DOI: 10.31359/2079-6242-2022-46-21
УДК 004.8; 342.56; 342.6; 342.7; 343.21

М. В. Карчевський, доктор юридичних наук, професор, головний науковий співробітник відділу дослідження проблем кримінального права Науково-дослідного інституту вивчення проблем злочинності імені академіка В. В. Сташиса Національної академії правових наук України; професор кафедри кримінального права і кримінології Львівського державного університету внутрішніх справ

Д. О. Куковинець, молодша наукова співробітниця відділу дослідження проблем кримінального права Науково-дослідного інституту вивчення проблем злочинності імені акад. В. В. Сташиса Національної академії правових наук України

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ ПРАВООХОРОННИМИ ТА СУДОВИМИ ОРГАНАМИ: СВІТОВИЙ ДОСВІД ТА НАПРЯМИ РОЗВИТКУ НАЦІОНАЛЬНОГО ЗАКОНОДАВСТВА*

Постановка проблеми. За останні двадцять років штучний інтелект пройшов шлях від наукової абстракції та концептуальних моделей до практичних задач та повсякденного використання. Системи штучного інтелекту використовуються практично в усіх сферах діяльності людини. Відбулися зміни у науковій рефлексії та правовому регулюванні соціалізації штучного інтелекту.

Збройна агресія рф прискорила практичне впровадження технологій штучного інтелекту в роботу національних правоохоронних органів. Розслідування воєнних злочинів, діяльності колаборантів, пропаганди на користь агресора вимагають оперативного опрацювання значних масивів даних. Правоохоронці активно використовують системи розпізнавання обличчя, відеоаналітику, транскрибування відео- та аудіозаписів. Водночас *використання штучного інтелекту правоохоронними органами без належного правового регулювання та комплексу організаційно-технічних заходів із дотримання нормативних приписів може призвести до системних порушень прав людини та, як наслідок, ускладнити євроінтеграційні процеси, діалог із міжнародними партнерами України*. Тому дослідження цієї теми вбачається вкрай актуальним.

Аналіз останніх досліджень і публікацій. Наукове підґрунтя дослідження склали праці, присвячені вивченню та розробленню актуальних проблем застосування технологій штучного інтелекту таких вітчизняних науковців, як: О. А. Баранов, Т. Г. Каткова, Ю. В. Кривицький, В. А. Мисливий, О. В. Плахотнік, О. Е. Радутний, В. А. Рекун, Н. А. Савінова, О. А. Теличко, К. С. Токарєва, В. М. Шевчук та ін. Висвітленню цих питань присвячені і розробки деяких зарубіжних дослідників, серед яких Вільям Айзек (William Isaac), Фредерік Зюйдервін Борхес (Frederik Zuiderveen Borgesius), Майкл Віл (Michael Veale), Ганнес Вестерманн (Hannes Westermann), Тор Грепел (Thore Graepel), Майкл Джойс (Michael Joyce),Бенуа Дюпон (Benoît Dupont), Андреас Енгель (Andreas Engel), Маурітц Коп (Mauritz Kop), Міхал Косінскі (Michal Kosinski), Крістіан Лум (Kristian Lum), Девід Марко Мауер (Marco Mauer), Мінделл (David Mindell),

[1] *Примітка.* Статтю підготовлено в межах розробки фундаментальної теми «Теоретичні, законодавчі та правозастосовні проблеми кримінально-правової охорони інформаційної безпеки в Україні» (номер державної реєстрації 0121U114324).

Юань Стівенс (Yuan Stevens), Девід Стіллвелл (David Stillwell), Філіп Хакер (Philipp Hacker) та ін.

Метою статті є здійснення аналізу соціалізації технологій штучного інтелекту та дослідження тенденцій правового регулювання використання відповідних систем правоохоронними органами.

Виклад основного матеріалу дослідження. Наукова дискусія щодо правового регулювання технологій штучного інтелекту є багаторівневою. Будемо говорити про стратегічний та тактичний рівні. На першому здійснюється усвідомлення глобальних проблем і перспектив, на другому отримують розв'язання практичні питання, пов'язані з використанням технологій штучного інтелекту.

Стратегічний рівень. Найбільш радикальним негативним прогнозом перспектив людства в контексті розвитку комп'ютерів є концепція технологічної сингулярності. Її автор, В. Віндж, уважає, що після появи інтелекту, який перевершить людський, швидкість прогресу стане надвеликою. Людство опиниться в «режимі, який відрізняється від нашого минулого не менш радикально, ніж ми, люди, самі відрізняємося від нижчих тварин. Така подія анулює через непотрібність всі людські закони, можливо, в одну мить. Некерована ланцюгова реакція почне розвиватися за експонентим типом без будь-якої надії на відновлення контролю за ситуацією»[1]. На думку В. Вінджа, до цього приведуть або технології штучного інтелекту (artificial intelligence, AI), або технології підсилення інтелекту людини (intelligence amplification, IA).

Сучасний рівень розвитку робототехніки актуалізує проблематику безпечної соціалізації технологій, але штучний інтелект – не єдина гіпотеза. У науці також широко представлена проблема трансгуманізму, розвитку здібностей людини за рахунок технологічних змін у її організмі. Трансгуманізм розглядається як «інтелектуальний та культурний рух, який відстоює можливість і бажаність принципового поліпшення стану людини через застосування, розвиток та надання широкого доступу до технологій ліквідації старіння, посилення людських інтелектуальних, фізичних і психологічних можливостей». Крім цього, трансгуманізм може розглядатися як «вивчення наслідків, потенційних переваг та небезпек технологій, які дають можливість подолати основні людські обмеження, а також пов'язане вивчення етичних питань, зумовлених розробкою та використанням таких технології»[2].

Попри аргументи про невідворотність поганого сценарію, людство має можливість зберегти контроль над ситуацією. Важливе значення для цього має ефективне правове регулювання. На нашу думку, перспективні завдання правового регулювання, зумовлені як гіпотезою розвитку штучного інтелекту, так і гіпотезою розвитку технологій трансгуманізму, можуть бути сформульовані у такий спосіб:

1. Розвиток технологій неможливо заборонити. Попри ризик небезпек абсолютна заборона розробки систем штучного інтелекту чи трансгуманістичних технологій є неможливою. Правове регулювання у цій сфері має забезпечувати стимулювання соціально ефективного використання технологій та мінімізацію ризиків зловживання технологією.

2. Правове регулювання має забезпечити максимальну диверсифікацію технологічних рішень та бути технологічно нейтральним. Технологія має не обмежуватися, а, навпаки, стати якомога різноманітнішою. Якщо право буде містити умови/вимоги для створення якомога більшої кількості різноманітних рішень у сфері технологій, ризик «глобальної відмови» буде мінімізований. Наприклад, відомий негативний сценарій «епідемії» імплантатів (заподіяння шкоди людству через порушення роботи всіх імплантованих пристроїв) буде просто неможливими через гарантовану наявність альтернативних технічних рішень. Водночас швидкість розвитку технологій вимагає відмовлятися від законодавчих формулювань, що вказують на певні види технологій. Будь- який закон, пов'язаний із конкретною технологією, матиме дуже обмежений час корисного існування.

3. Актуальною та затребуваною для сучасного рівня технологій є класична схема «розробник-власник-користувач». Ускладнення технологій вимагатиме переходу до нової, більш складної, схеми правового регулювання. Напевно, правове регулювання соціалізації штучного інтелекту пройде шлях від розгляду робота як об'єкта відносин до наділення його правами, обов'язками та відпо-

[1] Vinge V, 'The Coming Technological Singularity' (*Acceleration Studies Foundation*, 1993) <http://www.accelerating.org/articles/comingtechsingularity.html> (in English)

[2] Bostrom N, *The transhumanist frequently asked questions: a general introduction* (Faculty of Philosophy Oxford University, 2003) 56.

відальністю. Так само потребуватиме розв'язання проблема правового статусу фізичної особи, здатності якої підсилені за допомогою технологій трансгуманізму.

4. На додаток до традиційної юстиції, будуть формуватися дві нові сфери, які умовно назвемо «змішана юстиція» та «юстиція штучного інтелекту». Їх функціонування буде забезпечувати протидію роботам, що є загрозою для соціального розвитку й стабільності. Юстиція штучного інтелекту буде створена на основі роботів. Така система передбачатиме узагальнення в чітких алгоритмах досвіду, отриманого за час існування традиційної юстиції.

5. Оскільки контроль за розвитком та використанням певних технологій вимагатиме ефективної системи моніторингу, аналіз юридично значимої інформації стане набагато складнішим та вимагатиме принципово нових професійних компетенцій. Традиційний розподіл завдань між юристами та спеціалістами стане вкрай неефективним. Буде спостерігатися конвергенція юридичних та технічних наук. Потребуватиме розв'язання питання визначення та розвитку нових видів юридичних професій. При цьому формулювання нових законодавчих положень мають бути технологічно нейтральними, це необхідно для забезпечення стабільності правового регулювання в умовах швидкоплинних змін технологічної реальності.

6. Значення глобальної проблеми набуває питання правових гарантій ефективного розвитку інформаційного навколишнього середовища. Величезні обсяги інформації, яка накопичується людством у процесі життєдіяльності, очевидно, потребуватимуть нових правових механізмів. Існуючі зараз право власності на інформацію та право інтелектуальної власності скоріше за все будуть доповнені новими інститутами, що нагадуватимуть право використання надр та право археологічної діяльності. Комплекс означених питань слід розглядати як установлення координатної системи для майбутньої правової оцінки як штучного інтелекту, так і технологічно вдосконалених людей, оскільки саме в цій сфері відбуватиметься переважна частина їх соціально значимої активності.

Означені питання, характеризують, як ми вказали, стратегічний рівень наукової дискусії та переважно стосуються гіпотетичних технологій, тих, які поки що не існують, хоча у науці все помітнішою стає скептична думка щодо реальності повноцінного автономного штучного інтелекту. Наприклад, Девід Мінделл на підставі емпіричного дослідження з питань застосування сучасних роботів формулює три «міфи» як хибні уявлення про перспективи робототехніки. Перший міф – про лінійний прогрес, тобто припущення про те, що техніка пройде шлях від безпосереднього керування людиною до цілком автономних роботів. Другий міф – про заміщення, тобто йдеться про те, що машини поступово будуть виконувати всі людські завдання. Третій міф – про повну автономність, себто «утопічна ідея про те, що сьогодні або в майбутньому роботи зможуть діяти повністю самостійно»[1].

Дискусія щодо актуальних проблем правового регулювання технологій штучного інтелекту – *тактичний рівень* – стосується технологій, які вже використовуються та набули певного поширення. Достатньо поширеним є поділ технологій на «сильний» та «слабкий» штучний інтелект. Сильний являє собою гіпотетичний пристрій, який має здатність мислити, усвідомлювати оточуючий світ та себе як особистість, виконувати всі завдання, як і людина, або навіть перевищувати її інтелектуальні здібності. Слабкий штучний інтелект – фактично існуючі технології, орієнтовані на автоматизацію певних видів діяльності людини або кількох завдань, які виконує людина. Наприклад, керування транспортним засобом, гра в шахи, розпізнавання обличчя, голосу, рукописного тексту тощо.

Експерти визначають чотири загальні групи ризиків[2] використання таких технологій. Проблеми їх мінімізації переважно складають зміст тактичного рівня наукової дискусії щодо соціалізації технологій штучного інтелекту.

Нова якість порушення таємниці приватного життя. Автоматизована обробка даних про людину створює новий рівень загроз для людини. Аналіз уподобань у соціальних мережах[3], історії по-

[1] Mindell D, *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking, 2015) 272.

[2] Dupont B, Stevens Y, Westermann H and Joyce M, *Artificial Intelligence in the Context of Crime and Criminal Justice* (Korean Institute of Criminology, Canada Research Chair in Cybersecurity, ICCC, Université de Montréal, 2018) 228 <https://www.cicc-iccc.org/public/media/files/prod/publication_files/Artificial-Intelligence-in-the-Context-of-Crime-and-Criminal-Justice_KICICCC_2019.pdf>

[3] Kosinski M, Stillwell D and Graepel T, 'Private traits and attributes are predictable from digital records of human behavior' (2013) 110(15) *Proceedings of the National Academy of Sciences* 5802–5805 <http://dx.doi.org/10.1073/pnas.1218772110> (in English)

купок[1], інтернет з'єднань[2] із використанням технологій штучного інтелекту здатен більш ніж істотно порушити таємницю приватного життя конкретної людини.

Маніпулювання поведінкою. Технології «слабкого» штучного інтелекту вже сьогодні чинять істотний вплив на поведінку споживачів шляхом таргетованої реклами, індивідуалізованих рекомендацій пошукових сервісів, персоналізованих стрічок новин тощо. Значною є небезпека маніпуляцій із використанням штучного інтелекту у політичній діяльності[3]. Існує навіть спеціальний термін – «астротурфінг»[4], яким позначають імітацію громадської підтримки ініціатив[5].

Дискримінація. Через особливості машинного навчання технологія, яка лежить в основі «слабкого» штучного інтелекту, недостатня якість даних, використаних у процесі розробки системи, можуть призвести до системних порушень функціонування цієї технології. Прикладом означеної проблеми може слугувати упередженість автоматизованих систем відбору персоналу. «Навчальний» набір даних для таких систем, як правило, представляє собою відомості щодо успішних рішень з підбору персоналу. Оскільки цей процес у багатьох сферах не є гендерно нейтральним, мали місце випадки уведення в експлуатацію систем, які помножували гендерну нерівність під час функціонування[6].

Непрозорість. Правові гарантії інтелектуальної власності та конкурентна боротьба на ринку інформаційних технологій зумовлюють закритість алгоритмів систем штучного інтелекту, що унеможливлює перевірку правильності рішень та ефективний контроль за їх станом. У тих сферах, де неправильна робота систем штучного інтелекту здатна заподіяти значну шкоду, така ситуація створює небезпеку.

Означені ризики досить чітко окреслюють проблеми застосування систем штучного інтелекту для протидії злочинності. Використання систем штучного інтелекту правоохоронними та судовими органами здатне забезпечити якісне оновлення їх діяльності. У зарубіжних країнах у практику правоохоронних органів упроваджені проєкти, пов'язані із класифікацією та розпізнаванням об'єктів, розпізнаванням звукових сигналів (мови або, наприклад, системи визначення пострілів). Запропоновані технічні рішення для аналізу великих обсягів даних на основі алгоритмів машинного навчання. У такий спосіб здійснюється аналіз відомостей про телефонні або інтернет-з'єднання, про використання платіжних систем, віртуальних активів тощо. Такі рішення використовуються як потужні інструменти розслідування злочинів. Розробляються системи прогнозування злочинності та оцінки ризику індивідуальної протиправної поведінки на основі штучного інтелекту.

Водночас ризики використання таких систем не обмежуються небезпекою порушень таємниці приватного життя. Значними є ризики дискримінації та непрозорості. Алгоритми оцінки кримінального ризику (criminal risk assessment algorithms) використовуються деякими судами для прийняття рішень щодо визначення виду покарання, доцільності перебування у в'язниці до суду, суворості вироків. Теоретично це зменшує упередженість, оскільки судді приймають рішення на основі обробки даних, а не власних, можливо, суб'єктивних, переконань. При цьому постає надзвичайно важлива проблема. Через те, що базою для алгоритму є прийняті раніше рішення, він (алгоритм) може посилювати й увічнювати упередження, генерувати ще більшу кількість упереджених даних для

[1] Hill K, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' (*Forbes*, 16 February 2012) <www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>

[2] Zalnieriute M, 'Big Brother Watch and Others v. the United Kingdom' (2022) 116(3) *American Journal of International Law* 585–592 <http://dx.doi.org/10.1017/ajil.2022.35>

[3] 'Yuval Noah Harari argues that AI has hacked the operating system of human civilisation' (*The Economist*, 28 April 2023) <www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation>; Guggenberger N, Salib P, 'From Fake News to Fake Views: New Challenges Posed by ChatGPT-Like AI' (*Default,* 20 January 2023) <www.lawfaremedia.org/article/fake-news-fake-views-new-challenges-posed-chatgpt-ai>.

[4] Grassroots (з англ. «коріння трави») – термін сучасної американської політології; так у США називають спонтанні руху «знизу». Під grassroots розуміються, умовно кажучи, «справжні» рухи, організовані громадянами для боротьби за свої права. Імітацію ж «кореневого руху» називають astroturfing; у цьому випадку за псевдонародним рухом є політичне лобі ('Grassroots' (*Wikipedia, the free encyclopedia*, 2 February 2004) <https://en.wikipedia.org/wiki/Grassroots>)

[5] Dupont B, Stevens Y, Westermann H and Joyce M, *Artificial Intelligence in the Context of Crime and Criminal Justice* (Korean Institute of Criminology, Canada Research Chair in Cybersecurity, ICCC, Université de Montréal, 2018) 228 <https://www.cicc-iccc.org/public/media/files/prod/publication_files/Artificial-Intelligence-in-the-Context-of-Crime-and-Criminal-Justice_KICICCC_2019.pdf>

[6] '5 Examples of Biased Artificial Intelligence' (*Home | Misinformation-Fighting, AI-powered News & Fact Checking*, 30 July 2019) <www.logically.ai/articles/5-examples-of-biased-ai>

подальших циклів ще більш упереджених рішень[1]. Наприклад, якщо перед суддею особа з невеликим доходом, алгоритм з дуже великою вірогідністю буде радити застосувати ув'язнення до суду. Наступного разу у подібній ситуації алгоритм буде ще категоричніший, наступного – ще й ще…

Подібні проблеми існують і під час упровадження систем прогнозування злочинності[2]. Ідея полягає у тому, що на підставі аналізу даних про зареєстровані кримінальні правопорушення системи визначають райони, що потребують посиленої уваги з боку правоохоронців. У ці райони направляється більша кількість патрулів, чим має забезпечуватися більш ефективне використання ресурсів та досягатися необхідний рівень безпеки громадян. Результати впровадження таких системи показали зворотній бік проблеми. Чим більше поліцейських направлялося у заданий район, тим більшою була кількість виявлених у цьому районі правопорушень. Алгоритм фіксував прийняте рішення як правильне і продовжував рекомендувати посилені наряди для визначених районів. У такий спосіб увічнювався «кримінальний» статус таких районів, але загальне використання ресурсів поліції не ставало більш ефективним, загальний рівень безпеки громадян не підвищувався.

Наприклад, щоб оцінити упередженість поліцейського прогнозування, Група аналізу даних з прав людини (HRDAG) проаналізувала зареєстровані поліцейським управлінням Окленду злочини, пов'язані з наркотиками. Управління використовувало спеціальний алгоритм обробки великих даних для прогнозування наркозлочинів. Звичайно, HRDAG виявила, що прогностична модель майже виключно зосередилася на неєвропеоїдних спільнотах із низьким рівнем доходу. Але дані громадської охорони здоров'я щодо споживачів наркотиків у поєднанні з даними перепису населення США показали, що розподіл споживачів наркотиків не корелює з прогнозами програми, демонструючи, що прогнози алгоритму базувалися на упередженості, а не на реальності[3].

Показовим є те, що у червні 2020 р. міська рада Санта Круз, американського міста, яке одним із перших почало застосовувати для потреб поліції технології розпізнавання обличчя та прогнозування злочинів, відмовилася від використання таких систем, ураховуючи численні прояви упередженості їх роботи та недостатню ефективність. Рішення полягало у забороні використовувати обидві технології, за винятком схвалення міською радою на основі «висновків про те, що технологія і дані, які використовуються для цієї технології, відповідають науково підтвердженим та рецензованим дослідженням, захищають і охороняють цивільне населення, права і свободи всіх людей і не увічнюють упередженість»[4].

Крім зазначених проблем із дискримінацією, багато питань виникає й через непрозорість функціонування зазначених систем. Як зазначалося раніше, покрокове відстеження рішень, що приймається подібними системами, є доволі складною проблемою. І якщо такі ризики є допустимим під час, наприклад, автоматизованого перекладу текстів з остаточним їх редагуванням людиною, то в ситуації, коли такі алгоритми використовуються в сфері юстиції, вони мають бути максимально відкритими та прозорими[5].

Подібні небезпеки властиві не лише сфері охорони правопорядку. Накопичений досвід та «критична» маса загроз неконтрольованого розширення сфери застосування систем штучного інтелекту зумовили появу законодавчих ініціатив, спрямованих на створення комплексної нормативно-правової бази для забезпечення відповідального розвитку штучного інтелекту, захисту основних прав і сприяння інноваціям. Як найбільш актуальні слід відзначити розпочаті Урядом США 13 квітня 2023 р. громадські обговорення щодо «політики підзвітного штучного інтелекту»[6], а також обговорення проєкту, презентованого Європейською Комісією у квітні 2021 р., під назвою «The proposal

[1] Hao K, 'AI is sending people to jail – and getting it wrong' (*MIT Technology Review*, 21 January 2019) <www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/>

[2] Trejo O, 'What Is Bias in Machine Learning?' (*Scalable Path*, 4 May 2020) <www.scalablepath.com/machine-learning/bias-machine-learning>

[3] Lum K and Isaac W, 'To predict and serve? ' (2016) 13(5) *Significance* 14–19 <http://dx.doi.org/10.1111/j.1740–9713.2016.00960.x>

[4] Ibarra N, 'Santa Cruz becomes first U. S. city to approve ban on predictive policing' (*Santa Cruz Sentinel*, 24 June 2020) <www.santacruzsentinel.com/2020/06/23/santa-cruz-becomes-first-u-s-city-to-approve-ban-on-predictive-policing/>

[5] 'Technology Can't Predict Crime, It Can Only Weaponize Proximity to Policing' (*Electronic Frontier Foundation,* 3 September 2020) <www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing>

[6] 'Accountability Policy Request for Comment' (*Federal Register,* 13 April 2023) <www.federalregister.gov/documents/2023/04/13/2023–07776/ai-accountability-policy-request-for-comment>

for a regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act, AIA) and Amending Certain Union Legislative Acts»[1].

AIA використовує поняття «система штучного інтелекту» (ШІ) та визначає його у такий спосіб: *програмне забезпечення*, яке

а) розроблене з використанням одного або кількох підходів, що належать до:

• методів машинного навчання, включаючи контрольоване, неконтрольоване, та навчання з підкріпленням або використанням різноманітних методів, у тому числі глибокого навчання;

• методів, що ґрунтуються на логіці та знаннях, включаючи представлення знань, індуктивне (логічне) програмування, бази знань, логічні та дедуктивні механізми, (символічні) міркування та експертні системи;

• статистичних методів, включаючи байєсовську оцінку, методи пошуку та оптимізації;

б) може, для заданого набору визначених людиною цілей, генерувати результати, такі як контент, прогнози, рекомендації, або рішення, що впливають на середовище, з яким вони взаємодіють.

Зазначене з усією очевидністю свідчить про фокус європейського законодавця на суто практичних питаннях використання продуктів, які вже існують або можуть бути створені. За нашою класифікацією, розробка означеного проєкту належить до тактичного рівня наукової дискусії із питань соціалізації штучного інтелекту.

У AIA програми штучного інтелекту класифікуються на основі потенційних рівнів ризику. Категорія «неприйнятний ризик штучного інтелекту» забороняє розробку та використання певних програм штучного інтелекту, наприклад, систем соціального скорингу. До «штучного інтелекту високого ризику» віднесено системи, які можуть поставити під загрозу безпеку людей або порушити їх основні права.

Авторами AIA реалізовано ідею нормативної мінімізації указаних раніше соціальних ризиків впровадження ШІ. Порушення приватності пропонується контролювати у спосіб класифікованих за рівнем ризику вимог щодо розробки, введення в експлуатацію та використання систем ШІ.

Небезпеки впливу на поведінку людини мінімізуються шляхом заборони окремих видів ШІ, які пропонується відносити до «Prohibited Artificial Intelligence Practices» (Article 5, AIA). Такі системи характеризуються «неприйнятним ризиком» та поділяються на чотири категорії: дві з них стосуються когнітивного поведінкового маніпулювання людьми або певними вразливими групами; інші дві заборонені категорії – системи соціального скорингу та системи біометричної ідентифікації у режимі реального часу та на відстані. Однак для кожної категорії є винятки з основного правила[2].

Непрозорість пропонується долати шляхом обов'язкового документування створення, використання та вдосконалення високоризикованих систем ШІ, постійної актуалізації технічної документації таких систем, наявністю обов'язку виробника надавати контролюючим органам вичерпну інформацію щодо поточного стану системи ШІ, які віднесено до високоризикованих.

Нарешті, мінімізація упередженості забезпечується шляхом контролю за змістом та репрезентативністю навчальних, валідаційних і тестових наборів даних.

Важливими є положення AIA щодо підтримки досліджень та інновації в області ШІ. Зокрема, пропонується механізм «регуляторних пісочниць» (regulatory sandboxes), регуляторних інструментів, які дозволяють підприємствам тестувати та експериментувати з новими та інноваційними продуктами чи послугами під наглядом регулятора протягом обмеженого періоду часу. Регуляторні пісочниці виконують подвійну роль: 1) сприяють бізнес-навчанню, тобто розробці та тестуванню інновацій у реальному середовищі, та 2) забезпечують підтримку регуляторного навчання, тобто формулювання експериментальних правових режимів для керівництва та підтримки бізнесу в їх інноваційній діяльності під наглядом регуляторного органу. На практиці підхід спрямований на те, щоб уможливити експериментальні інновації у рамках контрольованих ризиків і нагляду, а також покращити розуміння регуляторами нових технологій[3].

[1] Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 2021/0106(COD), 2021. *EUR-lex*: of website. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[2] Kop M, 'EU Artificial Intelligence Act: The European Approach to AI' (2021) 2 *Transatlantic Antitrust and IPR Developments* 8–18 <https://law.stanford.edu/publications/euartificial-intelligence-act-the-european-approach-to-ai/>

[3] Madiega T and Van De Pol A, 'Artificial intelligence act and regulatory sandboxes' (*European Parliamentary Research*

Значну увагу AIA приділяє системам ШІ, які використовують правоохоронні органи. Певні інструменти ШІ у правоохоронних органах віднесено до категорії «високого ризику» (AIA Annex III). Йдеться про системи, призначені для:

– віддаленої біометричної ідентифікації;

– індивідуальної оцінки ризику вчинення правопорушення або ризику потенційних жертв кримінальних правопорушень;

– використання правоохоронними органами поліграфів та подібних інструментів для виявлення емоційного стану фізичної особи;

– виявлення діпфейків;

– оцінки достовірності доказів під час розслідування кримінальних правопорушень;

– прогнозування вчинення кримінального правопорушення на основі профілювання фізичних осіб, оцінки рис особистості, минулої злочинної поведінки фізичних осіб або груп;

– профілювання фізичних осіб під час виявлення, розслідування або судового розгляду кримінальних правопорушень;

– кримінального аналізу, що дозволяє правоохоронним органам здійснювати пошук у складних пов'язаних і непов'язаних великих наборах даних, доступних у різних джерелах даних або в різних форматах даних із метою виявлення невідомих закономірностей або виявлення прихованих зв'язків у даних[1].

Оскільки такі системи віднесено до високоризикових, їх користувачі, постачальники, розробники та продавці повинні дотримуватися передбачених AIA вимог. Зокрема, кожна програма повинна пройти вичерпний процес оцінки та зменшення ризиків (оцінка відповідності). Певні вимоги висуваються до даних, які використовуються для навчання цих систем штучного інтелекту. Так, їх набори мають бути достатніми, щоб попередити дискримінаційні результати та алгоритмічні упередження[2]. Сертифікація, оцінка та моніторинг високоризикованих систем ШІ має здійснюватися спеціальним уповноваженим органом; такі системи мають бути зареєстрованими та включеними у відповідну базу даних. Крім того, має бути забезпечений постійний процес запису та зберігання відомостей щодо всіх подій, які відбуваються у системі. Необхідним також є забезпечення надійності функціонування та кіберзахисту системи тощо.

Ключовою вимогою до постачальників високоризикованих ШІ є вимога створення комплексної системи управління якістю (Art. 17 AIA), що має включати:

– стратегію дотримання вимог AIA, зокрема, щодо виконання процедур оцінки відповідності та внесення змін до системи;

– техніку контролю якості та забезпечення якості;

– процедури випробування, тестування та підтвердження, які мають бути виконані перед, під час і після розробки системи;

– технічні специфікації, зокрема, стандарти, які мають бути застосовані;

– системи та процедури управління даними, зокрема, збір, аналіз, маркування, зберігання, фільтрацію, добування, агрегацію, збереження даних та будь-яку іншу операцію, пов'язану з даними, яка виконується для введення в експлуатацію високоризикових систем штучного інтелекту;

– систему управління ризиками, як безперервний ітераційний процес, що виконується протягом усього життєвого циклу системи, що систематично оновлюється;

– системи моніторингу після введення в експлуатацію;

– процедури, повідомлення про серйозні інциденти;

– процедури взаємодії із національними уповноваженими органами;

– системи та процедури зберігання всієї відповідної документації та інформації;

– управління ресурсами, зокрема, заходи забезпечення безпеки постачання;

– кадрову політику, зокрема, визначення відповідальності керівництва за напрямами системи забезпечення якості[3].

Ми приділяємо багато уваги деталям, оскільки вважаємо, що вони найкраще демонструють,

Service, June 2022) 6 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf>

[1] Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 2021/0106(COD), 2021. *EUR-lex*: of website. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[2] 'The EU AI act' (*Welcome | AP4AI*) <https://ap4ai.eu/eu-ai-act>

[3] 'Accountability Policy Request for Comment' (*Federal Register,* 13 April 2023) <www.federalregister.gov/documents/2023/04/13/2023–07776/ai-accountability-policy-request-for-comment>

як змінилася дискусія щодо правового регулювання соціалізації штучного інтелекту. Цілком природньо, що починалася вона з питань стратегічного рівня, але фактичні інформаційні технології, їх використання та розширення сфери застосування достатньо чітко визначили напрями подальшого розвитку дискурсу. Фокус змістився на чіткі прикладні задачі, дискусія набула традиційного юридичного характеру та змісту.

Законопроєкт отримав переважно схвальні відгуки науковців, водночас були представлені й критичні позиції. Наприклад, на думку М. Веле та Ф. Зуйдервена Боргесіуса, AIA був «зібраний із законодавства про безпеку продукції 1980-х років, захисту основних прав, нагляду та захисту споживачів». Проте такий підхід не дозволяє розглядати законопроєкт як всеосяжний та такий, що позбавлений істотних пробілів. Наприклад, положення про прозорість або мало відповідають чинному законодавству, або викликають більше запитань, ніж відповідей, коли розглядаються їх наслідки[1].

Новий виток дискусії з'явився з появою ChatGPT. Виникли запитання: чи може генеративний штучний інтелект загального призначення бути використаним для заподіяння шкоди; чи може він стати частиною злочинного використання ШІ та, відповідно, чи не підлягатиме він забороні як один із видів «Prohibited Artificial Intelligence Practices». Гіпотетично, представлені у відкритому доступі системи ШІ, такі як ChatGPT, Midjourney або DALLE, можуть, та, скоріше за все, будуть використовуватися для вчинення злочинів. Чи означає це, що вони мають бути забороненими? Зрозуміло, що ні. Європейський законодавець, без сумніву, знайде збалансоване рішення. Дискусія триватиме[2]. На нашу думку, вихід полягає у відмові від правил для систем та у формулюванні правил використання систем у певних сферах діяльності людини.

Висновки.

1. Встановлення тенденцій та закономірностей соціалізації штучного інтелекту відбувається у наукові дискусії, що має стратегічний і тактичний рівні.

2. Сучасна дискусія характеризується значною увагою до практичних і прикладних проблем. Правове регулювання використання технологій ШІ розглядається одночасно як засіб мінімізації ризиків та як засіб стимулювання позитивних економічних трансформацій.

3. Існує потреба правового регулювання використання систем ШІ в Україні. Чинна система норм видається недостатньою[3]. Бажано, щоб український закон про використання систем штучного інтелекту містив положення про:

- визначення, побудоване на основі європейського підходу, яке чітко обмежило б сферу нормативного впливу, структурувало національний юридичний та технічний дискурси;
- класифікацію сфер використання штучного інтелекту за небезпекою можливих ризиків;
- залежний від цієї класифікації розподіл вимог до використання систем ШІ;
- можливість як створювати системи ШІ, орієнтовані на конкретні сфери застосування, так і проводити локалізацію систем ШІ загального призначення;
- обов'язкову диверсифікацію систем ШІ, сфери використання яких характеризуються найбільшим ризиком;
- гнучкий механізм підтвердження відповідності системи ШІ вимогам, що пред'являються до її використання у певній сфері. Такий підхід стимулюватиме розроблення нових технічних рішень та забезпечуватиме необхідну динаміку використовуваних технологій;
- можливість періоду експериментального правового регулювання систем ШІ. Протягом такого періоду контролюючі органи зобов'язуються надати пропозиції розробникам щодо проходження процедури відповідності (regulatory sandboxes);
- правові засоби лібералізації інвестиційної діяльності в сфері використання систем ШІ.

4. Регулювання використання систем ШІ національними правоохоронними органами має відбуватися у спосіб формулювання спеціальних норм до загальних правил, які названі вище. Відсутність чітких та зрозумілих законодавчих по-

[1] Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach' (2021) 21(4) *Computer Law Review International* 97–112 <http://dx.doi.org/10.9785/cri-2021-220402>

[2] Hacker P, Engel A and Mauer M Regulating 'Regulating ChatGPT and other Large Generative AI Models' *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, June 2023) 1112–1123 <https://doi.org/10.1145/3593013.3594067>

[3] Karchevskyi M and Radutniy O, 'Ukrainian Report on Traditional Criminal Law Categories and AI (Artificial Intelligence)' *Traditional Criminal Law Categories and AI: Crisis or Palingenesis?* (International Colloquium Section I, 15–16 September 2022) 363–383

ложень про можливі обмеження приватності громадян під час використання технологій ШІ для протидії злочинам створює реальну небезпеку визнання діяльності правоохоронців незаконною навіть за формальною ознакою (не «у відповідності до закону»).

REFERENCES

List of legal documents

Legislation

1. Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 2021/0106(COD), 2021. *EUR-lex*: of website. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (in English)

Bibliography

Authored books

1. Bostrom N, *The transhumanist frequently asked questions: a general introduction* (Faculty of Philosophy Oxford University, 2003) 56 (in English)
2. Dupont B, Stevens Y, Westermann H and Joyce M, *Artificial Intelligence in the Context of Crime and Criminal Justice* (Korean Institute of Criminology, Canada Research Chair in Cybersecurity, ICCC, Université de Montréal, 2018) 228 <https://www.cicc-iccc.org/public/media/files/prod/publication_files/Artificial-Intelligence-in-the-Context-of-Crime-and-Criminal-Justice_KICICCC_2019.pdf> (in English)
3. Mindell D, *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking, 2015) 272 (in English)

Journal articles

4. Karchevskyi M, Losych S and Germanov S, 'Socialization of artificial intelligence and transhumanism: legal and economic aspects' (2023) 9(1) *Baltic Journal of Economic Studies* 61–70 <http://dx.doi.org/10.30525/2256–0742/2023-9-1-61-70> (in English)
5. Kop M, 'EU Artificial Intelligence Act: The European Approach to AI' (2021) 2 *Transatlantic Antitrust and IPR Developments* 8–18 <https://law.stanford.edu/publications/euartificial-intelligence-act-the-european-approach-to-ai/> (in English)
6. Kosinski M, Stillwell D and Graepel T, 'Private traits and attributes are predictable from digital records of human behavior' (2013) 110(15) *Proceedings of the National Academy of Sciences* 5802–5805 <http://dx.doi.org/10.1073/pnas.1218772110> (in English)
7. Lum K and Isaac W, 'To predict and serve? ' (2016) 13(5) *Significance* 14–19 <http://dx.doi.org/10.1111/j.1740–9713.2016.00960.x> (in English)
8. Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach' (2021) 21(4) *Computer Law Review International* 97–112 <http://dx.doi.org/10.9785/cri-2021-220402> (in English)
9. Zalnieriute M, 'Big Brother Watch and Others v. the United Kingdom' (2022) 116(3) *American Journal of International Law* 585–592 <http://dx.doi.org/10.1017/ajil.2022.35> (in English)

Conference paper

10. Hacker P, Engel A and Mauer M Regulating 'Regulating ChatGPT and other Large Generative AI Models' *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, June 2023) 1112–1123 <https://doi.org/10.1145/3593013.3594067> (in English)
11. Karchevskyi M and Radutniy O, 'Ukrainian Report on Traditional Criminal Law Categories and AI (Artificial Intelligence)' *Traditional Criminal Law Categories and AI: Crisis or Palingenesis?* (International Colloquium Section I, 15–16 September 2022) 363–383 (in English)

Websites

12. '5 Examples of Biased Artificial Intelligence' (*Home | Misinformation-Fighting, AI-powered News & Fact Checking*, 30 July 2019) <www.logically.ai/articles/5-examples-of-biased-ai> (in English)
13. 'Accountability Policy Request for Comment' (*Federal Register,* 13 April 2023) <www.federalregister.gov/documents/2023/04/13/2023–07776/ai-accountability-policy-request-for-comment> (in English)
14. 'Grassroots' (*Wikipedia, the free encyclopedia*, 2 February 2004) <https://en.wikipedia.org/wiki/Grassroots> (in English)

15. Guggenberger N, Salib P, 'From Fake News to Fake Views: New Challenges Posed by ChatGPT-Like AI' (*Default,* 20 January 2023) <www.lawfaremedia.org/article/fake-news-fake-views-new-challenges-posed-chatgpt-ai> (in English)
16. Hao K, 'AI is sending people to jail – and getting it wrong' (*MIT Technology Review*, 21 January 2019) <www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/> (in English)
17. Hill K, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' (*Forbes*, 16 February 2012) <www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> (in English)
18. Ibarra N, 'Santa Cruz becomes first U. S. city to approve ban on predictive policing' (*Santa Cruz Sentinel*, 24 June 2020) <www.santacruzsentinel.com/2020/06/23/santa-cruz-becomes-first-u-s-city-to-approve-ban-on-predictive-policing/> (in English)
19. Madiega T and Van De Pol A, 'Artificial intelligence act and regulatory sandboxes' (*European Parliamentary Research Service*, June 2022) 6 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf> (in English)
20. McGrory K, Bedi N and Clifford DR, 'Targeted' (*The Tampa Bay Times*, 3 September 2020) <https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing/> (in English)
21. 'Technology Can't Predict Crime, It Can Only Weaponize Proximity to Policing' (*Electronic Frontier Foundation,* 3 September 2020) <www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing> (in English)
22. 'The EU AI act' (*Welcome | AP4AI*) <https://ap4ai.eu/eu-ai-act> (in English)
23. Trejo O, 'What Is Bias in Machine Learning?' (*Scalable Path*, 4 May 2020) <www.scalablepath.com/machine-learning/bias-machine-learning> (in English)
24. Vinge V, 'The Coming Technological Singularity' (*Acceleration Studies Foundation*, 1993) <http://www.accelerating.org/articles/comingtechsingularity.html> (in English)
25. 'Yuval Noah Harari argues that AI has hacked the operating system of human civilisation' (*The Economist*, 28 April 2023) <www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation> (in English)

Карчевський М. В., Куковинець Д. О.
Використання технологій штучного інтелекту правоохоронними та судовими органами: світовий досвід та напрями розвитку національного законодавства

У статті розглядаються питання стратегічного та тактичного рівня впровадження технологій штучного інтелекту в суспільну практику, особливий акцент робиться на їх використанні у діяльності правоохоронних та судових органів.

Сформульовано перспективні завдання правового регулювання штучного інтелекту, серед яких: 1) стимулювання соціально ефективного використання технологій та мінімізації ризиків зловживання ними; 2) забезпечення максимальної диверсифікації технологічних рішень та технологічної нейтральності; 3) перехід до нової схеми правового регулювання; 4) формування «змішаної юстиції» та «юстиції штучного інтелекту»; 5) конвергування юридичних і технічних наук; 6) забезпечення правових гарантій ефективного розвитку інформаційного навколишнього середовища.

Окреслено ризики використання штучного інтелекту, такі як порушення приватності, маніпулювання поведінкою, дискримінація та непрозорість, що кореспондують проблемам застосування систем штучного інтелекту для протидії злочинності.

Здійснено аналіз положень проєкту Artificial Intelligence Act, презентованого Європейською Комісією у 2021 р., на предмет класифікації програм штучного інтелекту залежно від потенційних рівнів ризику (у тому числі окремо окреслено певні інструменти, що використовуються правоохоронними органами), висвітлення механізмів підтримки досліджень та інновацій у сфері штучного інтелекту, визначення вимог до постачальників високоризикованого штучного інтелекту.

Висловлюється твердження, що наразі в Україні існує проблема в правовому регулюванні використання систем штучного інтелекту. Запропоновано низку положень, що бажано відобразити в національному законодавстві з цього приводу, зокрема, наявність чіткого визначення, побудованого на основі європейського підходу; класифікація сфер використання штучного інтелекту за рівнем ризику та розподіл вимог до використання систем

штучного інтелекту; гнучкий механізм підтвердження відповідності вимогам для стимулювання та розвитку нових технологій; тощо.

***Ключові слова**: штучний інтелект, соціалізація штучного інтелекту, неприйнятний ризик штучного інтелекту, штучний інтелект високого ризику, кримінальна відповідальність, правоохоронні органи, судові органи.*

Karchevskyi M. V., Kukovynets D. O.

The Use of Artificial Intelligence Technologies by Law Enforcement and Judicial Authorities: International Experience and Directions for the Development of National Legislation

The article deals with the issues of strategic and tactical level of introduction of artificial intelligence technologies into public practice, with a special emphasis on their use in law enforcement and judicial activities.

The author formulates the perspective tasks of legal regulation of artificial intelligence, including: 1) stimulating socially efficient use of technologies and minimizing the risks of their misuse; 2) ensuring maximum diversification of technological solutions and technological neutrality; 3) transition to a new legal regulation scheme; 4) formation of «mixed justice» and «artificial intelligence justice»; 5) convergence of legal and technical sciences; 6) ensuring legal guarantees for the effective development of the information environment.

The author outlines the risks of using artificial intelligence, such as privacy violations, behavioral manipulation, discrimination, and opacity, which correspond to the problems of using artificial intelligence systems to combat crime.

The author analyzes the provisions of the draft AIA to classify artificial intelligence programs depending on potential risk levels, highlight mechanisms for supporting research and innovation in the field of artificial intelligence, and define requirements for high-risk of artificial intelligence providers.

The author argues that currently in Ukraine there is a problem with legal regulation of the use of artificial intelligence systems. The author suggests a number of provisions which should be reflected in national legislation in this regard, in particular, a clear definition based on the European approach; classification of the areas of artificial intelligence use by the level of risk and distribution of requirements for the use of artificial intelligence systems; a flexible mechanism for confirming compliance with the requirements to stimulate and develop new technologies, etc.

***Keywords**: artificial intelligence, socialization of artificial intelligence, unacceptable risk of artificial intelligence, high-risk of artificial intelligence, criminal liability, law enforcement agencies, judicial authorities.*

Стаття надійшла до редакції:10.10.2023 р.
Прийнята до друку: 20.11.2023 р.