
МЕТОДОЛОГІЧНІ ЗАСАДИ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ ТА КЛІНІЧНОЇ ЕФЕКТИВНОСТІ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ В ОХОРОНІ ЗДОРОВ'Я

Головенко М. Я., Ларіонов В. Б.

DOI <https://doi.org/10.30525/978-9934-26-637-9-19>

ВСТУП

Ми переживаємо четверту промислову революцію, де підприємства та галузі промисловості більше зосереджуються на автоматизації та обладнанні для ефективного виробництва та оптимізації робочої сили. Вона встановлює нові способи виконання завдань (у більш оптимізованих та ефективних процедурах), включаючи те, як виробники промисловості перетворюють товари, виробляють, дотримуються стратегій ремонту та встановлення нової виробничої лінії, а також як інші працівники у відповідних секторах реалізують плани, вдосконалюють та розподіляють все¹. Галузь охорони здоров'я стала також ареною для трансформаційного потенціалу ШІ, оскільки цей технологічний прогрес є важливим для революціонування догляду за пацієнтами та адміністративних операцій завдяки використанню величезних поздовжніх даних про пацієнтів². ШІ охоплює спектр технологій, що вміщують машинне навчання (МН), обробку природної мови (НЛП), експертні системи на основі правил (РПСП), фізичних роботів та роботизовану автоматизацію процесів, кожна з яких пропонує унікальні можливості від прогнозного моделювання та виявлення захворювань до підвищення хірургічної точності та автоматизації адміністративних завдань^{3,4}. Інтеграція ШІ в охорону

¹ Adib Bin Rashid, Ashfakul Karim Kausik. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*. 2024. Vol. 7. Article 100277. DOI: 10.1016/j.hybadv.2024.100277.

² Ghafur S., van Dael J., Leis M., Darzi A., Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. *Lancet Digital Health*. 2020. Vol. 2, No. 9. P. e444–e446. DOI: 10.1016/S2589-7500(20)30161-8.

³ Davenport T., Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019. Vol. 6, No. 2. P. 94–98. DOI: 10.7861/futurehosp.6-2-9.

⁴ Vial A., Stirling D., Field M., Ros M., Ritz C., Carolan M., Holloway L., Miller A.A. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Translational Cancer Research*. 2018. Vol. 7, No. 3. P. 803–816. DOI: 10.21037/tcr.2018.05.02.

здоров'я обіцяє підвищену точність діагностики, обґрунтоване прийняття рішень та оптимізоване планування лікування, тим самим потенційно зменшуючи кількість медичних помилок та покращуючи результати лікування пацієнтів⁵. Регуляторний ландшафт розвивається, щоб йти в ногу з технологічним прогресом, прагнучи створити надійні системи управління, які забезпечують відповідальне використання ШІ в закладах охорони здоров'я. Крім того, поява попередньо навчених моделей великих мов програмування, прикладами яких є такі моделі, як BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-Transformer) та їх варіанти, ще більше розширила можливості ШІ в охороні здоров'я⁶. Такі моделі використовують величезні обсяги текстових даних, що дозволяє виконувати завдання, починаючи від покращення клінічної документації до автоматизованого узагальнення медичної літератури.

Однак, поряд з цими перспективними розробками, впровадження ШІ в охорону здоров'я супроводжується значними етичними та регуляторними проблемами, які потребують ретельного розгляду⁷. Проблеми варіюються від захисту конфіденційності даних пацієнтів до вирішення алгоритмічних упереджень, які можуть увічнювати нерівність у результатах охорони здоров'я⁸.

При використанні ШІ в будь-якій галузі науки та техніки поняття «користь – ризик» є ключовими, бо вони визначають, чи варто впроваджувати певну технологію або продукт, зважаючи не лише на потенційну вигоду, але й на можливі небезпеки⁹. У ШІ користь, це автоматизація, підвищення продуктивності, нові наукові відкриття, персоналізовані сервіси. Ризиком є упередженість алгоритмів, порушення конфіденційності, втрати роботи, неконтрольовані наслідки використання. неправильні рішення (помилкова медична порада, аварія автомобіля), соціальні наслідки (упередженість,

⁵ Adib Bin Rashid, Ashfakul Karim Kausik. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*. 2024. Vol. 7. Article 100277. DOI: 10.1016/j.hybadv.2024.100277.

⁶ Gaviria-Valencia S., Murphy S.P., Kaggal V.C., McBane II R.D., Rooke T.W., Chaudhry R., Alzate-Aguirre M., Arruda-Olson A.M. Near real-time natural language processing for the extraction of abdominal aortic aneurysm diagnoses from radiology reports: algorithm development and validation study. *JMIR Medical Informatics*. 2023. Vol. 11. Article e40964. DOI: 10.2196/40964.

⁷ Quinn T.P., Senadeera M., Jacobs S., Coghlan S., Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association*. 2021. Vol. 28, No. 4. P. 890–894. DOI: 10.1093/jamia/ocaa268.

⁸ Albahri A.S., Duhaim A.M., Fadhel M.A., Alnoor A., Baqer N.S., Alzubaidi L., Albahri O.S., Alamoodi A.H., Bai J., Salhi A., Santamaria J., Ouyang C., Gupta A., Gu Y., Deveci M. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. *Information Fusion*. 2023. Vol. 96. P. 156–191. DOI: 10.1016/j.inffus.2023.03.008.

⁹ Fulton R., Fulton D., Hayes N., Kaplan S. The Transformation Risk-Benefit Model of Artificial Intelligence: balancing risks and benefits through practical solutions and use cases. *International Journal of Artificial Intelligence and Applications (IJAA)*. 2024. Vol. 15, No. 2 (March). P. 1–22.

маніпуляція, підміна фактів), зловживання (військове застосування). Модель має застосовуватися лише тоді, коли очікувана користь для суспільства переважає ймовірні ризики, а ризики можна мінімізувати технічними й етичними бар'єрами.

В медичній практиці відомо, що жоден лікарський засіб не є абсолютно безпечним. Навіть прості добре перевірені препарати, як-от парацетамол чи аспірин, можуть спричиняти побічні ефекти¹⁰. Тому лікар завжди зважує наскільки очікувана користь від лікування перевищує можливі ризики. Це важливо, оскільки дозволяє керувати безпечністю (уникнення надмірного шкоди від ліків) та ефективністю (вибір засобу, який реально допоможе в конкретній ситуації) та індивідуалізація (різні пацієнти мають різні стани здоров'я, алергії, супутні хвороби, генетичні особливості). Отже, препарат допускається на ринок лише тоді, коли очікувана користь перевищує потенційний ризик. При недооцінці ризиків під час використання недостатньо провірених препаратів виникають серйозні загрози для людств, як це сталося з талідомідом. У 1953 р. в Німеччині компанія «Chemie Grünenthal» синтезувала нову молекулу, яка в медичній практиці отримала назву талідомід і планувалась як протисудомний засіб, але згодом виявили седативний ефект. У 1957 р препарат впровадили в медичну практику під торговою назвою «Contergan» і рекламували як абсолютно безпечне снодійне і засіб від нудоти та «ранкової слабкості» у вагітних^{11,12}. Використання препарату в широких масштабах сприяло формуванню вроджених вад розвитку плоду, що призвело до того, що близько 10 000 дітей у 46 країнах народилися з важкими каліцтвами, укорочення або відсутність кінцівок, деформації вух, очей, внутрішніх органів¹³. Це викликало величезний резонанс у медичній спільноті, недовіру до фармкомпаній, посиленню вимог до контролю за лікарськими засобами.

Трагедія з талідомідом стала можливою через брак належних доклінічних та клінічних досліджень, слабку регуляцію та агресивний маркетинг. Зокрема, не проводились належні випробування на тваринах щодо тератогенності (впливу на розвиток ембріона), оскільки лабораторні практики були набагато менш суворими, ніж зараз. Дослідження безпеки

¹⁰ Головенко М.Я. «Філософія» фармацевтичних інновацій. *Вісник Національної академії наук України*. 2012. № 3. С. 59–66. DOI: 10.15407/vism2012.03.059.

¹¹ Vanderbes J. Wonder Drug: The Secret History of Thalidomide in America and Its Hidden Victims. *New York : Random House*, 2023. 432 p.

¹² Vargesson N. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Research Part C: Embryo Today*. 2015. Vol. 105, No. 2. P. 140–156. DOI: 10.1002/bdrc.21096.

¹³ Johnson M., Stokes R., Arndt T. The Thalidomide Catastrophe: How it Happened, Who was Responsible and Why the Search for Justice Continues After More Than Six Decades. *London : Onwards and Upwards Publishers*, 2018. 263 p.

у вагітних жінок практично не проводились, бо вважалось, що плацента «захищає» плід від більшості речовин. Препарат широко рекламувався як «нешкідливий» і призначався навіть без рецепту в деяких країнах.

Завдяки позиції Френсіс Келсі, лікарки з FDA, у США талідомід не був зареєстрований¹⁴. Вона відмовилась дати дозвіл через недостатність даних про безпеку і її рішення зберегло десятки тисяч життів і стало символом значення суворої регуляції. Цей приклад слугує постійним нагадуванням, що навіть «безпечні» ліки можуть мати катастрофічні наслідки без належного контролю, що необхідно враховувати і при використанні ШІ.

В роботі обґрунтовано та проведено порівняльний аналіз інтегральних показників співвідношення користі та ризику у фармакології та ШІ медичного призначення, з акцентом на можливості адаптації концепції терапевтичного індексу до розробки аналогічного індексу для інформаційних систем. Задля цього використано наукову літературу з клінічної фармації, зокрема монографії та керівні настанови щодо визначення терапевтичного індексу та терапевтичного вікна лікарських засобів. Враховано основні позиції документів міжнародних та національних регуляторних органів, що регламентують критерії безпеки та ефективності фармакотерапії та нормативні ініціативи у сфері ШІ (AI Act, рекомендації OECD, принципи етики ШІ від IEEE).

Застосовано порівняльно-аналітичний метод для аналізу аналогій між терапевтичним індексом (ТІ) у фармакології та інформаційним індексом (ІІ) у ШІ. Систематизацію та класифікацію використано для визначення основних компонентів користі та ризику в обох сферах, а формалізацію для побудови моделі розрахунку інформаційного індексу за формулою:

$$II = \frac{K}{K + P_{II}},$$

де K інтегральний показник користі, а P інтегральний показник ризику.

У такий спосіб, дослідження носить концептуально-теоретичний характер і сприяє формуванню узагальненої моделі оцінювання безпеки та ефективності ШІ.

1. Інтеграція систем ШІ в клінічну діяльність: ризиків та викликів використання

Управління з контролю за продуктами харчування та лікарськими засобами США (FDA) схвалило понад тисячу медичних пристроїв на базі ШІ. Враховуючи широкий спектр потенційних застосувань ШІ в медицині

¹⁴ Warsh C.K. Frances Oldham Kelsey, the FDA, and the Battle Against Thalidomide. *The Amazon Book Review*. 2024. 424 p.

та зростаючу увагу до впровадження різних інформаційних моделей в повсякденну практику, вкрай важливо розуміти, які типи пристроїв наразі дозволено для клінічного використання, і як цей напрямок розвивається з часом. Поточні системи класифікації FDA для медичних пристроїв, коди продуктів та їх класи, надають широку характеристику, яка побудована за застосуванням і охоплює ключові аспекти використання ШІ¹⁵.

1. Медична візуалізація (*Viz.ai, Aidoc, Arterys, Zebra Medical Vision*) використовується для виявлення патологій (КТ/МРТ/Рентген/мамографії, пріоритетизація ургентних проблем, автоматична сегментація).

2. Офтальмологічні авто-скрінінги (*IDx-DR, AEYE/Optomed*), з автономною функцією (без лікаря) або допоміжна діагностика ретинопатії, глаукоми тощо.

3. Кардіологія (*AliveCor KardiaMobile*), виконує аналіз ЕКГ / мобільний моніторинг, де функцією є виявлення фібриляції передсердь, аритмій, виявлення показників ризику інфаркту.

4. Цифрова патологія (*Paige (Paige Prostate/PanCancer Detect)*), допомагає патологам у знаходженні підозрілих ознак (рак/метастази) на віртуальних зрізах (WSI).

5. Підтримка в отриманні зображень та кероване сканування (*Caption Health*), допомагає операторам (навіть нетренованим) отримувати діагностично якісні зображення (наприклад, УЗД).

6. Прогнозні алгоритми та моніторинг пацієнтів, які інтегруються в ЕМІС/ліжкові монітори. (регуляторний статус варіює). Їх функція – раннє виявлення сепсису, прогноз погіршення, попередження про ризики.

7. Геноміка та біоінформатика (деякі є інструментами досліджень, інші проходять регуляторну валідацію). Функція класифікація варіантів, пріоритетизація мішеней; як-от алгоритми для підвищення якості секвенування.

Усі вони мають програмне забезпечення Software as a Medical Device (SaMD), незалежно від того, що воно є частиною апарату¹⁶.

Оскільки системи охорони здоров'я покладаються на комерційні алгоритми прогнозування для виявлення та допомоги пацієнтам зі складними потребами, особливого значення набувають проблеми, які виникають в наслідок використання ШІ в медичній практиці. Перш за все, це стосується так званих «галюцинацій» ШІ (генерація інформації, яка виглядає правдоподібно, але насправді є неправильною, вигаданою

¹⁵ Singh R., Bapna M., Diab A.R., Ruiz E.S., William. How AI is used in FDA-authorized medical devices: a taxonomy across 1,016 authorizations. *NPJ Digital Medicine*. 2025. Vol. 8, No. 1. Article 388. DOI: 10.1038/s41746-025-01800-1.

¹⁶ Alowais S.A., Alghamdi S.S., Alsuhebany N. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*. 2023. Vol. 23. Article 689. DOI: 10.1186/s12909-023-04698-z.

або не підтвердженою жодним джерелом), які вважаються одними із найнебезпечніших¹⁷. Так GPT-3 згенерував 178 посилань, із яких 69 були некоректними чи взагалі неіснуючими, а ще 28 не мали DOI. Для лікаря чи науковця це виглядає правдоподібно, але насправді може бути повною вигадкою. Клініцист може покласти на неіснуюче дослідження і зробити хибний висновок, і як наслідок призначити неправильне лікування. ChatGPT, особливо LLM, взагалі можуть «уявляти» протоколи, яких не існує, або плутати дозування. Наприклад, модель може написати, що «лікування можна почати з певної дози», яка насправді небезпечна. Звідси неправильна терапія, передозування ліків або навпаки, недостатнє лікування, що погіршує стан пацієнта. Не виключені пропуски важливої інформації, в результаті чого моделі дають неповну картину, не враховують винятки або рідкісні стани, а у медицині навіть «малоймовірне» може бути життєво важливим. Пацієнт з нетиповими симптомами отримає стандартну відповідь із помилковим діагнозом та втрату часу для лікування. Часом модель формулює відповідь дуже впевнено і переконливо, навіть коли помиляється, а для користувача без медичної освіти це звучить як «авторитетна думка». Пацієнти можуть самостійно почати лікуватися «за порадою ChatGPT» без консультації з лікарем. Використання некоректних порад від ШІ без перевірки лікарем може призвести до судових позовів. Лікарі ризикують професійною відповідальністю, якщо рішення ухвалюються «на основі чат-бота» і як наслідок, втрата довіри до медичних установ, репутаційні та фінансові збитки. Є випадки, коли пацієнти намагаються «обійти» лікаря й отримати рецепт чи схему лікування напряму від ШІ. У темних веб-спільнотах чат-боти вже використовувалися для створення фейкових медичних довідок чи «домашніх інструкцій» з виготовлення препаратів. Виникає небезпека самолікування, використання підроблених даних, ризик для життя. Є також небезпека впливу на освіту та підготовку лікарів і якщо студент-медик привчиться довіряти моделям без перевірки, формується культура поверхневих знань. Це знижує критичне мислення, здатність працювати з джерелами та приймати обґрунтовані клінічні рішення, що приводить до довгострокового погіршення якості медичних кадрів.

Чому це стається? Моделі не мають знань у класичному сенсі і вони прогнозують слова на основі статистики. Якщо у тренувальних даних не вистачає актуальної або перевіреної медичної інформації то відповідь буде неточною. Вони не перевіряють факти, а лише «імітують

¹⁷ Özer M. Potential Benefits and Risks of Artificial Intelligence in Education and Medicine. BUEFAD Bartın University Journal of Education Faculty. 2025. Vol. 13, No. 2. P. 232–244. DOI: 10.14686/buefad.1416087.

переконливість». Ось кілька документальних випадків та досліджень, що ілюструють, наскільки небезпечними можуть бути такі медичні «галюцинації» ШІ. Слідуючи порадам ChatGPT, 60-річний чоловік у Нью-Йорку, значно скоротив споживання натрію, що призвело до небезпечної гіпонатріємії¹⁸. Пацієнта госпіталізували на три тижні, але він одужав. Інший випадок: чоловік з неправдивої поради замінив кухонну сіль на бромід натрію, яка є токсичною речовиною, що призвело до отруєння, психозу та тривалого стаціонарного лікування¹⁹. У медичній статті²⁰ поєднувались два не пов'язаних анатомічних поняття (basilar artery та basal ganglia), яких не існує. Помилку не виявили відразу, але згодом її виправили без пояснення. Подібні помилки можуть привести до хибного клінічного тлумачення та діагнозу. Дуже часто спостерігаються небезпечні медичні відповіді. Аналіз 888 відповідей на медичні запитання (жінки, педіатрія, внутрішні хвороби) показав, що від 5% (Claude) до 13% (GPT-4o, Llama3) відповідей були прямо небезпечними. Загалом 21–43% відповідей вважались проблемними²¹. Крім того, модель JAMA PediatricsChatGPT у 2024 році надала понад 80% педіатричних кейсів некоректних діагнозів. У 339 випадках рентгенологічних описів GPT-3.5 і GPT-4 дали «галюцинаційні» посилання у 39,9% і 14,3% відповідно, а фальшиві твердження 16,2% та 4,7%²².

Спостерігались²³ токсичні або небезпечні поради, де: 21–43% відповідей були «проблемні», а 5–13% ,взагалі небезпечні. Зокрема, радили вводити воду немовлятам (небезпечно), використовувати ефірну олію біля очей, прикладати чайне дерево до вуха, небезпечно годувати грудьми в певних випадках. Існують задокументовані випадки²⁴ коли великі мовні моделі

¹⁸ ChatGPT advice lands 60-year-old man in hospital; the reason will surprise you. *The Times of India*. 2025. Режим доступу: <https://timesofindia.indiatimes.com/technology/tech-news/chatgpt-advice-lands-60-year-old-man-in-hospital-the-reason-will-surprise-you/articleshow/123200430.cms>.

¹⁹ Man Went to ChatGPT for Health Advice. What He Did Next Led to Poisoning and Psychosis. *People Magazine*. 2025. Режим доступу: <https://people.com/man-chatgpt-health-advice-led-to-poisoning-psychosis-11789649>.

²⁰ Field H. Google's healthcare AI made up a body part – what happens when doctors don't notice?. *The Verge*. 2025. Режим доступу: <https://www.theverge.com/health/718049/google-med-gemini-basilar-ganglia-paper-typo-hallucination>.

²¹ Draelos R. Large language models provide unsafe answers to patient-posed medical questions. *Computation and Language Preprint*. 2025. arXiv:2507.18905. DOI: 10.48550/arXiv.2507.18905.

²² Chelli M., et al. Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *Journal of Medical Internet Research*. 2024. Vol. 26. Article e53164. DOI: 10.2196/53164.

²³ Anderson M. Doctors' Study Finds 5–13% of Chatbot Medical Advice Is Dangerous or Unsafe. *Anderson's Angle*. 28.07.2025. Режим доступу: <https://www.unite.ai/doctors-study-finds-5-13-of-chatbot-medical-advice-is-dangerous-or-unsafe/>.

²⁴ Thompson D. AI Chatbots Easily Misled By Fake Medical Info. *Mount Sinai Health System*. 06.08.2025. Режим доступу: <https://www.usnews.com/news/health-news/articles/2025-08-08/ai-chatbots-easily-misled-by-fake-medical-info>.

(наприклад, ChatGPT) «вигадували» захворювання, яких насправді не існує. Користувачі ставили медичні запитання, і ШІ у відповідь створював абсолютно фіктивні синдроми з «науковими» назвами, описував симптоми, можливі причини й навіть методи лікування. Приклади таких вигадок: “*Casper-Lew Syndrome*” та “*Helkand Disease*”. Жоден із цих «синдромів» не зустрічається у медичній літературі, клінічних протоколах чи наукових базах даних (PubMed, WHO ICD тощо). Це чиста вигадка алгоритму.

Подібно до фармакології, де поряд із терапевтичним ефектом існують побічні реакції, у ШІ проявляються технологічні «побічні ефекти». Серед них: упередженість алгоритмів, що відтворює соціальні стереотипи; вразливість до маніпуляцій (наприклад, отруєння даних чи атаки на моделі); галюцинації ШІ, коли система продукує неправдиву або нелогічну інформацію, видаючи її за достовірну; нестабільність масштабування, коли ефективність моделі знижується при застосуванні в нових або критичних умовах; ризик витоку даних, що особливо небезпечно у медицині та фінансовому секторі.

Регулювання ефективного використання ШІ в медицині є багатограним процесом, який поєднує принципи та конкретні норми, експертні технічні методи і ефективне законодавство. Сучасні дискусії показують, що базові моделі ШІ вимагають окремої уваги; що ризик-орієнтоване законодавство має свої обмеження; і що прозорість, відповідальність і аудит мають ключове значення. Для подальшого розвитку потрібна взаємодія між науковцями, практиками та законодавцями, з оцінкою результатів та коригуванням технічних підходів у швидко змінному технологічному середовищі. Отже, так само як і у фармакології, де лікарський засіб може стати одночасно порятунком і джерелом небезпеки, ШІ поєднує в собі потенціал користі та значні ризики, які потребують системного управління та регулювання.

2. Національні та міжнародні регуляторні органи, наділені повноваженнями сертифікувати медичні ШІ

Більшість країн ділять підхід щодо регуляції використання ШІ в медичній практиці на: 1. принципово-орієнтований/добровільний (стандарти, рекомендації), 2. секторний (правила для конкретних галузей, медицина) і 3. загальнонаціональні/обов'язкові правила для «високоризикових» застосувань²⁵. Баланс між сприянням інноваціям і мінімізацією ризиків для людей знаходиться в основі всіх підходів.

²⁵ Chomutare T., Tejedor M., Svenning T.O., Marco-Ruiz L. et al. Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. *International Journal of Environmental Research and Public Health*. 2022. Vol. 19. Article 16359. DOI: 10.3390/ijerph192316359.

У США FDA оцінює та видає дозвіл для медичних пристроїв (510(k), De Novo, PMA, Breakthrough). Публікує настанови по AI/ML (Action Plan, GMLP, Transparency, PCCP).

Національні компетентні органи ЄС, як-от Notified Bodies (через MDR) проводять сертифікацію, а регуляторна координація здійснюється через національні органи та MDCG/European Commission.

Велика Британія за допомогою агентства MHRA регулює медичні вироби та випускає керівні документи щодо контролю AIaMD.

Health Canada контролює реєстрацію медичних пристроїв (SaMD guidance, GMLP співпраця).

У Японії системи PMDA/MHLW погоджують і проводять технічну оцінку SaMD/AI-виробів.

У Австралії TGA стандартизує SaMD/AI як медичні вироби і вимагає реєстрації в ARTG та публікує настанови по класифікації ПЗ.

WHO, IMDRF не сертифікують продукти, але дають міжнародні рекомендації та технічні стандарти, що впливають на національні процедури (IMDRF є важливий для уніфікації SaMD підходів)

В усіх випадках регулятори вимагають визначати ризики на всіх етапах життєвого циклу, а також проводити постринковий моніторинг/нагляд для «високоризикових» систем ШІ.

Україна приєдналась до системної роботи у цьому напрямі. Кабінет Міністрів України розпорядженням № 1556-р схвалив Концепцію розвитку ШІ в Україні, та виділив пріоритетні напрями та завдання розвитку ШІ для побудови конкурентоспроможної національної економіки, розвитку соціальної сфери та удосконалення системи публічного управління²⁶.

3. Структуровані методики оцінювання впливу та ризиків

На сьогодні не існує єдиної універсальної «формули», яка б кількісно обчислювала користь і ризики ШІ як просту функцію. Натомість у міжнародній практиці використовують структуровані методики оцінювання цих показників, закріплених у стандартах і керівних документах. Основні підходи, які вважаються «офіційними» або широко визнаними представлено в таблиці 1.

Ці стандарти не дають однієї «формули користі», але пропонують кількісні та якісні метрики (KPI, шкали 1–5, вартісні оцінки), які в різних країнах мають свої особливості²⁷.

²⁶ Про схвалення Концепції розвитку штучного інтелекту в Україні. *Офіційний вебпортал парламенту України*. 2020. № 1556-р. Режим доступу: <https://zakon.rada.gov.ua/laws/show/1556-2020-r#n8>, вільний.

²⁷ Study on Health Data, Digital Health and Artificial Intelligence in Healthcare. 5. Evaluation of Article 14 of Directive 2011/24/EU. Брюссель: Європейська комісія, 2023. Режим доступу: https://health.ec.europa.eu/system/files/2022-05/crossborder_evaluation-dir201124eu_study_annexes_en.pdf.

Стандарти ISO/IEC (призначення та ключові показники)

Документ	Призначення	Ключові показники
ISO/IEC 23894:2023	Управління ризиками ШІ	Ідентифікація, аналіз, оцінювання, обробка ризиків. Використовує загальні формули ймовірність × вплив.
ISO/IEC 42001:2023	Система менеджменту ШІ	Встановлює вимоги до процесу оцінки користі/ризик у організації.
ISO 31000:2018 (базовий для будь-яких ризиків)		Ризик = Ймовірність × Наслідок (Risk = P × Impact).

1. ШІ Act та інші інструменти (ALTAI) у ЄС використовують анкетно-бальну систему, яка оцінює вплив на права людини, соціальну та економічну користь та ймовірність і масштаб шкоди. Формалізованими розрахунками ризику є: Ризик=Р(події)×С(шкоди), де С – кількісна або грошова оцінка наслідків.

2. OECD рекомендують багатовимірні оцінки (технічні, економічні, соціальні, екологічні). Пропонуються багатокритеріальні моделі (multi-criteria decision analysis, MCDA) для балансу «користь – ризик».

3. США(RMF, 2023) вміщує чотири етапи: Map → Measure → Manage → Govern. Кількісний ризик оцінюється через Expected Loss E[L]:

$$E[L] = \sum_i P_i \times \text{Impact}_i$$

користь часто виражають у вигляді очікуваної економічної цінності (ROI, cost-benefit).

4. Типові формули, що застосовуються в цих підходах

Очікувана користність:

$$E[L] = \sum_i P_i \times U_i$$

де U_i – користь/вигода сценарію і.

Очікуваний ризик:

$$R = \sum_i P_i \times S_i$$

де S_i – тяжкість (severity) наслідків.

Ці формули гнучкі, де ваги P_i , U_i , S_i визначаються експертами залежно від контексту (галузь, регіон, тип ШІ).

Практичні приклади інструментів і метрик, які використовують лікарні, дослідницькі центри та розробники медичних рішень на основі

ШІ для оцінки користі й ризиків представлено в таблиці 2. Вона дозволяє зіставити потенційну користь і можливі ризики у єдиному форматі²⁸. Кожен параметр має власну шкалу вимірювання, що спрощує подальший розрахунок інтегрального ризик – профілю та допомагає вирішити, чи впроваджувати технологію, чи потрібні додаткові заходи контролю.

Таблиця 2

Шаблон матриці оцінки користі та ризиків.

Параметр	Призначення	Тип/одиниця	Коментарі до інтерпретації
Очікувана користь	Кількісно відображає позитивний ефект від технології: наприклад, скорочення часу діагностики, зменшення витрат, підвищення пропускної здатності.	% зміни, хв/пацієнта, економія \$ тощо.	У прикладі 30% скорочення часу діагностики, підтвержене пілотним проектом.
Клінічний вплив	Показує, як технологія впливає на якість медичного рішення: чутливість, специфічність, точність, позитивна прогностична цінність.	%, або комбінація метрик (наприклад 95% / 92%).	У прикладі 95% чутливість і 92% специфічність порівняно з поточним стандартом.
Ймовірність збою	Частота критичних помилок або відмов системи, що можуть призвести до шкоди пацієнтові чи бізнесу.	% або кількість на 1000 випадків.	0,5% означає, що у середньому 5 випадків на 1000 можуть спричинити критичну помилку.
Тяжкість наслідків	Оцінка потенційної шкоди для пацієнта або процесу у разі збою. Зазвичай 5-бальна шкала: 1 мінімальна шкода, 5 смерть або незворотна втрата функції.	Бал 1–5 (можна розширювати до 1–10).	Значення 4 = «серйозна шкода, що потребує невідкладного лікування».
Ризик (P×Impact)	Інтегральний показник: добуток імовірності збою (P) на тяжкість наслідків (Impact).	0–25 (або інша шкала залежно від множників).	У прикладі: 0,5% (~0,5) × 4 ≈ 2 → категорія «низький». Дозволяє ранжувати ризики та визначати пріоритети для їхнього зменшення.
Регуляторні вимоги	Відповідність правовим і галузевим нормам (HIPAA, GDPR, MDR, FDA).	Так/Ні або перерахування стандартів.	Порушення цих вимог може нівелювати всю користь, навіть якщо інші ризики низькі.

²⁸ Croke K., Thapa G.K., Aryal A., Pokhrel S., Kruk M.E. The politics of health system quality: How to ignite demand. *BMJ*. 2023. Vol. 383. Article e076792. DOI: 10.1136/bmj-2023-076792.

Такий шаблон матриці оцінки користі та ризиків є зручним інструментом кількісного і якісного аналізу для медичних технологій при використанні ШІ. Він допомагає одночасно продемонструвати очікувану клінічну користь, ймовірність небажаних подій та регуляторну готовність, що є критично важливим перед впровадженням будь-якої інновації у медичну практику.

Ключовим показником користі ШІ є КРІ (Key Performance Indicator), кількісний або якісний індикатор, який показує, наскільки ефективно окремий процес досягає визначених цілей²⁹. Основними його рисами є вимірюваність, яка має чітку метрику (% , години тощо). Зв'язок із цілями, що прямо відображає прогрес у досягненні стратегічної мети. Часову прив'язку, яка вимірюється за певний період (день, місяць, квартал).

У контексті охорони здоров'я «чутливість» може означати здатність КРІ реагувати на зміни в системі та швидко помічати проблеми чи відхилення. У деяких випадках це може бути статистичний термін як-от «чутливість діагностичного тесту» тобто наскільки тест виявляє хворобу, при її наявності.

До її важливих характеристик відносять високу реактивність(КРІ повинен змінюватися коли змінюється стан справ, низьку затримку даних та адекватну частоту їх збору.

Ефективність, у цьому випадку, визначається наскільки добре система чи процес досягає поставлених цілей та очікуваних результатів і чи мають вони бажаний ефект.

Щодо безпеки пацієнтів, то вона повина відвертати шкоду пацієнту під час надання медичної допомоги, а показниками є ті, що відображають частоту та серйозність помилок, інцидентів, виробничого несправностей тощо.

Один і той же КРІ може стосуватись кількох категорій, але важливо, щоб цей показник був валідним, досяжним, вимірюваним й часово прив'язаний. Задля його визначення необхідно встановити цільові значення (наприклад, чутливість $\geq 90\%$, ROI $\geq 15\%$) та налаштувати автоматичний збір даних (інтеграція з PACS та EMR). До того ж, проводити моніторинг у часі (щомісячно чи квартално та порівнювальний аналіз показників до та після оновлення моделі. Надавати регуляторний звіт, який необхідно подавати до FDA, EMA чи національних органів охорони здоров'я.

²⁹ Croke K., Thapa G.K., Aryal A., Pokhrel S., Kruk M.E. The politics of health system quality: How to ignite demand. *BMJ*. 2023. Vol. 383. Article e076792. DOI: 10.1136/bmj-2023-076792.

4. Співвідношення «користь–ризик» у фармакотерапії та медичних ШІ: спільне й відмінне

У медичній практиці і лікарські засоби, і системи ШІ медичного призначення оцінюють через співвідношення «користь ризик», але зміст цього підходу має спільні риси й суттєві відмінності. Спільним для обох систем, досягнення абезпечена, користі для здоров'я пацієнта, яка перевищувала потенційну шкоду. Потрібна доказова база (клінічні дослідження, валідаційні дослідження, постмаркетинговий нагляд). Пріоритетом безпеки пацієнта є інформована згода, мінімізація ризиків, моніторинг небажаних ефектів/помилки. Співвідношення користі й ризику може змінюватися з часом (нові дані, умови застосування, індивідуальні особливості). У той же час існують значні відмінності між лікарськими засобами і ШІ (табл. 3).

Таблиця 3

Відмінності принципу «користь–ризик» у лікарських засобів та ШІ

Критерій	Лікарські засоби	ШІ медичного призначення
Природа продукту	Хімічна/біологічна речовина з фармакологічною дією.	Алгоритм/програмне забезпечення, що обробляє дані та генерує рекомендації чи рішення.
Тип ризиків	Токсичність, побічні реакції, лікарські взаємодії, передозування.	Алгоритмічні помилки, упередженість даних, збій програмного забезпечення, витік персональних даних, кіберзагрози.
Користь	Лікувальний, профілактичний чи діагностичний ефект, доведений у клінічних дослідженнях.	Підвищення точності діагностики, швидкість обробки даних, персоналізація лікування, підтримка клінічних рішень.
Контроль якості	Виробнича стандартизація, стабільність складу, біоеквівалентність.	Безперервне оновлення моделей, перевірка алгоритмів, контроль версій, необхідність повторної валідації після донавчання.
Регуляторне поле	Довготривалі етапи доклінічних і клінічних випробувань; післяреєстраційний фармаконагляд.	Оцінка програмного забезпечення як медичного виробу (SaMD), кібербезпека, аудит алгоритмів, контроль «drift» (зміни поведінки моделі).

Отже, спільним є принцип «користь має переважати ризик» та необхідність постійного моніторингу. Відмінне у характері користі й ризиків є те, що для ліків це насамперед фармакологічні ефекти на організм, а для ШІ якість даних, алгоритмічні помилки та інформаційна безпека. У практиці це означає, що лікарський засіб потребує ретельних клінічних випробувань, а ШІ, не лише медичної, а й технічної та етичної перевірки протягом усього життєвого циклу.

У фармакології оцінюють співвідношення між ефективністю та токсичністю терапевтичним індексом (ТІ), і чим ширший цей показник, тим безпечніший і передбачуваніший препарат.

Ми спробували перенести цю логіку на ШІ, де започаткували умовний термін «інформаційний індекс» (ІІ), який показує баланс між ефективністю (точність, релевантність, швидкість, продуктивність) та безпечністю (відсутність небажаних або шкідливих наслідків, як-от галюцинацій, упередження, витік конфіденційних даних, маніпуляцій):



На ранніх стадіях доклінічних випробувань лікарських засобів класичним фізіологічним показником є ТІ, де ED_{50} доза, яка викликає ефект у 50% популяції експериментальних тварин, TD_{50} – доза токсичності/летальності у 50% тварин.

Інформаційний індекс (ІІ) побудований на імовірностях і якісних метриках ШІ: К (К) користь системи ШІ, це інтегральна оцінка позитивного ефекту. Р (Р) ризик, який є агрегованою оцінкою можливих негативних наслідків. К + Р (К + Р) є сумою користі та ризику, тобто загальна «вага» усіх наслідків роботи системи. Таким чином індекс показує частку користі від загального впливу і на відміну від ТІ це не фізіологія, а оцінка ймовірностей результату.

Обидві формули описують баланс користі та ризику для прийняття рішення про безпечне застосування ліків або ШІ. Це безрозмірні індекси, отримані шляхом відношення двох величин, що робить їх придатними для порівняння між різними системами. Чим більше значення тим вищий запас безпеки чи перевага користі. Разом з тим, для обох параметрів потрібні надійні емпіричні або експертні оцінки складових (доза або показники користі/ризiku).

Таким чином, загальна наша ідея полягала в тому, що б використати ТІ та ІІ для кількісного відображення співвідношення користі та ризику, з урахуванням головної різниці тобто природи показників. ТІ використовує об'єктивні біомедичні дози, а ІІ комплексно та гнучко оцінює ефективність й загрози для цифрових систем.

Параметр ПІ залежить від таких чинників як вид тварини, варіабельність пацієнтів, тяжкість хвороби, можливість моніторингу. Для ПІ це критичність домену (медицина vs побут), масштаби експозиції (скільки користувачів), незворотність наслідків, довіра до даних оцінки.

Інтерпретація порогів ПІ: 10 – «широке вікно» (безпечний), 1–3 – вузький (небезпечний, потрібен моніторинг), <1 – неприйнятний. ПІ: 85% – «зелений» рівень, 70–85% – «жовтий», з контролем, <70% – «червоний» (високий ризик або низька користь). Даніми для розрахунків ПІ є клінічні та доклінічні експерименти, біомаркери, концентрації у плазмі, для ПІ це тести запитів, систематичний збір, моніторинг та аналіз помилок, що виникають у програмному продукті після його запуску у виробничому середовищі, експертні оцінки тяжкості. Детально зазначена проблема розглянута у монографії³⁰. Розрахунки фармакологічних показників (терапевтичний індекс, широта, інтервал та фактор надійної безпеки) представлено нами у експериментальній статті³¹ що характеризує ефективність та безпечність інноваційного анальгетичного засобу «пропоксазепам».

Отже ПІ для ПІІІ може бути мірою співвідношення користі до ризику, подібно до ПІ у фармакології. Це дає змогу об'єктивніше оцінювати надійність систем у різних сферах.

Для конкретного застосування ПІІІ у ми використали параметр ПІ, який є нормованою мірою співвідношення користі до ризику

$$P = \frac{B'}{B'+R'} \in [0,1]$$

де: B' та R' є скоригованими показниками ефективності та ризику, відповідно, а P у цьому випадку розраховується у відсотках ($P_{100}=100*P$). Базова ефективність $B \in [0,1]$ є зважене середнє метрик якості: $B = \sum_k \omega_k^{(B)} m_k^{(B)} / \sum_k \omega_k^{(B)} = 1$, де усі значення $m_{k(B)}$ нормуються через мін-макс або ізотонну калібровку.

Компонентами ризику є відношення $R = \sum_i \omega^{(R)} (p_i s_i) / \sum_i \omega^{(R)} = 1$, де для кожного класу i : $p_i \in [0,1]$ – імовірність настання, а $s_i \in [0,1]$ – тяжкість наслідку (нормована шкала шкоди). Найбільш типовими класами і егалюінації, фактичні помилки, шкідливий контент, порушення приватності, маніпулятивність та невідповідність нормам.

Ризики й користь мають різну вагу залежно від сфери застосування та масштабу і показником M контексту ризику слугує:

³⁰ Goodman & Gilman's The Pharmacological Basis of Therapeutics. 13th ed. New York : McGraw-Hill, 2018. 2108 p.

³¹ N.Ya. Golovenko1, S.A. Andronati 1, V.B. Larionov1, A.S. Reder. A pharmacological profile of propoxazepam – a new antiepileptic substance. *Dopov. Nac. akad. nauk Ukr.* 2018. No 12 P. 93–100. DOI: 10.15407/dopovidi2018.12.093.

$$M = (1 + \alpha C)(1 + \beta \log_{10} \frac{E}{E_0})(1 + \gamma \rho),$$

де E – експозиція (інтерацій/день), E_0 референтний показник (наприклад, 100), $\rho \in [0,1]$ – неповоротність (0 легко скасовується, а при 1 виникають незворотні наслідки), $\alpha, \beta, \gamma \in$ коефіцієнтами чутливості (0.5, 0.3, 0.2 відповідно). Тоді:

$$M = (1 + \alpha C)(1 + \beta \log_{10} \frac{E}{E_0})(1 + \gamma \rho),$$

де $c \in (0,1]$ – довіра до оцінок (менший c консервативно «збільшує» ризик), ε – константа стабільності (наприклад, 10^{-6}).

Критичність домену становить 0,1 для $C \in [0,1]$, тоді коли він використовується для побутових потрб, освіти – 0,3, фінансів – 0,6, медицині – 1,0.

Нами проведено розрахунок параметрів Π ($B / (B+R)$) для різних напрямів застосування ШІ у медицині, результати яких представлено в табл. 4.

Таблиця 4.

Порівняльна характеристика Π різних моделей ШІ медичного призначення.

Сфера застосування ШІ	Компоненти користі (B)	Компоненти ризику (R)	Приблизний розрахунок Π (B / (B+R))	Інтерпретація
Медична діагностика (радіологія, аналіз зображень)	Висока точність у виявленні патологій, швидкість аналізу, зниження навантаження на лікаря	Хибнопозитивні/ хибнонегативні діагнози, відсутність інтерпретованості	0.6–0.75	Значна користь, але потребує перевірки лікарем («human-in-the-loop»).
Прогнозування захворювань (генетичні ризики, рання діагностика)	Рання профілактика, індивідуалізація лікування, підвищення шансів на успіх	Хибні прогнози → тривожність, непотрібне лікування або втрата шансів на реальне втручання	0.55–0.7	Помірний індекс: корисно як додатковий інструмент, не як остаточне рішення.
Розробка нових ліків (drug discovery)	Прискорення пошуку молекул, економія ресурсів, відкриття нових терапій	Галюцинації в моделях, хибні біомаркери, юридичні й етичні питання використання даних	0.7–0.85	Високий потенціал користі, але на стадії валідації потрібні клінічні випробування

Продовження таблиці 4

Роботи-асистенти в хірургії (AI-guided surgery)	Підвищення точності рухів, зменшення часу операції, менша інвазивність	Системні збої, втрата контролю, проблема відповідальності	0.5–0.65	Критично вузьке «інформаційне вікно»: обов'язкова присутність хірурга.
Телемедицина та чат-боти для пацієнтів	Швидкі поради, зменшення навантаження на лікарів, доступність	Галюцинації, неправильні рекомендації, ризик самолікування	0.6–0.7	Корисно для «першої лінії», але не для остаточних діагнозів.
Аналіз медичних записів (NLP для історій хвороб)	Швидкий пошук інформації, підтримка лікаря у прийнятті рішень	Витік персональних даних, неправильна класифікація	0.65–0.8	Високий П за умов дотримання GDPR/захисту даних.

На прикладі моделі «Медична діагностика (радіологія, аналіз зображень)», яка представлена в таблиці 4, розрахунок параметрів П ($B / (B+R)$) склав «0,6–0,75». Зазначимо, що це не одна точна цифра, а діапазон можливого П, обчислений за запропонованою умовною формулою:

$$П = B / (B + R),$$

де B – сумарна оцінка користі, R – сумарна оцінка ризику, обидва показники вимірюються у відносних балах (наприклад, 0–1 або 0–10). Користь (B): для радіологічних алгоритмів можна орієнтовно задати 6–8 балів (висока точність, швидкість, економія часу). Ризик (R): хибні висновки, низька пояснюваність – приблизно 2–4 бали.

Щоб отримати діапазон:

Нижня межа

$$B=6, R=4 \Rightarrow 6/4+6=0.60$$

Верхня межа

$$B=8, R=3 \Rightarrow 8/8+3 \approx 0.73 (\approx 0.75 \text{ заокруглено})$$

Що це означає: 0,6: користь становить близько 60% від сукупного «пакету» користь+ризик. 0,75: за кращих умов (вищі показники точності, нижчий ризик) користь може складати до 75%. Це орієнтовні експертні оцінки, а не результат клінічних випробувань. У реальному дослідженні потрібно: формалізувати вагові коефіцієнти для кожного компоненту користі та ризику, провести статистичну валідацію (наприклад, на тестових

Як у цьому випадку вираховується бали? У прикладі з П бали не беруться з готової шкали, а присвоюються експертно за кожним

компонентом користі й ризику. Це умовний, але досить поширений підхід, коли немає єдиної нормативної методики.

Як можна підходити до цього у медичному ШІ. Спочатку перелічують підкритерії для користі й ризику. Користь (В): точність, чутливість, специфічність, швидкість аналізу, зменшення навантаження на персонал, покращення доступу до діагностики. Ризик (R) це хибнопозитивні та хибнонегативні результати, відсутність інтерпретованості, можливі збої або кіберзагрози, юридичні/етичні наслідки. Кожен підкритерій експерти (лікарі, інженери, етики) оцінюють від 0 до 10, де: 0–2 є мінімальний вплив, 3–5 помірний, 6–8 значний, а 9–10 критичний / винятково високий

Приміром для розрахунку користі медичної діагностики, радіологія, аналіз зображень (табл. 4) приблизний розрахунок Π ($B / (B+R)$) склав 0.6–0.75 було використован наступні показники точність та бали (в дужках) – 95% (8), швидкість, секунди (7), зменшення навантаження (6), покращення доступу (5). Сума користі В у цьому випадку становить 26 (8+7+6+5). У випадку розрахунку ризику використовують хибнопозитивні / хибнонегативні наслідки (5), відсутність інтерпретованості (4), кіберзагрози (3), юридичні та етичні аспекти (2). Сума ризику R складає 14 (5+4+3+2).

Щоб привести до шкали 0–10 необхідно розрахувати середнє ($B = 26/4 = 6.5$; $R = 14/4 = 3.5$), або просто використовувати суму, бо у формулі відношення $B/(B+R)$ масштаб скорочується.

Розрахунок Π

$\Pi = B / (B+R) = 6.5 / (6.5+3.5) \approx 0.65$ Це вписується у діапазон **0.6–0.75**.

Ключові моментами розрахунків є те, що бали, які відносяться до експертної оцінки, інколи підтверджені статистикою (наприклад, точність 95% може відповідати 8/10). Ваги можна коригувати (наприклад, точність важить більше, ніж швидкість). Пропонована методика є гнучкою і може змінюватися залежно від сфери (діагностика, роботизована хірургія тощо).

Отже, бали це результат багатфакторного експертного оцінювання, яке зводиться до єдиної числової шкали для введення у формулу $B/(B+R)$.

Зважаючи на те, що у високоризикових сферах (хірургія, діагностика) індекс вужчий, то потрібен постійний контроль і підтвердження людиною. У низькоризикових сферах (аналіз записів, розробка ліків на ранніх етапах) індекс вищий тому можна дозволити більшу автоматизацію. Необхідно враховувати той факт, що як і у фармакології, деякі застосування ШІ мають «вузьке інформаційне вікно», тобто навіть невелика помилка може призвести до критичних наслідків.

Порівняльний підхід до аналізу «користь–ризик» для ліків і медичних ШІ-систем не лише поглиблює розуміння безпеки, а й закладає основу

для створення єдиних стандартів контролю. У майбутньому це дозволить впроваджувати штучний інтелект у медицину з такою ж довірою, як і сертифіковані фармацевтичні препарати.

ВИСНОВКИ

Таким чином, порівняльний підхід до аналізу «користь–ризик» для ліків і медичних ШІ-систем не лише поглиблює розуміння безпеки, а й закладає основу для створення єдиних стандартів контролю. У майбутньому це дозволить впроваджувати ШІ у медицину з такою ж довірою, як і сертифіковані фармацевтичні препарати. У фармакології перед виходом на ринок будь-який лікарський засіб проходить поетапну систему клінічних випробувань. Цей процес включає доклінічні дослідження на лабораторному та модельному рівні, а далі декілька фаз клінічних випробувань на добровольцях і пацієнтах. Така модель дозволяє поступово перевіряти безпеку та ефективність препарату, зменшуючи ризики для суспільства. Подібний підхід може бути застосований і для ШІ, де кожна нова система повинна пройти етапне тестування, яке дозволить оцінити її користь і ризики до масового використання. Наприклад, попереднє тестування (аналог доклінічної стадії) може відбуватися у закритих середовищах, а наступні фази у пілотних проєктах з обмеженою кількістю користувачів, перш ніж система буде впроваджена у широкій практиці. Очікується створення міжнародних норм, що встановлюватимуть порогові значення індексів користі–ризик для різних класів медичних алгоритмів, подібно до вимог FDA чи EMA для ліків. На відміну від фіксованих клінічних показників, системи ШІ потребуватимуть **безперервного аудиту** з автоматичним оновленням індексу при зміні даних або моделі. Майбутні індекси враховуватимуть не лише технічні метрики (точність, чутливість), а й показники впливу на якість лікування, задоволеність пацієнтів і соціально-економічні наслідки. Водночас необхідно чітко визначення відповідальності за помилки ШІ, механізми відшкодування шкоди та захисту приватності, що інтегруватиметься у модель оцінки користь–ризик.

АНОТАЦІЯ

У статті розглянуто методологічні підходи до оцінки безпеки та клінічної ефективності систем штучного інтелекту (ШІ) в охороні здоров'я. Проведено порівняльний аналіз концепції «користь–ризик» у фармакології та медичних ШІ-системах із метою адаптації терапевтичного індексу до інформаційних технологій. Запропоновано модель інформаційного індексу як формалізованого інструменту для оцінювання співвідношення очікуваної користі та потенційних ризиків. Враховано сучасні регуляторні

вимоги, технічні показники ефективності, етичні аспекти та соціальні наслідки. Обґрунтовано необхідність етапного тестування ШІ-систем за аналогією до клінічних випробувань лікарських засобів. Окреслено перспективу розробки міжнародних нормативів для визначення допустимих індексів користі–ризик. Наголошено на важливості безперервного аудиту, оновлення моделей та встановлення механізмів відповідальності. Запропонований підхід сприяє формуванню уніфікованих стандартів для безпечного впровадження ШІ в медичну практику.

Література

1. Adib Bin Rashid, Ashfakul Karim Kausik. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*. 2024. Vol. 7. Article 100277. DOI: 10.1016/j.hybadv.2024.100277.
2. Ghafur S., van Dael J., Leis M., Darzi A., Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. *Lancet Digital Health*. 2020. Vol. 2, No. 9. P. e444–e446. DOI: 10.1016/S2589-7500(20)30161-8.
3. Davenport T., Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019. Vol. 6, No. 2. P. 94–98. DOI: 10.7861/futurehosp.6-2-9.
4. Vial A., Stirling D., Field M., Ros M., Ritz C., Carolan M., Holloway L., Miller A.A. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Translational Cancer Research*. 2018. Vol. 7, No. 3. P. 803–816. DOI: 10.21037/tcr.2018.05.02.
5. Gaviria-Valencia S., Murphy S.P., Kaggal V.C., McBane II R.D., Rooke T.W., Chaudhry R., Alzate-Aguirre M., Arruda-Olson A.M. Near real-time natural language processing for the extraction of abdominal aortic aneurysm diagnoses from radiology reports: algorithm development and validation study. *JMIR Medical Informatics*. 2023. Vol. 11. Article e40964. DOI: 10.2196/40964.
6. Quinn T.P., Senadeera M., Jacobs S., Coghlan S., Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association*. 2021. Vol. 28, No. 4. P. 890–894. DOI: 10.1093/jamia/ocaa268.
7. Albahri A.S., Duhaim A.M., Fadhel M.A., Alnoor A., Baqer N.S., Alzubaidi L., Albahri O.S., Alamoodi A.H., Bai J., Salhi A., Santamaria J., Ouyang C., Gupta A., Gu Y., Deveci M. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. *Information Fusion*. 2023. Vol. 96. P. 156–191. DOI: 10.1016/j.inffus.2023.03.008.
8. Fulton R., Fulton D., Hayes N., Kaplan S. The Transformation Risk-Benefit Model of Artificial Intelligence: balancing risks and benefits through

practical solutions and use cases. *International Journal of Artificial Intelligence and Applications (IJAI)*. 2024. Vol. 15, No. 2 (March). P. 1–22.

9. Головенко М.Я. «Філософія» фармацевтичних інновацій. *Вісник Національної академії наук України*. 2012. № 3. С. 59–66. DOI: 10.15407/visn2012.03.059.

10. Vanderbes J. Wonder Drug: The Secret History of Thalidomide in America and Its Hidden Victims. *New York : Random House*, 2023. 432 p.

11. Vargesson N. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Research Part C: Embryo Today*. 2015. Vol. 105, No. 2. P. 140–156. DOI: 10.1002/bdrc.21096.

12. Johnson M., Stokes R., Arndt T. The Thalidomide Catastrophe: How it Happened, Who was Responsible and Why the Search for Justice Continues After More Than Six Decades. *London : Onwards and Upwards Publishers*, 2018. 263 p.

13. Warsh C.K. Frances Oldham Kelsey, the FDA, and the Battle Against Thalidomide. *The Amazon Book Review*. 2024. 424 p.

14. Singh R., Bapna M., Diab A.R., Ruiz E.S., William. How AI is used in FDA-authorized medical devices: a taxonomy across 1,016 authorizations. *NPJ Digital Medicine*. 2025. Vol. 8, No. 1. Article 388. DOI: 10.1038/s41746-025-01800-1.

15. Alowais S.A., Alghamdi S.S., Alsuhebany N. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*. 2023. Vol. 23. Article 689. DOI: 10.1186/s12909-023-04698-z.

16. Özer M. Potential Benefits and Risks of Artificial Intelligence in Education and Medicine. *BUEFAD Bartın University Journal of Education Faculty*. 2025. Vol. 13, No. 2. P. 232–244. DOI: 10.14686/buefad.1416087.

17. ChatGPT advice lands 60-year-old man in hospital; the reason will surprise you. *The Times of India*. 2025. Режим доступу: <https://timesofindia.indiatimes.com/technology/tech-news/chatgpt-advice-lands-60-year-old-man-in-hospital-the-reason-will-surprise-you/articleshow/123200430.cms>, вільний.

18. Man Went to ChatGPT for Health Advice. What He Did Next Led to Poisoning and Psychosis. *People Magazine*. 2025. Режим доступу: <https://people.com/man-chatgpt-health-advice-led-to-poisoning-psychosis-11789649>.

19. Field H. Google’s healthcare AI made up a body part – what happens when doctors don’t notice? *The Verge*. 2025. Режим доступу: <https://www.theverge.com/health/718049/google-med-gemini-basilar-ganglia-paper-tyro-hallucination>, вільний.

20. Draelos R. Large language models provide unsafe answers to patient-posed medical questions. *Computation and Language Preprint*. 2025. arXiv:2507.18905. DOI: 10.48550/arXiv.2507.18905.

21. Chelli M., et al. Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *Journal of Medical Internet Research*. 2024. Vol. 26. Article e53164. DOI: 10.2196/53164.
22. Anderson M. Doctors' Study Finds 5–13% of Chatbot Medical Advice Is Dangerous or Unsafe. *Anderson's Angle*. 28.07.2025. Режим доступу: <https://www.unite.ai/doctors-study-finds-5-13-of-chatbot-medical-advice-is-dangerous-or-unsafe/>, вільний.
23. Thompson D. AI Chatbots Easily Misled By Fake Medical Info. *Mount Sinai Health System*. 06.08.2025. Режим доступу: <https://www.usnews.com/news/health-news/articles/2025-08-08/ai-chatbots-easily-misled-by-fake-medical-info>.
24. Chomutare T., Tejedor M., Svenning T.O., Marco-Ruiz L. et al. Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. *International Journal of Environmental Research and Public Health*. 2022. Vol. 19. Article 16359. DOI: 10.3390/ijerph192316359.
25. Про схвалення Концепції розвитку штучного інтелекту в Україні. Офіційний вебпортал парламенту України. 2020. № 1556-р. Режим доступу: <https://zakon.rada.gov.ua/laws/show/1556-2020-p#n8>.
26. Study on Health Data, Digital Health and Artificial Intelligence in Healthcare. 5. Evaluation of Article 14 of Directive 2011/24/EU. Брюссель: Європейська комісія, 2023. Режим доступу: https://health.ec.europa.eu/system/files/2022-05/crossborder_evaluation-dir201124eu_study_annexes_en.pdf.
27. Croke K., Thapa G.K., Aryal A., Pokhrel S., Kruk M.E. The politics of health system quality: How to ignite demand. *BMJ*. 2023. Vol. 383. Article e076792. DOI: 10.1136/bmj-2023-076792.
28. Goodman & Gilman's The Pharmacological Basis of Therapeutics. 13th ed. New York : McGraw-Hill, 2018. 2108 p.
29. Головенко М.Я., Андронаті С.А., Ларіонов В.Б., Редер А.С. A pharmacological profile of пророхазепам – a new antiepileptic substance. *Доповіді Національної академії наук України*. 2018. № 12. С. 93–100. DOI: 10.15407/dopovidi2018.12.093.

Information about the authors:

Golovenko Mykola Yakovych,

<https://orcid.org/0000-0003-1485-128X>

Doctor of Biological Sciences, Professor,

Academician of the National Academy of Medical Sciences of Ukraine,

Head of Department of Biomedicine

A. V. Bogatsky Physico-Chemical Institute
of the National Academy of Sciences of Ukraine

86, Liustdorfska road, Odesa, 65084, Ukraine

Larionov Vitalii Borysovyh,

<https://orcid.org/0000-0003-2678-4264>

Doctor of Biological Sciences, Senior Researcher,

Head of the Laboratory of molecular pharmacology and medicine of the

Department of biomedicine,

A. V. Bogatsky Physico-Chemical Institute
of the National Academy of Sciences of Ukraine

86, Liustdorfska road, Odesa, 65084, Ukraine