УДК 343.9.018
DOI https://doi.org/10.32782/2524-0374/2023-7/94

ТЕХНОЛОГІЇ «ШТУЧНОГО ІНТЕЛЕКТУ» У ПРАВООХОРОННІЙ ДІЯЛЬНОСТІ: ДОСЯГНЕННЯ, ПРОБЛЕМИ, ПЕРСПЕКТИВИ ВИКОРИСТАННЯ

ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN LAW ENFORCEMENT ACTIVITIES: ACHIEVEMENTS, PROBLEMS, PROSPECTS OF USE

Федоров О.В., аспірант кафедри кримінально-правової політики
Національний юридичний університет імені Ярослава Мудрого

Революційне значення інформаційних технологій «штучного інтелекту» (скорочено – ШІ, англ. «artificial intelligence», скорочено AI) для людської цивілізації стає дедалі більш очевидним. У своєму різноманітті саме вони є унікальними інструментами узагальнення великих масивів даних, функціональні можливості яких недосяжні для інших технологічних застосунків людей. Втім, у виняткових спроможностях «штучного інтелекту» дослідники вбачають і небезпеку. Відтак має бути вирішено питання про подальше використання цього ресурсу у різних сферах соціального життя. Насамперед – у сфері захисту прав громадян (правоохоронній сфері).

Відповідно, мета роботи – вдосконалення заходів запобігання злочинності, а завдання – окреслити прийнятність і перспективи застосування технологій «штучного інтелекту» у правоохоронній діяльності.

Задля оцінки перспектив впливу технологій «штучного інтелекту» на суспільні процеси за допомогою методу історичного аналізу у статті досліджено багаторічну історію винаходів-технологій, що використовувались в опрацюванні інформації. За результатами узагальнення першоджерел стверджується, що спроможності людства у сфері інформаційно-аналітичних технологій визначались двома аспектами: 1) вдосконалення технологій опрацювання даних; 2)вдосконалення обчислювальних приладів, які їх використовували.

У статті надано визначення та розкрито історію вдосконалення технологій «штучного інтелекту». Встановлено, що вони знайшли своє належне використання у правоохоронній діяльності в якості технологічної основи так званих «аналітичних інструментів».

Водночас вбачається небезпечний потенціал технологій «штучного інтелекту» як технологічного ресурсу злочинних дій. Стверджується, що як і щодо інших технологічних застосунків, небезпечними є не самі технології «штучного інтелекту», а їх злочинне використання. Особливу увагу пропонується приділити небезпечній інтеграції «штучного інтелекту» у знаряддя – джерела підвищеної небезпеки з передаванням функцій керування ними.

Відповідно до цього пропонується створювати запобіжники, серед яких головним має стати заборона на використання технологій «штучного інтелекту» для керування зброєю. Обґрунтовується в якості превентивного заходу впровадження етичних принципів використання технологій ШІ: 1) відповідальність; 2) неупередженість; 3) відстеження; 4) надійність; 5) підпорядкування.

Ключові слова: злочинність, запобігання злочинності, «штучний інтелект», наукові методи, методи досліджень, методи кримінології.

The revolutionary significance of information technologies artificial intelligence (AI) for human civilization is becoming more and more obvious.

In their diversity, they are unique tools for summarizing large data sets, the functionality of which is unattainable for other technological applications of people. However, researchers also see danger in the exceptional capabilities of artificial intelligence.

Therefore, the issue of further use of this resource in various spheres of social life should be resolved. First of all, in the field of protection of citizens' rights (law enforcement field).

Accordingly, the purpose of the work is to improve measures to prevent crime, and the task is to outline the acceptability and prospects of using artificial intelligence technologies in law enforcement activities.

To assess the perspectives on the influence of artificial intelligence technologies on social processes using the method of historical analysis, the article examines the long-term history of technological inventions used in information processing. Based on the results of summarizing primary sources, it is claimed that humanity's capabilities in the field of information and analytical technologies were determined by two aspects:

1) improvement of data processing technologies;

2) improvement of computing devices that used them.

The article provides a definition and reveals the history of the improvement of artificial intelligence technologies. It is established that they have found their proper use in law enforcement activities as the technological basis of the so-called analytical tools.

At the same time, the dangerous potential of «artificial intelligence» technologies as a technological resource for criminal activities is seen. It is argued that, as with other technological applications, it is not the artificial intelligence technologies themselves that are dangerous, but their criminal use. It is proposed to pay special attention to the dangerous integration of artificial intelligence into tools – sources of increased danger with the transfer of their control functions.

By this, it is proposed to create safeguards, the main of which should be a ban on the use of artificial intelligence technologies to control weapons. The introduction of ethical principles for the use of AI technologies is justified as a preventive measure: 1) responsibility; 2) impartiality; 3) tracking; 4) reliability; 5) subordination.

Key words: crime, crime prevention, "artificial intelligence", scientific methods, research methods, methods of criminology.

Вступ. Для людської цивілізації революційне значення інформаційних технологій «штучного інтелекту» (скорочено ШІ, англ. «artificial intelligence», AI) стає дедалі більш очевидним. У своєму різноманітті саме вони є унікальними інструментами узагальнення великих масивів даних, функціональні можливості яких недосяжні для інших технологічних застосунків. AI охоплює різні технології, включаючи машинне навчання, обробку природної мови, комп'ютерне бачення, робототехніку та експертні системи [1, с. 3]. Наприклад, платформа «штучного інтелекту» ChatGPT здатна генерувати текстовий матеріал (відповіді на запитання, тексти різних жанрів, мовні переклади) та обробляти зображення за запитами користувачів. А у межах проекту DeepMind створено систему AlphaCode, яка пише комп'ютерні програми на конкурентному рівні [2]. У економіці впровадження цифрових технологій трансформує бізнесмоделі, заміщує людську працю машинами. Водночас діджиталізація та технологічні інновації створюють нові робочі місця, що розкривають креативний потенціал особи [3, с. 137].

Втім, у виняткових спроможностях «штучного інтелекту» дослідники вбачають і небезпечнй потенціал. Так, консалтингова компанія Eurasia Group у своєму звіті про основні ризики 2023 року окремо виділила небезпеку цієї технології [4].

Генеральний директор Tesla Elon Reeve Musk назвав ресурси «штучного інтелекту» найбільшим ризиком, з яким стикається цивілізація, і закликав до швидкого та рішучого втручання уряду для контролю цієї сфери [5].

У березні 2023 р. було опублікований відкритий лист [6], в якому дослідники штучного інтелекту закли-

кали терміново призупинити навчання систем ШІ більш функціональних, ніж GPT-4. У відповідь на побоювання, окреслених у цьому листі, Генеральний директор OpenAI Samuel H. Altman заявив, що його компанія поки не працює над GPT-5 і не буде цього робити певний час [7]. Інші фахівці мають сумніви щодо гіпотетичних загроз майбутнього, пов'язаних з розвитком ШІ. Вони стверджують, що «Ми повинні будувати машини, які працюють на нас, замість того, щоб «пристосовувати» суспільство до машинного читання та запису» [8]. Генеральний директор Facebook Mark Elliot Zuckerberg наполягає, що прогнози щодо штучного інтелекту про кінець світу є досить безвідповідальні [9]. A William Henry 'Bill' Gates вважає, що призупинення досліджень ШІ на окремий термін не вирішить проблему, тут треба визначити проблемні аспекти та врегулювати їх [10].

Щодо окресленої дискусії показовим є рішення Національного управління із захисту персональних даних Італії від 30 березня 2023 р. встановити щодо OpenAI LLC (американської компанії, яка розробляє та керує ChatGPT) тимчасові обмеження обробки персональних даних суб'єктів, зареєстрованих на території Італії [11].

Проблеми застосування новітніх технологій у вітчизняній правоохоронній свері окреслив Оболенцев В.Ф. [12].

Тож дискусія про прийнятну межу застосування технологій «штучного інтелекту» до цього часу триває. Відтак має бути вирішено питання про подальше використання цього ресурсу у різних сферах соціального життя. Насамперед – щодо охорони прав громадян.

Відповідно, метою роботи ми визначили вдосконалення заходів запобігання злочинності, а завданням – окреслити прийнятність і перспективи застосування технологій «штучного інтелекту» у правоохоронній діяльності.

Огляд літератури. Теоретичною базою для розробки технологій «штучного інтелекту» стала велика кількість ґрунтовних напрацювань фахівців. Серед таких маємо згадати роботи українського науковця Кравчука Михайла Пилиповича[13]. Фундаментальною працею за окресленою проблематикою обґрунтовано вважається публікація Turing A. «Computing machinery and intelligence», де автор дослідив математичну спроможність машин опрацьовувати інформацію та приймати рішення подібно до інтелекту людини[14].

На цей час пошуки фахівців продовжуються щодо вдосконалення технологій AI з урахуванням закономірностей еволюційних систем, вирішення завдань, що мають суб'єктивно-емоційні рішення, оптимізації розрахунків у обчислювальних алгоритмах.

Claes Strannegård, Wen Xu, Niklas Engsner, John A. Endler у статті [15] дослідили поєднання еволюції та навчання в моделях обчислювальних екосистем.

Ryan J. McCall, Stan Franklin, Usef Faghihi, Javier Snaider, Sean Kugele у статті "Motivation for Cognitive Software Agents" [16] розкрили розробку біологічно інспірованої системи мотивації, заснованої на почуттях (включаючи емоції), інтегрованих у когнітивну архітектуру LIDA на фундаментальному рівні. Ця мотиваційна система забезпечує репертуар мотиваційних можливостей (відчуття тривоги, бажання й відторгення та ін.), що діють у діапазоні часових шкал зростаючої складності. Цій проблематиці присвячена і піблікаціїя H. Georg Schulze [17], де автор розглядає можливість «розумних машин» за результатами свого функціонування створювати значущі психічні стани та реагувати на них у спосіб, подібний до того, як вони функціонують у людей.

Samuel Allen Alexander у статті "The Archimedean trap: Why traditional reinforcement learning will probably not yield AGI" [18] продемонстрував, що дійсні числа не можна використовувати для точного вимірювання неархімедових структур. Автор описав два способи, якими можна змінити традиційне навчання з підкріпленням, щоб усунути цю перешкоду.

Wladimir Stalski запропонував новий метод «навчання» для вдосконалення алгоритмів «штучного інтелекту», що ґрунтується на застосуванні людської мови у «спілкуванні» людини та людиноподібних роботів (ресурсів) [19].

Узагальнюючи напрацювання у досліджуваній сфері можемо стверджувати, що за наявності суттєвого просування окреслених технологій і до цього часу для фахівців пріоритетом фундаментальних досліджень залишається тематика безпеки ресурсів «штучного інтелекту». Прикладом може бути стаття Peter Eckersley, Anders Sandberg «Is Brain Emulation Dangerous?» у якій автори стверджують: ризики, пов'язані з емуляцією мозку (зокрема, технологіями ШІ) залежать від балансу сил між зловмисниками та захисниками в змаганнях з комп'ютерної безпеки. Дослідники пропонують метод зниження таких ризиків: безпечніше, щоб емуляція мозку відбулася раніше, оскільки повільніші процесори зроблять вплив технології більш контрольовним [20].

Матеріали та методи. Задля оцінки перспектив впливу технологій «штучного інтелекту» на суспільні процеси за допомогою методу історичного аналізу було досліджено багаторічну історію винаходів-технологій, що використовувались в опрацюванні інформації.

За допомогою методу аналізу документів досліджено Статут OpenAI [21], де окреслено місію цього ресурсу. Цим документом мають керуватися дослідників задля забезпечення найкращих інтересів людства під час усього терміну проекту.

З метою вирішення завдань нашого дослідження та оцінки небезпеки ШІ нами було встановлені новітні застосунки цього ресурсу. Перша версія GPT-1 (2018 р.) вважається проривом у технологіях опрацювання великих обсягів даних. У 2019 р. версія під назвою GPT-2 вже була здатна опрацьовувати текстову інформацію за півтора мільярдом параметрів та писати невеликі оригінальні статті з логічним викладенням матеріалу. Чат-бот з назвою GPT-3 (2020 р.) мав спроможність перекладати тексти іншими мовами, вирішувати математичні задачі. Технологічним проривом стала також функція цієї версії застосунку генерувати нові взаємозв'язки всередині «тренувальних» даних. У 2022 р. OpenAI оприлюднили демонстраційну версію ChatGPT версії 3.5. Тут відповіді генерувались вже з урахуванням загальновизнаних суспільно-етичних принципів.

Методом експерименту нами було досліджено роботу ChatGPT версії 4 [].Цей ресурс являє собою аналітичну платформу опрацювання природної (людської) мови відповідно до алгоритмів машинного навчання «Generative Pre-trained Transformer» (GPT). Відомо, що за ознакою особливостей алгоритмів «навчання» виділяють такі види технологій «штучного інтелекту»: «навчання з учителем» (Supervised Learning), «навчання без вчителя» (Unsupervised Learning); «навчання з підкріпленням» (Reinforcement Learning). «Навчання з вчителем» застосовують якщо потрібно навчити машину розпізнавати сигнальну ідентифікуючу інформацію щодо об'єктів. «Навчання без вчителя» використовують задля виявлення аномалій (незвичайного або несхожого). «Навчання з підкріпленням» використовують задля виконання завдань з великою кількістю можливих варіантів розвитку подій. ChatGPT версії 4 працює на алгоритмах т.зв. «навчання з підкріпленням із зворотним зв'язком від людини» (Reinforcement Learning with Human Feedback – RLHF). В ньому бере участь людина, яка через «навчання» уточнює результати роботи цього налаштунку, забезпечуючи таким чином дотримання вказівок і генерації відповідей з високою схожістю на авторські тексти.

Алгоритми «навчання», що є методичним функціоналом ШІ, вочевидь обумовлюють і функціональні обме-

ження цього ресурсу. По-перше, у конкретних випадках здатність систем цього виду лімітується операційною специфікою «алгоритмів» навчання та змістом первинних (т.зв. «еталонних») даних. По-друге, для «навчання» та практичного вирішення конкретних завдань потрібен час. По-третє, при суттєвих змінах у форматі та середовищі використаних даних треба здійснювати повторне навчання з урахуванням нових умов. І при тому вирішальне значення має людина, яка визначає завдання системи, обирає формат навчання, формує базу еталонних даних, визначає соціальні правила ресурсу, запускає його у експлуатацію та зупиняє у разі потреби (після отримання результатів або при суттєвій зміні середовища опрацьованих даних).

Розробники ChatGPT станом на 20 червня 2023 р. вказують такі обмеження цього ресурсу: 1) час від часу може генерувати невірну інформацію; 2) час від часу може створювати шкідливі інструкції або упереджений вміст; 3 обмежені знання про світ і події після 2021 року [21].

Методологічне значення для досліджень ШІ має факт відсутності єдиного визначення поняття «штучного інтелекту» [23] І це при тому, що термін "Artificial Intelligence" було використано у «Пропозиціях щодо дартмутського літнього проекту дослідження штучного інтелекту» [24] ще у 1955 р. [23].

Слушною ми вважаємо точку зору Європолу, який визначає «штучний інтелект» як широку галузь інформатики, що охоплює створення інтелектуальних машин, які можуть виконувати завдання, що зазвичай вимагають людського інтелекту (наприклад, розуміння природної мови, розпізнавання зображень і прийняття рішень) [1, с. 3].

У цій роботі ми будем ототожнювати терміни «технології «штучного інтелекту» та «штучний інтелект», надаючи змістовний пріоритет саме першій категорії. Скороченим варіантом цього терміну буде вказуватись абревіатура ШІ (англ. AI).

Результати та обговорення. За результатами узагальнення першоджерел можемо стверджувати, що спроможності людства у сфері інформаційно-аналітичних технологій визначались двома аспектами: 1) вдосконаленням технологій опрацювання даних; 2)вдосконаленням обчислювальних приладів, які їх використовували.

У сенсі практичної значущості для інформаційно-аналітичних технологій визначним вважаємо винахід чисел (V тисячоліттям до н.е.) та писемності (у найпростіших форматах – на перетині V–IV тисячоліття до н.е., і т.зв. «справжньої» з алфавітом – приблизно III тисячоліття до н.е.). Окрім іншого, це досягнення стало фундаментальним щодо передачі інформації та збереження даних.

Перші механічні лічильні прилади спрощували людям рахункові операцій через використання їх конструктивних складових (каміння, вузли на мотузці і т.ін.), які вважались еквівалентом кількісних характеристик об'єктів зовнішнього світу (довжини, ваги, обсягу та ін.). Наприклад, абак (III тисячоліття до нашої ери, Месопотамія) являв собою рамку з паралельними дротами, на яких розташовувались вільно ковзаючі бусини (камінці, інші невеличкі подібні елементи). Кожен ряд намистин відповідав розрядам чисел (одиниці, десятки, сотні). Ряди дротів абаку поділялись на дві частини. Першу застосовували для обрахування, а другу – для запам'ятовування результатів обрахувань.

І сторіччям до нашої ери датується використання т.зв. «Антикітерського механізму» – стародавнього механічного пристрою, призначеного для розрахунків розташувань небесних тіл.

У XVII сторіччі було винайдено кругову логарифмічну лінійку та механічну обчислювальну машину – прообраз майбутнього електронного калькулятора.

Узагальнюючи наведене, можемо стверджувати:

1. Швидкість обрахування за допомогою згаданих приладів не перевищувала швидкість аналогічних операції у людському мозоку.

2. Застосовування цих приладів обмежувалось сферою обрахунків.

3. Ці приладдя не створювали інформаційні результати, які не були б передбачувані для людини.

4. Алгоритми обрахування були доволі простими і передбачали обов'язкові механічні маніпуляції людини з приладдями.

5. Ці технологічні приладдя не мали спроможності безпосередньо керувати механічними приладами, небезпечними для людей.

Проривом у технологіях опрацювання даних став винахід комп'ютерів (середина XX століття). Їх використання характеризується такими обставинами.

1. Продуктивність опрацювання інформації значно перевищувала аналогічні здібності мозоку людини.

2. Ці прилади опрацьовували первинну інформацію та створювали інформаційний результат відповідно до запрограмованих надскладних алгоритмів.

3. Функціональність цих приладів відтворювала не тільки операції безпосереднього обрахування, але й вирішувала завдання розпізнавання, логічні операції та ін.

4. Участь людини в опрацюванні інформації програмуванням обчислювальних алгоритмів та введенням даних для їх опрацювання.

5. Комп'ютерні технології забезпечують можливість тривалого збереження результатів та обмін цими даними між користувачами.

6. Технічні прилади, в яких використовувались комп'ютерні технології, дозволяли не тільки аналізувати інформацію, але й керувати технічними процесами (зокрема – джерелами техногенної небезпеки).

Подальше вдосконалення комп'ютерізації відбувалось, зокрема, з розробкою технологій «штучного інтелекту». Newell и Simon (1956) винайшли Logic Theorist – комп'ютерну програму, яка могла доводити теореми символічної логіки. Цей застосунок вважається передвісником «штучного інтелекту», оскільки він моделював здатність людей вирішувати складні проблеми [25].

На відміну від первинних комп'ютерних технологій використання застосунків «штучного інтелекту» для опрацювання даних характеризується однією концептуальною обставиною – використанням алгоритмів так званого «навчання». Саме ці алгоритми суттєво зменшили участь людини у формуванні завдань, за якими має бути надано результат у конкретних випадках.

Завдяки функціональній спроможності «навчатись» системи штучного інтелекту можуть аналізувати дані, робити прогнози та обирати (пропонувати користувачу) варіанти неочевидних рішень. Щодо цього важливими є технології «нейронних мереж з глибоким навчанням» – найбільш сучасний підхід до машинного навчання. Вони застосовуються задля розпізнавання та генерації зображень, оцінки та прийняття багатоваріантних управлінських рішень, для машинного перекладу та вирішення інших надскладних задач.

Унікальні спроможності технологій штучного інтелекту знайшли своє належне використання у правоохоронній діяльності в якості методологічної основи т. зв. «аналітичних інструментів». «Аналітичним інструментом» Strukov V. M., Gnusov Y. V. розуміють «…методику, технічний засіб або програмний продукт (чи модуль), за допомогою яких виконується певна аналітична функція (операція)» [26, с. 64]. Дослідники визначають декілька видів аналітичних інструментів, що використовуються у сучасній практиці інтелектуального аналізу даних – кримінальному аналізі.

Щодо аналітичних інструментів, які працюють та технологіях «штучного інтелекту» та вже використовуються у правоохоронній практиці, показовою є функціональність вітчизняної платформи кримінального аналізу RICAS. Ця система дає змогу відшукати приховані зв'язки між зада-

ними об'єктами та відображати знайдену інформацію як у вигляді геоінформації, так і у вигляді хронологічної стрічки подій. Також тут втілено модуль аналізу неструктурованої інформації, що надає можливість здійснювати пошук за заданими критеріями в режимі реального часу фактично у будь-яких текстових масивах. RICAS побудовано як інтелектуальний інструмент аналізу даних, що дає змогу проводити пошук за неочевидними критеріями, аналізувати зв'язки та ступінь близькості об'єктів, осіб і подій з візуалізацією результатів аналізу. У застосунку використовуються засоби математичного моделювання, інтелектуального семантичного аналізу, наочного темпорального аналізу, аналізу поведінкового профілю та аналізу прихованих зв'язків. У режимі реального часу RICAS опрацьовує ресурси не лише інформаційних систем правоохоронних органів, а й інших відкритих державних реєстрів, систем обліку осіб, речей і подій, дозволяє прогнозувати криміногенну обстановку [26].

DeepStateUA. У цій аналітичній платформі автоматизована система нейронної обробки даних "Griselda" [27] дозволяє швидко та якісно опрацьовувати великі потоки даних. До опрацювання інформації залучені модулі автоматичного аналізу та нейронні мережі. Якість інформації забезпечується чотирма ступенями перевірки. У цьому проекті розроблено також застосунок «Броня» ("Armour") у якому потоки інформації з різних джерел перетворюються на дані про позиції злочинців-окупантів. Після опрацювання інформація передається військовим в застосунки «Кропива», «Броня», «ГісАрта», «Дельта» задля оперативного використання на полі бою. Окрім іншого ці дані дозволяють виявляти ворожу техніку, встановлювати ситуацію на деокупованих територіях, додатково перевіряти сумнівну інформацію, збільшувати обізнаність військових про місцевість та ін. [27]. А разом – запобігати злочинам країни-агресора.

Mathew Emeka Nwanga, Kennedy Chinedu Okafor запропонували технологію розпізнавання терористичної загрози. Йдеться про аналітичний функціонал, який за результатами зовнішніх даних-індикаторів генерує інформацію про загрозу. Фактично ця система пропонує службам безпеки приховану захищену розвідку даних [28].

Узагальнюючі наведені та інші факти маємо зауважити: унікальні можливості технологій «штучного інтелекту» є ефективним ресурсом правоохоронної практики. Технологіям «штучного інтелекту» притаманна унікальна властивість, яка відрізняє їх від інших інформаційно-аналітичних ресурсів людства – здатність опрацьовувати надвеликі обсяги даних з наданням неочевидних результатів. Але вочевидь, така непередбачуваність функціонування містить і потенційну небезпеку.

1. *Небезпека ШІ для демократії та прав громадян у політичній сфері*. У Статуті AI стверджується обов'язок розробників: «Ми зобов'язуємося … уникати використання штучного інтелекту або AGI, яке завдає шкоди людству або надмірно концентрує владу» [21]. Вважаємо це зобов'язання актуальним, адже функціональність технологій ШІ відкриває небезпечні можливості для зловживань щодо контроля населення. У цьому сенсі показовою є практика соціального рейтингу громадян (social credit system) під управлінням систем штучного інтелекту у Китаї. При тому, що в масштабах країни система соціального рейтингу громадян остаточно не сформована, населення обґрунтовано хвилюється [29].

Загрозу демократії становить й використання технологій ШІ задля дезінформації у формуванні суспільної думки через масове поширення упереджених публікації у соціальних мережах. Функціональні можливості ChatGPT дозволяють розробникам створювати текст для цілей пропаганди та дезінформації, який може далі поширюватись у виді оголошень у медіа.

Наприклад, за даними Комітету з розвідки Палати представників конгресу США [30] у 2015 р. мали місце проплачені «фабрикою тролей» публікації у Facebook, створені для впливу на вибори президента США у 2016 р. За даними розслідування у цей період було зроблено 3 393 рекламних оголошення, які побачили більше ніж 11,4 млн американських користувачів. «Фабрика тролей» створила 470 сторінок, які згенерували більше 80 тыс. одиниць органічного контенту, що впливало більше ніж на 126 млн. громадян США.

Вочевидь використання провокуючого контенту соціальних медіа, створених за допомогою технологій штучного інтелекту, може не тільки впливати на демократичні інституції, але й спричиняти соціальні конфлікти.

2. *Використання технологій «штучного інтелекту» для злочинних дій*.

У першому звіті Europol Tech Watch Flash «ChatGPT – вплив великих мовних моделей на правоохоронні органи»[1] наведено огляд можливого кримінального зловживання ChatGPT.

1. Шахрайство. Задля фішингу може бути використана здатність ChatGPT створювати надзвичайно реалістичний текст. А цією можливістю можна зловживати, щоб ввести в оману потенційних жертв та змусити їх довіритися злочинцям.

2. Кіберзлочинність. Функціональність ChatGPT дозволяє створювати код кількома різними мовами програмування. Це надає можливість потенційним злочинцям навіть з невеликими технічними знаннями створювати шкідливі програми та використовувати їх у кіберзлочинах.

Безумовно, з урахуванням непередбачуваного потенціалу функціональних можливостей постає питання про здатність ШІ вбити людину. На профільних інтернет-ресурсах наявна інформація про ChaosGPT – функціонал на основі технологій ШІ, для якого розробники визначили цілі: знищення людства, встановлення глобального панування, спричинення хаосу і руйнувань, контроль людей за допомогою маніпуляцій. Цей застосунок може писати та тестувати код комп'ютерних програм, збирати в Інтернеті інформацію. Втім, у мовному діалозі ChaosGPT стверджує, що на цей час у нього немає інструментів для знищення людства і відмовився від спроб розпалювати ядерну війну [31]. Ми ж погоджуємось з дослідниками, які стверджують: як мовний застосунок ChaosGPT здатний писати тексти, коди комп'ютерних програм і навіть публікувати твіти. Але станом на зараз цей ресурс не здатний механічно збирати зброю. Те, що говорить ChaosGPT, це фактично репліка науково-фантастичних текстів, форумів соціальних мереж та іншого текстового контенту Інтернету, який написали люди [30].

Втім, технології штучного інтелекту і надалі будуть вдосконалюватися. І тому критичного усвідомлення потребує потенціал їх протиправно-небезпечного застосування передусім у сфері охорони прав громадян. І щодо цього нам вбачається обґрунтованим превентивним заходом розробка та впровадження п'яти *етичних принципів,* визначених на початку 2020 р. міністерством оборони США для використання систем ШІ у військових цілях.

1. В*ідповідальність*. Військовий персонал повинен з належною увагою оцінювати дії ШІ, залишаючись повністю відповідальним за розробку, розгортання і використання систем ШІ.

2. Н*еупередженість*. Міністерство оборони США має робити кроки для мінімізації небажаних відхилень в можливостях систем ШІ.

3. *Відстеження*. Військові системи ШІ та їх можливості повинні розроблятися і розвиватися таким чином, щоб персонал мав належний рівень розуміння технології, процесів розробки та методів застосування. Для військового персоналу повинні бути доступні методології, дані й документація, що належить до використовуваних систем ШІ.

4. Н*адійність*. Можливості військових систем ШІ повинні бути однозначними, чітко сформульованими. Без-

пека та ефективність таких можливостей повинні перевірятися випробуваннями та підтверджуватися протягом усього терміну служби.

5. *Підпорядкування*. Військові системи ШІ повинні повністю виконувати призначені для них завдання, але військові повинні мати можливість виявляти та запобігати небажаним наслідкам використання ШІ. Військові також повинні мати можливість виводити з бою або вимикати системи ШІ у яких були помічені відхилення в роботі [32].

І принципово обґрунтованою є позиція керівництва Об'єднаного центру штучного інтелекту США: американські військові не будуть оснащувати системами ШІ центри управління стратегічним озброєнням, адже за запуски балістичних ракет мають завжди відповідати *тільки люди*, тобто рішення про застосування зброї масового ураження має бути прерогативою виключно людини»[32].

3. *ШІ як фактор криміногенних процесів у суспільстві*. Зокрема, застосування технологій «штучного інтелекту» у виробничих процесах обумовлює суттєві зміни на ринку трудових ресурсів: під загрозою скорочення попиту на робочу силу до повного її заміщення технологіями знаходяться робочі місця, які передбачають рутинні операції [3]. А безробіття провокує злочинність та соціальні конфлікти на релігійному, соціальному, економічному ґрунті.

Висновки:

1. В історії приладів опрацювання інформації новітнім напрацюванням є технології «штучного інтелекту». Вже зараз ці налаштунки використовуються у різних сферах суспільного життя, значно посилюючи і позитивні, і деструктивні спроможності людей.

2. Фактично ШІ став новітнім ресурсом цивілізаційного протистоянні між добром та злом, правомірним та протиправним. На нашу думку, проблемою є не технології ШІ, а потенціал їх злочинного використання. І тому у нормотворчості та правоохоронній сфері особливу увагу треба приділити рішенням про передавання «штучному інтелекту» функцій керування джерелами підвищеної небезпеки.

3. Людство вже має у своєму арсеналі знаряддя, здатні знищити усіх живих істот на планеті. Але свідома координація людської спільноти унеможливлювали цей катастрофічний варіант розвитку подій. Усвідомлення потенційної небезпеки технологій «штучного інтелекту» має стимулювати, по-перше, контроль застосування конкретних технологій ШІ у «вразливих» соціальних та технічних сферах щодо їх відповідності етичним принципам людської Цивілізації та нормативним приписам.

ЛІТЕРАТУРА

1. ChatGPT – The impact of Large Language Models on Law Enforcement. A Tech Watch Flash Report from the Europol Innovation Lab. Luxembourg, Publications Office of the European Union, 2023. 13 p. URL: https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf. (дата звернення 01.07.2023).
2. Competitive programming with AlphaCode. December 8, 2022. URL: https://www.deepmind.com/blog/competitive-programming-with-alphacode (дата звернення 01.07.2023).
3. Азьмук Н.А. Штучний інтелект у процесі праці у цифровій економіці: нові виклики та можливості. *Економічний вісник Донбасу.* 2019. № 3 (57). С. 137–147. URL: http://dspace.nbuv.gov.ua/bitstream/handle/123456789/161150/16-Azmuk.pdf?sequence=1 (дата звернення 01.07.2023).
4. Bremmer Ian, Kupchan Cliff. Eurasia group's top risks for 2023. Risk 3: Weapons of Mass Disruption. URL: https://netfreedom.org.ua/article/eksperti-nazvali-shtuchnij-intelekt-zagrozoyu-dlya-demokratiyi
5. David Z. Morris. Elon Musk Says Artificial Intelligence Is the 'Greatest Risk We Face as a Civilization'. *Fortune*. July 15, 2017. URL: https://fortune.com/2017/07/15/elon-musk-artificial-intelligence-2/ (дата звернення 01.07.2023).
6. Pause Giant AI Experiments: An Open Letter. *Future of Life*. March 22, 2023. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (дата звернення 01.07.2023).
7. Bijin Jose. GPT-5 is not in the works, Sam Altman opens up about a letter seeking a halt on AI projects. *The Indian Express*. April 17, 2023. URL: https://indianexpress.com/article/technology/artificial-intelligence/gpt-5-is-not-happening-says-sam-altman-8556272/ (дата звернення 01.07.2023).
8. Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, Margaret Mitchell. Statement from the listed authors of Stochastic Parrots on the "AI pause" letter. *DAIR*. March 31, 2023. URL: https://www.dair-institute.org/blog/letter-statement-March2023 (дата звернення 01.07.2023).
9. Catherine Clifford. Facebook CEO Mark Zuckerberg: Elon Musk's doomsday AI predictions are 'pretty irresponsible'. *Make it.* Jul. 24, 2017. URL: https://www.cnbc.com/2017/07/24/mark-zuckerberg-elon-musks-doomsday-ai-predictions-are-irresponsible.html (дата звернення 01.07.2023).
10. Jennifer Rigby. Bill Gates says calls to pause AI won't 'solve challenges. *Reuters.* April 4, 2023. URL: https://www.reuters.com/technology/bill-gates-says-calls-pause-ai-wont-solve-challenges-2023-04-04/ (дата звернення 01.07.2023).
11. Garante per la protezione dei dati personali. Provvedimento del 30 marzo 2023 [9870832]. https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870832 (дата звернення 01.07.2023).
12. Оболенцев В.Ф. Інформаційні технології захисту даних у правоохоронній діяльності. *Наукові іновації та передові технології. Серія «Право»*. 2022. № 10(12). С. 169–177. DOI: https://doi.org/10.52058/2786-5274-2022-10(12)-169-178
13. Вірченко Н. О. Кравчук Михайло Пилипович. *Енциклопедія Сучасної України* / редкол.: І. М. Дзюба, А. І. Жуковський, М. Г. Железняк [та ін.] ; НАН України, НТШ. К., Інститут енциклопедичних досліджень НАН України, 2014. URL: https://esu.com.ua/article-2611 (дата звернення 01.07.2023).
14. Turing A. M. I. – computing machinery and intelligence. *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460. URL: https://doi.org/10.1093/mind/LIX.236.433 https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf (дата звернення 01.07.2023).
15. Claes Strannegård, Wen Xu, Niklas Engsner, John A. Endler. Combining Evolution and Learning in Computational Ecosystems. *Journal of Artificial General Intelligence*, 11(1), 2020. P. 1–37. DOI: 10.2478/jagi-2020-0001
16. Ryan J. McCall, Stan Franklin, Usef Faghihi, Javier Snaider, Sean Kugel. Artificial Motivation for Cognitive Software Agents. *Journal of Artificial General Intelligence*, 2020 11(1):38–69 DOI:10.2478/jagi-2020-0002 (дата звернення 01.07.2023).
17. H. Georg Schulze. The Synthesis and Decoding of Meaning. *Journal of Artificial General Intelligence*, 2021, 12 (1), pp. 26–70. DOI:10.2478/jagi-2021-0002
18. Samuel Allen Alexander. The Archimedean trap: Why traditional reinforcement learning will probably not yield AGI. *Journal of Artificial General Intelligence*. 2020, 11(1) 70–85. DOI: https://doi.org/10.2478/jagi-2020-0004
19. Wladimir Stalski. A New Approach to Creation of an Artificial Intellect and Method of its Implementation. *Journal of Artificial General Intelligence*. 2021 12(1):87–110. DOI: https://doi.org/10.2478/jagi-2021-0004
20. Peter Eckersley Anderson Sandberg. Is Brain Emulation Dangerous? *Journal of Artificial General Intelligence*, 2013, 4(3). Pp. 171–195. DOI:10.2478/jagi-2013-0011
21. OpenAI Charter. OpenAI. URL: https://openai.com/charter (дата звернення 01.07.2023).
22. OpenAI. URL: https://chat.openai.com/ (дата звернення 01.07.2023).
23. Pei Wang. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*. 2019 10(2):1–37. DOI:10.2478/jagi-2019-0002

24. J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon. A proposal for the Dartmouth summer research project on artificial intelligence. August 31, 1955. URL: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (дата звернення 01.07.2023).

25. Gugerty, Leo. Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. .2006. 50. 880–884. DOI: https://doi.org/10.1177/154193120605000

26. Strukov V. M., Gnusov Y. V. Інструментальні інтелектуальні платформи для кримінального аналізу. *Law and Safety*, vol. 83, no. 4, pp.64, 2021. С. 64–79. P. 64–65. (дата звернення 01.07.2023).

27. Griselda. Автоматизована система внесення, обробки та передачі інформації з використанням штучного інтелекту. URL: https://www.griselda.com.ua/ (дата звернення 01.07.2023).

28. Mathew Emeka Nwanga, Kennedy Chinedu Okafor. Computational Robotics: An Alternative Approach for Predicting Terrorist Networks. *International Journal of Robotics and Automation Technology*. January 2021, 8(1):1–11. DOI: 10.15377/2409-9694.2021.08.1

29. Ван Цзюфен, Пан Є. Розслідування загального хаосу кредитного покарання: де межа? Як ним зловживають? *Xinhua Daily Telecom*.13 липня 2020 р. URL: https://www.chinanews.com.cn/gn/2020/07-13/9236503.shtml (дата звернення 01.07.2023).

30. 30.Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements. Washington, D.C., US House of Representatives. Democrats. Permanent Select Committee on Intelligence. URL: https://democrats-intelligence.house.gov/social-media-content/default.aspx (дата звернення 01.07.2023).

31. Chloe Xiang. AI Tasked With 'Destroying Humanity' Now 'Working on Control Over Humanity Through Manipulation. *VISE*. 12.04.2023. URL: https://www.vice.com/en/article/z3mxe3/ai-tasked-with-destroying-humanity-now-working-on-control-over-humanity-through-manipulation (дата звернення 01.07.2023).

32. Пацурія Н. Б. Впровадження технологій штучного інтелекту у забезпечення національної безпеки та обороноздатності України: проблеми та перспективи повоєнного періоду. Координата: Платформа стратегічної та законотворчої аналітики. URL: https://coordynata.com.ua/vprovadzenna-tehnologij-stucnogo-intelektu-u-zabezpecenna-nacionalnoi-bezpeki-ta-oboronozdatnosti-ukraini-problemi-ta-perspektivi-povoennogo-periodu (дата звернення 01.07.2023).