DOI: 10.31359/2079-6242-2024-47-50
УДК 343.98

Г. К. Авдєєва, кандидат юридичних наук, старший науковий співробітник, провідний науковий співробітник Науково-дослідного інституту вивчення проблем злочинності
імені академіка В. В. Сташиса
Національної академії правових наук України

РОЗВИТОК ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ І ДОТРИМАННЯ ПРАВ ЛЮДИНИ: ПРОБЛЕМИ ЗАБЕЗПЕЧЕННЯ БАЛАНСУ

Постановка проблеми. Технології штучного інтелекту (ШІ) багато в чому допомагають реалізації прав людини, надають їй можливість покращити та полегшити своє життя. Завдяки своїй потужності, масштабу і швидкості оброблення інформації системи ШІ підвищують ефективність і результативність роботи, замінюють працівників на небезпечних ділянках та виконують за них виснажливу роботу.

Швидкий прогрес у галузі ШІ призводить до підвищення рівня забезпечення прав осіб на освіту, достатній рівень життя, охорону здоров'я, соціальну та медичну допомогу та ін. Однак є багато прикладів стосовно помилок у роботі систем ШІ, що призводять до порушення фундаментальних прав і свобод людини[1]. Тому дослідження причин виникнення помилок у роботі систем ШІ та визначення балансу між розвитком технологій ШІ і дострименням прав людини є актуальними.

Аналіз останніх досліджень і публікацій. Проблеми визначення напрямів використання ШІ у різних сферах діяльності людини та їх правового регулювання досліджувались багатьма вітчизняними і закордонними дослідниками, а саме: М. Карчевським та О. Радутним (2023)[2], В. Шевчуком (2023)[3], В. Шепітьком (2024)[4], Турутою О. (2022)[5], Д. Леслі (2021)[6], Алі Ф. Кабола (2022)[7], Б. Коном (2023)[8] та ін. Незважаючи на значну кількість публікацій щодо зазначених питань, проблеми можливого порушення прав людини внаслідок некоректного використання технологій ШІ досліджено ще недостатньо.

М**етою статті** є дослідження причин виникнення помилок у роботі систем ШІ, виокремлення напрямів негативного впливу технологій ШІ на права людини та формулювання пропозицій щодо забезпечення балансу між розвитком технологій ШІ і збереженням прав людини.

---

[1] Гайворонська Б. 'Тріумф та загрози штучного інтелекту – як нейромережі впливають на наше життя і як вони законодавчо регулюються' (Cityhost.UA, 04.09.2023) <https://cityhost.ua/uk/blog/triumf-ta-zagrozi-shtuchnogo-intelektu-yak-neyromerezhi-vplivayut-na-nashe-zhittya-i-yak-ce-zakonodavcho-regulyu-tsya.html#the-top-5-major-failures-of-AI≥ (дата звернення: 01.05.2024).

[2] Karchevskyi Mykola, Radutniy Oleksandr. 'Artificial Intelligence in Ukrainian Traditional Categories of Criminal Law' (2023) 1 (19) Вісник Асоціації кримінального права України 1–25 https://doi.org/10.21564/2311–9640.2023.19.281123

[3] Шевчук В. М. 'Штучний інтелект в криміналістиці : проблеми, можливості, перспективи' *Інформаційне забезпечення розслідування злочинів*: *мат-ли X Міжнар. круглого столу* (Одеса, 19.05.2023) 89–97

[4] Shepitko, V., Shepitko, M., Latysh, K., Kapustina, M., & Demidova, E.. 'Artificial intelligence in crime counteraction: From legal regulation to implementation' (2024) 7(1) Social and Legal Studios 135–144. https://doi.org/10.32518/sals1.2024.135

[5] Турута О. В., Турута О. П. 'Штучний інтелект крізь призму фундаментальних прав людини' (2022) 71 Науковий вісник Ужгородського національного університету https://doi.org/10.24144/2307–3322.2022.71.7

[6] Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. A*rtificial intelligence, human rights, democracy, and the rule of law: a primer* (The Council of Europe., 2021) <https://www.researchgate.net/publication/350808360_Artificial_intelligence_human_rights_democracy_and_the_rule_of_law_a_primer≥ (дата звернення: 02.05.2024).

[7] Ali Faghiri Kabol. 'The Use Of Artificial Intelligence In The Criminal Justice System (A Comparative Study)' (Webology, 2022) <https:// www.researchgate.net/publication/365027297_The_Use_Of_Artificial_Intelligence_In_The_Criminal_Justice_System_A_Comparative_Study≥ (дата звернення: 02.05.2024).

[8] Kohn B., Pieper F-U. 'AI regulation about the world' (Taylor Wessing, 09.05.2023) <https://www.taylorwessing.com/en/interface/2023/ai---are-we-getting-the-balance-between-regulation-and-innovation-right/ai-regulation-around-the-world≥ (дата звернення: 02.05.2024).

Виклад основного матеріалу. Останні досягнення з приводу розроблення моделей машинного навчання (генеративного ШІ) показали неймовірні результати завдяки тому, що цей процес здійснюється безперервно (24/7) шляхом оброблення величезних масивів інформації і проходить швидко через відсутність у них людських емоцій, втоми, голоду та інших перепон для найбільш повного сприйняття інформації. Однією з основних ідей машинного навчання (ML) є те, що комп'ютер можна навчити автоматизувати процеси виконання завдань, які є вкрай складними або неможливими для розв'язання їх людиною[1]. Таке навчання являє собою поступовий аналіз інформації (вихідних даних) із різних джерел, виявлення певних закономірностей (подібності або відмінності) у даних для прогнозування результату у конкретному випадку[2]. Оскільки модуль ШІ навчається за заздалегідь заданим алгоритмом і не здатен оцінювати якість «навчальних» матеріалів, то від якості і повноти інформації, що аналізується, та надійності її джерел залежить якість «навчання» системи ШІ й правильність рішень, які нею приймаються.

Дослідники Великої Британії виокремили низку ризиків високого рівня внаслідок використання систем ШІ, серед яких найнебезпечнішим є ризики фізичної шкоди людям і майну та шкоди психічному здоров'ю[3]. Вченими доведено, що технології ШІ, які використовуються у правозастосовній діяльності, можуть спричинити або посилити дискримінацію, що призводить до порушення права на справедливе правосуддя[4].

ШІ ставить під загрозу свободу вираження поглядів, асоціацій та зібрань через збирання й аналіз інформації про діяльність людей (онлайн та офлайн), оброблення їх реєстраційних даних та аналіз дій під час відвідування веб-сайтів та соціальних мереж. Соціальні мережі використовують ШІ для аналізу індивідуальних вподобань користувачів із метою пропонування їм певних публікацій і реклами для підтримки взаємодії з ними. ШІ також застосовується для створення фальшивих фото- та відео-матеріалів, підроблених облікових записів та іншого сфабрикованого контенту, який може заважати людині критично мислити. Такі маніпуляції можуть призводити до формування спотворених уявлень та навіть до формування антидемократичних або насильницьких світоглядів. Безвідповідальне використання алгоритмів ШІ, які здійснюють моніторинг думок і поглядів людей, може призвести до формування у них «потрібних» думок та придушення інакомислення.

Системи ШІ можуть отримати доступ до величезних обсягів даних про людей і обробляти їх з неймовірною швидкістю. Штучний інтелект може робити прогнози щодо поведінки, душевного стану та особистостей людини. Системи відеспостереження з функцією розпізнавання обличчя можуть стежити за людьми без їх згоди, що порушуватиме права людини на особисте приватне життя та змушувати її змінювати свою поведінку при підозрі, що за нею спостерігають або аналізують її дії.

Науковці Ессекського університету дійшли висновку, що технології ШІ містять потенційну загрозу для конфіденційності та людської гідності через те, що частота помилок системи відеоспостереження британської поліції, яка за допомогою ШІ здійснює розпізнавання обличчя, становить 81%[5]. Тобто програма може помилково ідентифікувати як підозрюваних чотирьох із п'яти невинних осіб.

Американська спілка громадянських свобод виявила, що технологія розпізнавання обличчя Amazon частіше позначала людей із темним кольором шкіри як злочинців. Дослідники Техаського університету в Далласі встановили, що розроблені у Східній Азії системи ШІ краще розпізнають осіб східноазіатської зовнішності, а створені в західних країнах алгоритми точніше функціонують при виявленні осіб європейської зовнішності[6].

---

[1] Кларк Е. 'Підручник з машинного навчання для початківців: що таке, основи ML' (Guru99, 2024) <https://www.guru99.com/uk/machine-learning-tutorial.html> (дата звернення: 02.05.2024)

[2] Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. *Artificial intelligence, human rights, democracy, and the rule of law: a primer* (The Council of Europe, 2021) 7 <https://www.researchgate.net/publication/350808360_Artificial_intelligence_human_rights_democracy_and_the_rule_of_law_a_primer≥ (дата звернення: 02.05.2024)

[3] DeGrave, A. J., Janizek, J. D. & Lee, SI. 'AI for radiographic COVID-19 detection selects shortcuts over signal' (2021) 3 Nat Mach Intell 610–619 https://doi.org/10.1038/s42256-021-00338-7

[4] Micklitz H-W, Pollicino O, Reichman A, Simoncini A, Sartor G, De Gregorio G, eds. In: Constitutional Challenges in the Algorithmic Society. Cambridge University Press; 2021. 342 p. https://doi.org/10.1017/9781108914857

[5] England R. 'UK police's facial recognition system has an 81 percent error rate' (Engadget, 04.07.2019) <https://www.engadget.com/2019-07-04-uk-met-facial-recognition-failure-rate.html≥ (дата звернення: 05.06.2024)

[6] Daws R. 'AI is sentencing people based on their 'risk' assessment' (ALNEWS, 22.01.2019). <https://www.artificialintelligence-news.com/2019/01/22/ai-sentencing-people-risk-assessment/≥ (дата звернення: 02.05.2024)

Дослідники університету Пердью (США) довели, що більшість відповідей найбільш популярної системи ШІ ChatGPT є неправильними. За результатами аналізу 517 відповідей системи ШІ ChatGPT на запитання спеціалістів ІТ-сфери, ці спеціалісти встановили, що у 52% випадків відповіді, хоча і були структурованими і детальними, але містили помилки або некоректні рекомендації (дезінформацію)[1]. Наведене дослідження свідчить про необхідність застосування критичного підходу до використання ШІ під час вирішення важливих завдань, а відповіді системи ШІ ChatGPT краще використовувати для отримання посилань на надійне джерело інформації або для генерування та обґрунтування нових ідей і напрямів дослідження, щодо яких остаточне рішення приймає людина (фахівець).

Системи ШІ використовуються в правоохоронних органах різних країн для оцінки осіб, які мають судимість. Ці технології ґрунтуються на алгоритмах машинного навчання та призначені для прогнозування ризику повторного вчинення злочинів. Вони впливають на рішення суддів та співробітників органів пробації. Наприклад, у Великій Британії система оцінки ризику Oasys (Offender Assessment System) допомагає зменшити вплив людської упередженості при прийнятті рішень, але її алгоритми залишаються «чорною скринькою» через відсутність доступу до даних та недостатню прозорість[2]. Аналогічні системи використовуються у США та Україні[3].

Використання завідомо упереджених систем може призвести до помилкових рішень щодо застосування засобів примусу до невинних осіб та неправильного оцінювання «ризику» повторного вчинення злочину (повторного засудження протягом двох років після звільнення) під час прийняття судового рішення[4]. За результатами аналізу «профілю» обвинуваченого система ШІ здійснює оцінку такого ризику, а суддя на її основі визначає суворість судового вироку та приймає рішення щодо обрання особі запобіжного заходу (домашній арешт, тримання під вартою та ін.).

Алгоритмічне оцінювання «ризику» рецидиву злочину здійснюється на основі даних, отриманих під час допиту або опитування підозрюваних співробітниками правоохоронних органів, які можуть по-різному оцінювати показники представників різних етнічних груп щодо їх імпульсивності, характеру поведінки, соціально-побутових умов, матеріального стану та ін. Такі дані не завжди є об'єктивними через можливі несвідомі упередження, що є результатом культурних відмінностей окремих етнічних груп (люди з різних культурних середовищ можуть вважати одну й ту саму емоційну поведінку прийнятною або неприйнятною). Для можливості оскарження судових рішень алгоритм систем ШІ, які здійснюють оцінювання «ризику» рецидиву злочину, має бути доступним та зрозумілим для підсудного.

Помилки в оцінюванні «ризику» рецидиву злочину можуть привести до трагічних наслідків. Зокрема, у Великій Британії офіцер пробації зібрав дані про злочинця та за допомогою системи ШІ Oasys здійснив таке оцінювання щодо 32-річного Деміена Бендалла для судді, який виносив вирок у справі про підпал. Офіцер вважав малоймовірним, що Бендалл, який мав історію насильницьких правопорушень, міг би здійснити злочин повторно і класифікував «ризик» як середній, а не високий. На основі такої оцінки суд призначив злочинцю умовне покарання, а через три місяці він убив свою вагітну дружину та трьох малолітніх дітей[5]. Ще один засуджений за вбивство був звільнений із в'язниці в 2022 році з оцінкою середнього ризику рецидиву злочину. Через три дні він зґвалтував і жорстоко вбив молоду жінку. Суд установив, що оцінювання ризику здійснено неправильно[6]. Тобто помилки в оцінюванні ризику ре-

[1] Samia Kabir, David N. Udo-Imeh, Bonan Kou, and Tianyi Zhang. 'Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions' *Proceedings of the CHI Conference on Human Factors in Computing Systems* (May 11–16, 2024, Honolulu, HI, USA) 1–17 <https://doi.org/10.1145/3613904.3642596 > (дата звернення: 02.05.2024).

[2] Hamilton M., Ugwudike P. A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up' (The Conversation, 2023). <https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594> (дата звернення: 02.05.2024).

[3] Хрипун В. 'В Україні штучний інтелект оцінюватиме вірогідність порушення законів' (Судово-юридична газета в Україні, 21.09.2020) <https://sud.ua/ru/news/ukraine/179753-v-ukrayini-shtuchniy-intelekt-otsinyuvatime-virogidnist-porushennya-zakoniv > (дата звернення: 02.05.2024).

[4] Daws R. 'AI is sentencing people based on their «risk» assessment' (ALNEWS, 22.01.2019) <https://www.artificialintelligence-news.com/2019/01/22/ai-sentencing-people-risk-assessment/> (дата звернення: 03.05.2024).

[5] Jack Hardy 'Murderer was left free to kill after probation blunders' (The Telegraph, 22.12.2022) <https://www.telegraph.co.uk/news/2022/12/22/murderer-damien-bendall-left-free-kill-probation-blunders/≥ (дата звернення: 02.05.2024)

[6] Simi O'Neill. *Independent serious further offence review of Jordan McSweeney* (HM Inspectorate of Probation.

цидиву злочину можуть призводити до вчинення жахливих убивств.

Вчені Інституту Алана Тюрінга (Велика Британія) та Спеціального комітету зі штучного інтелекту (CAHAI) Ради Європи встановили, що застосування правоохоронними органами прогностичних моделей ризиків рецидиву злочину та алгоритмічно вдосконалених цифрових можливостей цифрового спостереження призводить до посилення та подальшого закріплення моделей структурної дискримінації, системної маргіналізації та нерівності[1].

Використання систем ШІ може призвести до формування необґрунтованих упереджень та дискримінаційних прогнозів стосовно певної особи або групи людей за такими ознаками: стать, вік, колір шкіри, політичні або релігійні переконання, стан здоров'я, сексуальна орієнтація, місце проживання, майновий стан, фізичні вади, мова тощо. Це, зокрема, може заважати реалізації права на справедливе правосуддя, включаючи право на рівність сторін, під час використання таких систем у кримінальному правосудді. Тобто уповноважені особи, які для підтримки прийняття процесуальних рішень використовують технології ШІ, мають самостійно докладно обґрунтувати свої рішення з метою забезпечення принципу підзвітності. При цьому обидві сторони кримінального провадження повинні мати доступ до системи ШІ, яка допомагає ухвалювати процесуальне рішення, що містить прогноз про можливість вчинення повторного або додаткового злочину певною особою. Це дозволить підсудному проаналізувати його й оскаржити.

Різні аспекти захисту прав людини під час використання ШІ є предметом особливої уваги юристів в усьому світі. Міжнародною групою експертів Європейської комісії у галузі ШІ (AI HLEG) опубліковано посібник «Етичні настанови для надійного ШІ»[2], в якому сформульовані концептуальні правила роботи систем ШІ (законність, етика, відповідальність, надійність, прозорість, безпечність та ін.).

Міжнародною організацією зі стандартизації (ISO) опублікований стандарт ISO/IEC 42001:2023[3], в якому містяться вимоги до створення, тестування, впровадження, підтримки та постійного вдосконалення систем ШІ. Положення цього стандарту застосовуються до будь-яких установ чи організацій, незалежно від напряму їх діяльності.

У 2019 р. у Сінгапурі запроваджена Національна стратегія штучного інтелекту[4], яка складається з Модельної рамкової програми управління ШІ у вигляді посібника, в якому висвітлені практичні організаційні аспекти управління ШІ. Урядом Китаю у 2023 р. із метою відповідності генеративного ШІ соціальному порядку та моралі, уникнення дискримінації та дотримання права на інтелектуальну власність прийняті «Тимчасові заходи з управління генеративними службами ШІ»[5]. Канада, Індія, Бразилія і Швейцарія активно працюють над регулюванням ШІ. У США сформовано федеральну політику щодо управління ШІ, а в деяких штатах вже прийнято відповідні нормативно-правові акти[6]. В опублікованій у США Концепції Закону про штучний інтелект наголошено на важливості використання ШІ для пришвидчення технологічного розвитку країни, але висловлено застереження щодо можливого його негативного впливу на права і свободи людини. У цьому документі запропоновано запровадити п'ять рівнів захисту від цих негативних впливів, а саме: 1) захист від небезпечних або неефективних систем; 2) захист від цифрової дискримінації; 3) захист персональних даних від надмірного втручання в приватне життя людини; 4) прозорість використання технологій ШІ, зокрема, уникнення їх потенційного негативного впливу на пра-

January, 2023) 6–7 <https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2023/01/FINAL-JM-report-HMI-Probation.pdf≥ (дата звернення: 02.05.2024)

[1] Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. *Artificial intelligence, human rights, democracy, and the rule of law: a primer* (The Council of Europe, 2021) <https://www.researchgate.net/publication/350808360_Artificial_intelligence_human_rights_democracy_and_the_rule_of_law_a_primer> (дата звернення: 02.05.2024)

[2] The Ethics Guidelines for Trustworthy Artificial Intelligence (AI) European Commsssion. (08.04.2019). URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai≥ (дата звернення: 02.05.2024)

[3] ISO/IEC 42001:2023 (E). Information technology – Artificial intelligence – Management system. First edition 2023–12. International standard. Geneva, 2023. URL: https://webstore.iec.ch/preview/info_isoiec42001%7Bed1.0%7Den.pdf (дата звернення: 02.05.2024)

[4] National AI Strategy.2019. Smart nation Singapure. URL: https://www.smartnation.gov.sg/nais/ (дата звернення: 02.05.2024)

[5] Temporary measures to manage artificial intelligence generative services. China Network Information. 2023. URL: https://www.chinalawtranslate.com/generative-ai-interim/ (дата звернення: 02.05.2024)

[6] Kohn B., Pieper F-U. 'Al regulation about the world' (Taylor Wessing, 09.05.2023) <https://www.taylorwessing.com/en/interface/2023/ai---are-we-getting-the-balance-between-regulation-and-innovation-right/ai-regulation-around-the-world > (дата звернення: 02.05.2024).

ва людини та навколишнє середовище; 5) захист від автоматизованого ухвалення некоректного рішення[1].

Перший у світі закон про штучний інтелект (Artificial Intelligence Act або AI Act)[2] був прийнятий парламентом ЄС у червні 2023 р. Оновлений текст закону опубліковано 2 лютого 2024 року[3]. Цей закон установлює правила та вимоги для розробників систем ШІ та спрямований на забезпечення безпеки, прозорості, екологічності й етичності використання ШІ в ЄС. Основні положення цього закону включають класифікацію систем ШІ за рівнями ризику, яким відповідають певні регуляторні заходи. Забороні підлягають приховані технології маніпуляції емоціями людей. Розробники таких систем зобов'язані надавати можливість користувачам вільно отримувати інформацію про те, як вони працюють і приймають рішення. В законі також передбачені заходи щодо запобігання дискримінації, упередженості та несправедливості в системах ШІ. Для систем ШІ, які, зокрема, здійснюють обробку персональних даних та інформації щодо об'єктів критичної інфраструктури, запроваджується обов'язкова сертифікація. До систем, які потенційно можуть становити ризик для прав людини або державних інтересів (дискримінація, дезінформація, маніпуляції в інформаційному просторі та ін.), застосовуються певні обмеження.

Закон про ШІ (AI Act), який має стати прикладом уніфікації юридичних правил у галузі ШІ, викликав безліч дискусій в юридичних спільнотах різних країн Європи та всього світу. Частина юристів наголошують на тому, що закон повинен обмежувати застосування деяких технологій ШІ, які можуть становити небезпеку для прав і свобод людини. Інші, навпаки, вважають, що будь-які обмеження у застосуванні технологій ШІ перешкоджають розвитку інновацій і соціальному прогресу загалом. Однак усі одностайні в тому, що технології ШІ потребують справедливого правового регулювання.

В Україні розпочався процес правового регулювання ШІ з урахуванням міжнародних стандартів у цій сфері. Міністерство цифрової трансформації України опублікувало «Дорожню карту з регулювання штучного інтелекту в Україні»[4], видало методичні матеріали «Права людини в епоху штучного інтелекту: виклики та правове регулювання»[5]. Спеціалісти у галузі ШІ дискутують з приводу визначення винної особи у разі завдання шкоди системою ШІ, яка працює автономно. Частина авторів вважає, якщо негативну подію спричинили дефекти конструкції пристрою зі ШІ, винним є виробник пристрою, а якщо стався програмний збій, то розробник ШІ. Дехто вважає, що обидва суб'єкти є винними одночасно. Однак для кожного випадку має бути розроблена правова основа для розв'язання цього питання та опрацьовані відповідні підходи залежно від ступеня небезпеки й інших характеристик системи ШІ[6].

Непрозорість і складність систем ШІ може перешкоджати реалізації права на справедливий судовий розгляд, включаючи й право на рівність сторін. Тому сторона, яка підлягає алгоритмічному рішенню, може оскаржити процесуальне рішення. Хоча використання ШІ у кримінальному провадженні може зменшити свавілля та дискримінаційні дії слідчого або судді, судові рішення, підтримані або обґрунтовані ШІ, можуть негативно вплинути на незалежність судової системи. Тому учасники кримінального провадження повинні мати достатній рівень розуміння ШІ, який вико-

[1] Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. The White House. Washington, 2022. URL: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf (дата звернення: 04.05.2024).

[2] Artificial Intelligence Act. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html#sdocta6 (дата звернення: 04.05.2024)

[3] Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world. Council of the EU : Press release. 2 February 2024. URL: https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/ (дата звернення: 04.05.2024)

[4] 'Дорожня карта з регулювання штучного інтелекту в Україні' (Міністерство цифрової трансформації України)<https://cms.thedigital.gov.ua/storage/uploads/files/page/community/docs/%D0%94%D0%BE%D1%80%D0%BE%D0%B6%D0%BD%D1%8F_%D0%BA%D0%B0%D1%80%D1%82%D0%B0_%D0%B7_%D1%80%D0%B5%D0%B3%D1%83%D0%BB%D1%8E%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F_%D0%A8%D0%86_%D0%B2_%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%96_compressed.pdf> (дата звернення: 01.05.2024)

[5] 'Права людини в епоху штучного інтелекту: виклики та правове регулювання : Методичний матеріал' (Міністерство цифрової трансформації України, 2024) 44 с<https://drive.google.com/file/d/1YLb1X8wCMQi3 g8LjPsERa2b58GM1fRS2/view> (дата звернення: 01.05.2024)

[6] --

ристовується для підтримки прийняття процесуальних рішень та забезпечення підзвітності рішень, прийнятих з його допомогою.

Підзвітність системи ШІ означає оцінку її потенційної шкоди та переваг для окремих груп осіб. Така процедура гальмується відмовою розробників ШІ, які захищають свої авторські права, оприлюднювати його алгоритми. При цьому виявити потенційну несправедливу упередженість у роботі системи ШІ внаслідок її непрозорості та технічної складності дуже непросто.

Для забезпечення прозорого та підзвітного використання систем ШІ вони мають створюватися на основі міжнародних стандартів та проходити сертифікацію й постійний (протягом усього життєвого циклу) моніторинг із боку незалежних експертних органів, відповідальних за нагляд у різних галузях (охорона здоров'я, освіта, безпека дорожнього руху, правоохоронна діяльність та ін.).

Висновки. Технології ШІ багато в чому допомагають реалізації прав людини, надають їй можливість покращити та полегшити своє життя. При цьому у світі відомо багато прикладів про помилки у роботі систем ШІ, які привели до порушення фундаментальних прав і свобод людини.

На сьогодні в українському законодавстві відсутні норми, які б дозволяли контролювати системи ШІ та вимагати звіти від суб'єктів, що відповідають за їх розроблення або використання. Для формування законодавчої бази щодо використання ШІ в Україні важливо встановити чітку диференційовану систему відповідальності для осіб, які розробляють, впроваджують або використовують системи ШІ.

Для збереження балансу між розвитком ШІ і дотриманням прав людини потрібно розробити методологію оцінювання впливу ШІ на права людини, кодекси поведінки для розробників та користувачів систем ШІ, нормативно-правову базу щодо регулювання ШІ за міжнародними стандартами (захист персональних даних, прозорість алгоритмів, підзвітність і відповідальність, захист інтелектуальної власності). У системах ШІ на законодавчому рівні мають бути забезпечені критичний підхід до обрання джерел інформації та можливість людського втручання для виправлення помилок під час аналізу персональних даних із метою запобігання можливому порушенню прав людини. Такий підхід сприятиме розвитку систем ШІ, які розширюють людські можливості, а не замінюють їх, не утискають основні права людини, а навпаки, уникають дискримінації та сприяють справедливості та верховенству права.

REFERENCES

List of legal documents

Legislation

1. The Ethics Guidelines for Trustworthy Artificial Intelligence (AI) URL: European Commsssion. (08.04.2019) https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (in English)
2. ISO/IEC 42001:2023 (E). Information technology – Artificial intelligence – Management system. First edition 2023–12. International standard. Geneva, 2023. URL: https://webstore.iec.ch/preview/info_isoiec42001%7Bed1.0%7Den.pdf (in English)
3. National AI Strategy Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. The Council of Europe. URL: https://www.researchgate.net/publication/350808360_Artificial_intelligence_human_rights_democracy_and_the_rule_of_law_a_primer (in English)
4. National AI Strategy (2019) Smart nation Singapure. URL: https://www.smartnation.gov.sg/nais/ (in English)
5. Temporary measures to manage artificial intelligence generative services. China Network Information. 2023 URL: https://www.chinalawtranslate.com/generative-ai-interim/ (in English)
6. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. The White House. Washington, 2022. URL: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf (in English)
7. Artificial Intelligence Act. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html#sdocta6 (in English)

8. Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world. Council of the EU : Press release. 2 February 2024. URL: https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/ (in English)

Bibliography

Authored books

1. Simi O'Neill *Independent serious further offence review of Jordan McSweeney. HM Inspectorate of Probation. January* (HM Inspectorate of Probation, 2023) 35 <https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2023/01/FINAL-JM-report-HMI-Probation.pdf >(in English)
2. Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., and Briggs, M. *Artificial intelligence, human rights, democracy, and the rule of law: a primer* (The Council of Europe, 2021) 46<https://www.researchgate.net/publication/350808360_Artificial_intelligence_human_rights_democracy_and_the_rule_of_law_a_primer > (in English)

Edited books

3. Micklitz H-W, Pollicino O, Reichman A, Simoncini A, Sartor G, De Gregorio G, (eds) *In: Constitutional Challenges in the Algorithmic Society* (Cambridge University Press, 2021) 342 p. https://doi.org/10.1017/9781108914857 (in English)

Journal articles

4. Karchevskyi Mykola, Radutniy Oleksandr. 'Artificial Intelligence in Ukrainian Traditional Categories of Criminal Law' (2023) 1(19) Visnyk Asotsiatsii kryminalnoho prava Ukrainy 1–25 https://doi.org/10.21564/2311–9640.2023.19.281123 (in English)
5. Turuta O. V., Turuta O. P. 'Shtuchnyi intelekt kriz pryzmu fundamentalnykh prav liudyny' [Artificial Intelligence through the Prism of Fundamental Human Rights] (2022) 71 Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu https://doi.org/10.24144/2307–3322.2022.71.7 (in Ukrainian)
6. Ali Faghiri Kabol. 'The Use Of Artificial Intelligence In The Criminal Justice System (A Comparative Study)' (2022) 19 (5) Webology <https:// www.researchgate.net/publication/365027297_The_Use_Of_Artificial_Intelligence_In_The_Criminal_Justice_System_A_Comparative_Study > (in English)
7. DeGrave, A. J., Janizek, J. D. & Lee, SI. 'AI for radiographic COVID-19 detection selects shortcuts over signal' (2021) 3 Nat Mach Intell 610–619 https://doi.org/10.1038/s42256-021-00338-7 (in English)
8. Shepitko, V., Shepitko, M., Latysh, K., Kapustina, M., & Demidova, E.. 'Artificial intelligence in crime counteraction: From legal regulation to implementation' (2024) 7(1) Social and Legal Studios 135–144. https://doi.org/10.32518/sals1.2024.135

Conference paper

9. Shevchuk V. M. 'Shtuchnyi intelekt v kryminalistytsi : problemy, mozhlyvosti, perspektyvy' [Artificial Intelligence in Criminalistics: Problems, Opportunities, Prospects] *Informatsiine zabezpechennia rozsliduvannia zlochyniv*: matly Kh Mizhnar. kruhloho stolu (Odesa, 19.05.2023) 89–97 (in Ukrainian)
10. Samia Kabir, David N. Udo-Imeh, Bonan Kou, and Tianyi Zhang 'Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions' *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA May 11–16, 2024) 1–17< https://doi.org/10.1145/3613904.3642596 > (in English)

Websites

11. Daws R. 'AI is sentencing people based on their «risk» assessment' (ALNEWS) (22.01.2019) <https://www.artificialintelligence-news.com/2019/01/22/ai-sentencing-people-risk-assessment/> (in English)
12. 'Dorozhnia karta z rehuliuvannia shtuchnoho intelektu v Ukraini' [Roadmap for regulation of artificial intelligence in Ukraine] (Ministerstvo tsyfrovoi transformatsii Ukrainy) < https://cms.thedigital.gov.ua/storage/uploads/files/page/community/docs/%D0%94%D0%BE%D1%80%D0%BE%D0%B6%D0%BD%D1%8F_%D0%BA%D0%B0%D1%80%D1%82%D0%B0_%D0%B7_%D1%80%D0%B5%D0%B3%D1%83%D0%BB%D1%8E%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F_%D0%A8%D0%86_%D0%B2_%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%96_compressed.pdf > (in Ukrainian)
13. Haivoronska B. 'Triumf ta zahrozy shtuchnoho intelektu – yak neiromerezhi vplyvaiut na nashe zhyttia i yak vony zakonodavcho rehuliuiutsia' [The Triumph and Threats of Artificial Intelligence – How Neural Networks Affect Our Lives and How They Are Legally Regulated] (Cityhost.UA 04.09.2023) <https://cityhost.ua/uk/blog/triumf-ta-zagrozi-shtuchnogo-intelektu-yak-neyromerezhi-vplivayut-na-nashe-zhittya-i-yak-ce-zakonodavcho-regulyu-tsya.html#the-top-5-major-failures-of-AI> (in Ukrainian)
14. Hamilton M., Ugwudike P. A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up. (The Conversation 2023). <https://theconversation. com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594> (in English)

15. England R. 'UK police's facial recognition system has an 81 percent error rate' (Engadget, 04.07.2019). <https://www.engadget.com/2019-07-04-uk-met-facial-recognition-failure-rate.html>(in English)
16. Jack Hardy 'Murderer was left free to kill after probation blunders' (The Telegraph, 22.12.2022). <https://www.telegraph.co.uk/news/2022/12/22/murderer-damien-bendall-left-free-kill-probation-blunders/> (in English)
17. Klark E. 'Pidruchnyk z mashynnoho navchannia dlia pochatkivtsiv: shcho take, osnovy ML' [Machine Learning Tutorial for Beginners: What Is, ML Basics] (Guru99, 2024) <https://www.guru99.com/uk/machine-learning-tutorial.html> (in Ukrainian)
18. Kohn B., Pieper F-U. 'Al regulation about the world. Taylor Wessing' (Taylor Wessing, 09.05.2023) <https://www.taylorwessing.com/en/interface/2023/ai---are-we-getting-the-balance-between-regulation-and-innovation-right/ai-regulation-around-the-world> (in English)
19. Khrypun V. V 'Ukraini shtuchnyi intelekt otsiniuvatyme virohidnist porushennia zakoniv' [In Ukraine, artificial intelligence will assess the likelihood of violating laws] (Sudovo-yurydychna hazeta v Ukraini, 21.09.2020) <https://sud.ua/ru/news/ukraine/179753-v-ukrayini-shtuchniy-intelekt-otsinyuvatime-virogidnist-porushennya-zakoniv > (in Ukrainian)
20. 'Prava liudyny v epokhu shtuchnoho intelektu: vyklyky ta pravove rehuliuvannia' [Human rights in the era of artificial intelligence: challenges and legal regulation] : Metodychnyi material (Ministerstvo tsyfrovoi transformatsii Ukrainy, 2024) 44 <https://drive.google.com/file/d/1YLb1X8wCMQi3 g8LjPsERa2b58GM1fRS2/view> (in Ukrainian)

Авдєєва Г. К.

Розвиток технологій штучного інтелекту і дотримання прав людини: проблеми забезпечення балансу

Досліджено позитивний і негативний впливи технологій штучного інтелекту, що використовуються в різних сферах діяльності, на фундаментальні права людини. Здійснено критичний аналіз механізмів маніпулятивних дій щодо користувачів інтернет-ресурсів із боку їх власників за допомогою технологій штучного інтелекту. Встановлено причини виникнення помилок у роботі систем штучного інтелекту, які призводять до порушення прав людини на життя, свободу та особисту недоторканість, справедливе правосуддя та ін. Визначено потенційну загрозу систем відеоспостереження з елементами штучного інтелекту для конфіденційності та людської гідності.

На основі аналізу механізму машинного навчання систем штучного інтелекту встановлено, що якість «навчання» систем штучного інтелекту та правильність рішень, які ними приймаються, залежить від якості і повноти інформації, що ними аналізується, та надійності її джерел. На прикладах показано, що застосування правоохоронними органами прогностичних моделей ризиків рецидиву злочину та алгоритмічно вдосконалених цифрових можливостей цифрового спостереження призводить до посилення й подальшого закріплення моделей структурної дискримінації, системної маргіналізації та нерівності. Доведено, що використання систем штучного інтелекту може призвести до формування необґрунтованих упереджень та дискримінаційних прогнозів стосовно певної особи або групи людей за такими ознаками, як стать, вік, колір шкіри, політичні або релігійні переконання, стан здоров'я, сексуальна орієнтація, місце проживання, майновий стан, фізичні вади, мова тощо.

Для забезпечення прозорого та підзвітного використання систем штучного інтелекту вони мають створюватися і використовуватися на основі міжнародних стандартів (захист персональних даних, прозорість алгоритмів, підзвітність і відповідальність, захист інтелектуальної власності) та проходити сертифікацію й постійний (протягом усього життєвого циклу) моніторинг із боку незалежних експертних органів, відповідальних за нагляд у різних галузях (охорона здоров'я, освіта, безпека дорожнього руху, правоохоронна діяльність та ін.).

Ключові слова: *штучний інтелект, машинне навчання, права людини, потенційний ризик для прав людини, дискримінаційні рішення.*

Avdeeva G. K.

The Development of Artificial Intelligence Technologies and the Observance of Human Rights: Problems Ensuring Balance

The positive and negative impacts of artificial intelligence (AI) technologies applied across various fields on fundamental human rights have been studied. A critical analysis has been conducted on the mechanisms of manipulative actions towards internet resource users by their owners using AI technologies. The causes of errors in AI systems, which lead to violations of human rights such as the right to life, liberty, personal security, and fair justice, among others, have been identified. The potential threat to privacy and human dignity posed by AI-enhanced surveillance systems has been revealed.

Based on the analysis of the machine learning mechanisms of AI systems, it has been determined that the quality of AI «training» and the correctness of its decisions depend on the quality and completeness of the information analyzed and the reliability of its sources. The examples provided show that the use of predictive models for crime recidivism risks and algorithmically enhanced surveillance capabilities by law enforcement agencies leads to the reinforcement and further entrenchment of structural discrimination, systemic marginalization, and inequality. It has been proven that the use of AI systems can lead to the formation of unfounded prejudices and discriminatory forecasts against individuals or groups based on characteristics such as gender, age, skin color, political or religious beliefs, health status, sexual orientation, place of residence, property status, physical disabilities, language, etc.

To ensure the transparent and accountable use of AI systems, they must be created and utilized based on international standards (data protection, algorithm transparency, accountability, intellectual property protection) and undergo certification and continuous monitoring (throughout their entire lifecycle) by independent expert bodies responsible for oversight in various fields (healthcare, education, road safety, law enforcement, etc.).

Key words: artificial intelligence, machine learning, human rights, potential risk to human rights, discriminatory decisions

Стаття надійшла до редакції:25.10.2023 р.
Прийнята до друку: 20.11.2023 р.