed.ck.ua/download/polyclinic-stacionar/4.3.52>.

65. Fuzzy logic toolkit for SciPy scikit-fuzzy. Версія 0.5.0. 2024. URL: <https://pypi.org/project/scikit-fuzzy/>.

66. Бондарев С. Проблеми управління та виконання проєктів з розробки інформаційних систем штучного інтелекту в сфері охорони здоров'я. *Інноваційна наука: пошук відповідей на виклики сучасності* / chair Ю. Олександров. 2025. С. 392–396.

67. Bondarev S., Derevianko I. Requirements for the development of methods for analysis and quality evaluation of intelligent information systems in healthcare. *Modern Perspectives on Science and Economic Progress: Collection of Scientific Papers with Proceedings of the 3rd International Scientific and Practical Conference*. International Scientific Unity. Vilnius, Lithuania. April 29 – May 1, 2026. 204-208 p.

68. Bondarev S., Derevianko I. Development of the composite accuracy index as a method for evaluating artificial intelligence models, Technology and Economics: Collection of Scientific Papers with Proceedings of the 7th International Scientific and Practical Conference. International Scientific Unity. Luxembourg, Luxembourg. May 6-8, 2026. 245-249 p.

69. Bondarev S. Development of the user system perception index as a method for evaluating medical information systems. XI International Scientific Conference *Scientific space: current issues, achievements and innovations* . 2026. С. 328–337.

70. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи другого (магістерського) рівня вищої освіти за освітньо-науковою програмою «Управління проектами в галузі інформаційних технологій» / Упоряд.: Петров К.Е., Левикін В.М., Чалий С.Ф., Євланов М.В., Міхнов Д.К., Міхнова А.В., Чала О.В. – Харків: ХНУРЕ, 2024. – 24 с.