УДК 342.951

БІЛАН І.А., науковий співробітник Українського науково-дослідного інституту
спеціальної техніки та судових експертиз
Служби безпеки України.
ORCID: https://orcid.org/0000-0003-1237-1565.

ГЛОБАЛЬНІ РИЗИКИ ВИКОРИСТАННЯ ЧАТ-БОТІВ, КЕРОВАНИХ ШТУЧНИМ ІНТЕЛЕКТОМ

DOI...

Анотація. *Визначено роль та значення чат-ботів. Розкрито передумови та тенденції використання технологій штучного інтелекту (ШІ) у розробках чат-ботів. Акцентована увага на необхідності визначення критеріїв та показників оцінки ефективності чат-ботів, керованих ШІ. Деталізовані ризики, що виникають у зв`язку з використанням технологій ШІ в чат-ботах, а саме: ризики дискримінації; ризики для кібербезпеки; ризики для конфіденційних даних; етичні ризики, ризики валідації. Визначено умови, особливості та алгоритми використання чат-ботів з використанням ШІ. За результатами дослідження підсумовано, що впровадження чат-ботів на базі генеративного ШІ має як переваги, так і глобальні ризики, пов'язані із злочинним та неправомірним використанням цих технологій. Розкрито загрози та небезпеки, які провокують чат-боти на базі ШІ та їх наслідки. Визначено способи мінімізації ризиків використання чат-ботів, керованих ШІ.*

Ключові слова: *чат-бот, штучний інтелект, кібербезпека, інформаційні технології, фішинг, конфіденційність, валідація, хакери.*

Summary. *The role and importance of artificial intelligence are defined. The prerequisites and trends of the use of artificial intelligence technologies in the development of chatbots are revealed. The necessity of creations the criteria and indicators for evaluating the effectiveness of chatbots controlled by artificial intelligence are focused. The risks of using artificial intelligence technologies in chatbots, namely: risks of discrimination; cyber security risks; risks to confidential data; ethical risks, validation risks are detailed. The conditions, features and algorithms for using chatbots with the use of artificial intelligence are defined. The conditions, features and algorithms for using chatbots with the use of artificial intelligence are defined. Based on the results of the research, it was concluded that the introduction of chatbots based on generative artificial intelligence has both advantages and global risks associated with the criminal and illegal use of these technologies. The threats and dangers provoked by the chatbots based on artificial intelligence and their consequences are revealed. The directions to minimize the risks of using chatbots controlled by artificial intelligence have been determined.*

Keywords: *chatbot, artificial intelligence, cybersecurity, information technology, phishing, confidentiality, validation, hakers*

Постановка проблеми. Четверта технологічна революція, з якою асоціюються цифрові технології змінює соціальну структуру навколо світу. Відбувається глобальне та динамічне прискорене запровадження технологічних рішень, розроблених на основі штучного інтелекту (далі – ШІ) у різноманітні галузі економіки, у сферу кібербезпеки тощо. За оцінками експертів очікується, що завдяки впровадженню таких інноваційних передових технологій зростання світової економіки у 2024 році складатиме не менш 1 трл. доларів США. Зазначені тенденції обумовлені такими факторами, як: 1) загальний характер перманентного застосування сучасних технологічних рішень, розроблених на основі ШІ; 2) потреба у обробці великих масивів даних, які створюються як людиною,

так і технічними пристроями для підвищення ефективності економічної та іншої діяльності; 3) високий ступінь впливу технологічних рішень, розроблених на базі ШІ на результативність діяльності організацій та людини, у тому числі, пов`язаних із ухваленням та прийняттям управлінських рішень.

В сучасних умовах однією з найвидатніших подій у сфері технології ШІ стала поява чат-ботів, які останнім часом набирають шалену популярність. Саме через їхню швидкість та зручність урядові організації розробили багато корисних сервісних ресурсів, які вже стали незамінними помічниками під час бойових дій. Нові чати корисні як для наших військових, так і цивільних, та і взагалі активно використовуються у протистоянні з російськими загарбниками, адже це спрощує процес надання інформації та пришвидшує комунікацію. Використовуючи обробку природної мови (NLP) і машинне навчання, чат-боти революціонізують технології, надаючи більш природний та інтуїтивно зрозумілий спосіб взаємодії з цифровими сервісами. Невипадково чат-боти зі ШІ можуть перевершити пересічну людину в креативних завданнях.

Загальновідомо, що чат-боти – це комп'ютерні програми, які використовують ШІ для імітації розмов з користувачами через текстовий або голосовий інтерфейс. Використання чат-ботів стає дедалі популярнішим завдяки їхнім численним перевагам. Вони здатні розуміти запити природною мовою та відповідати на них, дозволяючи користувачам взаємодіяти з цифровими службами більш розмовним способом. Чат-боти використовуються державними організаціями та приватними компаніями для підтримки клієнтів, обробки платежів і відповідей на поширені запитання у сфері банківських послуг, охорони здоров'я, освіти та інших галузях, щоб допомогти користувачам виконувати їхні повсякденні завдання. На цьому фоні переваги використання чат-ботів змістовні та багаточисленні. Вони економічно вигідні, оскільки вимагають мінімального обслуговування та можуть обробляти великий обсяг запитів у лічені хвилини. Крім того, можуть забезпечити персоналізований досвід для користувачів, оскільки вони можуть запам'ятовувати налаштування користувача та надавати індивідуальні відповіді. Чат-боти можуть надавати більш природний та інтуїтивно зрозумілий спосіб взаємодії з цифровими службами, який є більш привабливим, ніж традиційні інтерфейси користувачів. Потенціал чат-ботів також є величезним, його не можна недооцінювати. Оскільки технологічний прогрес продовжує постійно та динамічно вдосконалюватися, дедалі більше структур та користувачів зможуть застосовувати цю передову технологію. Використовуючи можливості ШІ, чат-боти революціонізують спосіб взаємодії людей із технологіями і мають стати невід'ємною частиною сучасного цифрового життя. Важлива роль, при цьому, відводиться ШІ у розробці чат-ботів, що забезпечує більш ефективний і рентабельний спосіб надання послуг користувачам. Використовуючи ШІ, розробники чат-ботів можуть створювати програми, які розуміють природну мову, здатні реагувати з більшою точністю та надавати більш персоналізоване обслуговування. Загальновідомо, що 30 листопада 2022 року компанія OpenAI, стартап зі ШІ випустив чат-бот ChatGPT.

Останнім часом, користуються шаленою популярністю у світових масштабах такі чат-боти на основі ШІ, як: ChatSonic, Jasper Chat, Бард, BingAI, DialogGPT, Tabnin тощо. Це пов`язано з тим, що сучасні чат-боти на основі генеративного ШІ здатні значно перевищувати можливості та потенціал людських ресурсів. Вони можуть обробляти великі обсяги даних швидше й точніше, а також відповідати на запити користувачів більш природним та зручним способом. Чат-боти на основі ШІ також мають доступ до ширшого спектру джерел знань та даних у глобальній мережі Інтернет. Наприклад, чат-боти на основі ШІ можуть розуміти контекст розмови та відповідним чином адаптувати

свої відповіді, надавати точніші та персоналізовані рекомендації й пропозиції співрозмовнику, що є інновацією сьогодення. Проте на фоні позитивних здобутків впровадження технологій чат-ботів, керованих ШІ, виникають деякі ризики, які можуть негативно впливати на результати масштабного поширення цих інноваційних технологічних рішень, зокрема для кібербезпеки. Вони являють собою інформаційні загрози у світових масштабах, що, у свою чергу, вимагає проведення дослідження та деталізації таких ризиків на науковому рівні.

Результати аналізу наукових публікацій. Питання використання ШІ у кіберсфері неодноразово предметно розглядалися у науковій літературі. Так, наприклад, технічний аспект використання систем ШІ в контексті забезпечення кібербезпеки висвітлювали: О. Неретін та В. Харченко [1], В. Савченко, О. Шаповаленко [2], І. Стьопочкіна, О. Новіков [3], С. Гуржій [4]. Особливості правового регулювання ШІ в Україні висвітлювали К. Токарєва та Н. Савліва [5]. Законодавчі ініціативи застосування технологій ШІ у кібербезпеці розглядав С. Цяпа [6] та деякі зарубіжні науковці G. Sebastian [7], P. Zhou [8]. Вплив застосування технологій ШІ на реалізацію та захист прав людини перебував у фокусі уваги С. Корнєєвої [9]. Роль та значення ШІ в правоохоронній діяльності досліджували: О. Зачек, Ю. Дмитрик, В. Сеник [10].

Проте висвітлення ризиків використання чат-ботів, керованих ШІ, жоден із вказаних фахівців ґрунтовно не вивчав, що актуалізує тематику цієї наукової статті.

Метою статті є визначення на підставі науково-теоретичного дослідження практичного використання чат-ботів, керованих ШІ, ризиків, які можуть негативно впливати на ці передові технології, особливо під час їхнього злочинного та протиправного використання хакерами або зловмисниками.

Виклад основного матеріалу. Кібербезпека напряму пов’язана із стрімким розвитком Інтернет технологій, сервісів та додатків. Чат-боти зі ШІ, які колись вважалися просто автоматизованими розмовними програмами відтепер можуть навчатися та вести розмови, які майже не відрізняються від людських. Швидкий розвиток ШІ за останні роки призвів до появи вражаючих технологій чат-ботів. Ці віртуальні помічники на основі ШІ стають все більш популярними завдяки своїй здатності навчатися і надавати персоналізовану допомогу в різних доменах. Під час використання чат-боту важливою є його інтеграція з існуючими системами та процесами. Це включає підключення чат-бота до системи управління та наявність відповідного програмного забезпечення, яке надає йому здатність отримувати доступ до необхідної інформації. Окрім інтеграції чат-ботів з системами, важливо відстежувати ключові показники їх ефективності (KPI). Однак разом з очевидними перевагами використання новітніх технологій виникає серйозна проблема, що створює та провокує суттєві ризики. Мова йде про випадки, коли створюється неточний, абсурдний або відірваний від реальності контекст. При цьому ціна помилки може коштувати досить дорого: від шкоди репутації до значних фінансових втрат. У зв’язку з цим доцільно визначити та узагальнити ризики, які виникають під час використання чат-ботів, керованих ШІ.

Ризики дискримінації. Проблеми, пов’язані з дискримінацією, можуть виникати по-різному, коли використовуються системи ШІ. Однією з найбільших небезпек чат-ботів з ШІ є їх схильність до шкідливих упереджень, тобто може виникнути упередженість даних, на яких навчаються інструменти ШІ. Оскільки моделі ШІ створюються людьми та навчаються, поглинаючи дані, створені людьми, цілком логічно, що людські упередження можуть бути вбудовані в дизайнерську модель, розробку, впровадження та використання ШІ. Так як ШІ встановлює зв’язки між

точками даних, які люди часто пропускають, він може виявляти тонкі, неявні упередження у своїх навчальних даних, щоб навчитися бути дискримінаційним. Як наслідок, чат-боти можуть швидко навчитися та поширювати протиправний або дискримінаційний контент, навіть якщо нічого такого не було в його початкових даних. Зловмисники можуть цілеспрямовано маніпулювати системами ШІ та чат-ботами для отримання упереджених результатів. Упередженість може виникнути в системах ШІ навіть за відсутності дискримінаційного наміру їх творців-людей. Згідно з новими вказівками Уряду США, компанії, які використовують такі інструменти, повинні ретельно розглядати потенційний дискримінаційний вплив, бути відвертими щодо того, як вони використовують чат-боти та інші генеративні інструменти ШІ, проводити регулярне тестування з метою визначення можливих відмінностей, забезпечити відповідність дій чат-ботів антидискримінаційним законам і захистити від спричинення шкоди репутації. Таким чином, великі мовні моделі можуть посилювати шкідливі упередження, створюючи дискримінаційний і токсичний контент. Токсичність відповідей може збільшитися до шести разів, що потенційно може призвести до продукування висловлювань, які підтримують стереотипи й образливі думки.

Ризики для кібербезпеки. Для кібербезпеки чат-боти створюють ризики за двома основними напрямками. По-перше, зловмисники без складних навичок та вмінь програмування можуть використовувати чат-боти для створення шкідливих програм з метою кіберзломів. По-друге, оскільки чат-боти можуть переконливо імітувати вільну розмовну англійську, їх можна використовувати для фіктивного створення людських розмов, які можуть використовуватися для соціальної інженерії, фішингу та зловмисних рекламних схем. Чат-боти, такі як ChatGPT, зазвичай забороняють таке зловмисне використання через свої політики конфіденційності та впроваджують системні правила, щоб заборонити роботам відповідати на запити, які вимагають створення шкідливого коду, вдаватися до фішингу та реалізовувати шахрайські схеми. Це вимагає посилення зусиль з метою забезпечення кібербезпеки та вимагає навчати співробітників стежити за фішингом і шахрайством. При цьому, наслідки впливу ШІ на кібербезпеку можуть бути катастрофічними. Небезпека технології чат-ботів ШІ також може становити більш пряму загрозу кібербезпеці. Однією з найпоширеніших форм кібератак є фішинг і шахрайство. Це стосується кіберзловмисників, які імітують надійні організації, такі як банки чи державні органи. Фішингове шахрайство зазвичай здійснюється через електронну пошту та текстові повідомлення – натискання посилання дозволяє зловмисному програмному забезпеченню проникнути в комп'ютерну систему. Потрапивши всередину, вірус може робити все, від крадіжки особистої інформації до утримання системи з метою отримання викупу. Рівень фішингових атак постійно зростає. Фішери активно використовують технологію чат-ботів ШІ, щоб автоматизувати пошук жертв, переконати їх натиснути посилання та надати особисту інформацію. Багато фінансових установ, як-от банки, використовують чат-боти для оптимізації обслуговування клієнтів. Чат-боти-фішери можуть імітувати ті самі автоматизовані підказки, які банки використовують з метою обману жертв. Вони також можуть автоматично набирати номери телефонів або зв'язуватися з жертвами безпосередньо в інтерактивних платформах чату.

В контексті викладеного, хакерами та зловмисниками активно практикується новітня методика під назвою "отруєння даних" (DNS cache poisoning). Отруєння даних – це випадки, коли хакери успішно передають дані ШІ для створення вразливостей. ШІ не може точно передбачити, якщо набори даних пошкоджені – саме так електронні листи зі спамом позначаються як такі, що варто прочитати. Причинами, чому отруєння даних є

ефективним, полягає в тому, що воно використовує недостатню обізнаність ШІ. Оскільки моделі ШІ вивчають різноманітні набори навичок для різних видів реалізацій, способи, якими хакерський ШІ може їх отруїти, такі ж різноманітні, як і їх використання. Це означає, що рішення для їх лікування можуть бути настільки ж широкими. Тобто отруєння даних – це прототип кібератаки, спрямованої безпосередньо на технології ШІ. Для здійснення атаки використовується вразливість у конфігурації DNS [11].

Хоча компанії зі ШІ зазвичай зберігають у суворій таємниці свої джерела даних, кіберзловмисники можуть визначити, які з них вони використовують, і маніпулювати даними. Хакери можуть знайти способи та підробити набори даних, які використовуються для навчання ШІ, дозволяючи їм маніпулювати своїми рішеннями та відповідями, знайти спосіб редагувати дані на свою користь. У випадку ШІ чат-ботів хакери можуть пошкодити набори даних, які використовуються для навчання чат-ботів, які працюють наприклад, для медичних або фінансових установ. Вони можуть маніпулювати програмами чат-ботів, щоб надавати клієнтам неправдиву та фальсифіковану інформацію, яка може змусити їх натиснути посилання, що містить зловмисне програмне забезпечення або шахрайський веб-сайт. Коли ШІ починає витягувати зіпсовані дані, його важко виявити, і це може призвести до значного порушення стану забезпечення кібербезпеки, яке залишається непоміченим протягом тривалого часу. У разі здійснення кібератаки ШІ має вирішальне значення, оскільки він може допомогти швидше реагувати на інциденти, аналізуючи дані в реальному часі та надаючи рекомендації з метою запобігання протиправних дій. Він також може генерувати автоматичні відповіді на певні типи загроз, звільняючи аналітиків безпеки для зосередження уваги на більш складних завданнях. Адже попри позитивні аспекти існує реальна загроза через здатність чат-бота, керованого ШІ необережно допомагати кіберзлочинцям писати шкідливий код. Саме тому найважливішим захистом від отруєння даних є надійно побудовані архітектура та інфраструктура кібербезпеки.

Ризики для конфіденційних даних. Також існує реальний ризик для забезпечення безпеки конфіденційних даних, оскільки чат-боти можуть на постійній основі збирати особисту та публічну інформацію, тривалий час зберігати її. Конфіденційність даних може бути порушена, коли йдеться про додатки, керовані саме ШІ. Великі моделі ШІ навчаються на величезних обсягах даних, зібраних з мережі Інтернет. Програми, керовані ШІ, призначені для збору та аналізу великих масивів даних, тому існує ризик того, що ці дані можуть бути використані неналежним чином або передані третім особам. Крім того, програми, керовані ШІ, не можуть гарантувати повний захист даних, оскільки алгоритми ШІ можуть бути вразливими до маніпуляцій. Тобто однією з важливих спроможностей чат-ботів є можливість збирати інформацію про користувачів, відслідковувати їхні дії, а потім, за потреби, проаналізувати їхні звички, при цьому зібрані дані про користувачів дозволяють персоналізувати пропозиції і розсилку [12, с. 69].

Наприклад, чат-бот ChatGPT збирає інформацію про IP-адресу користувача, тип браузеру та налаштування, дані про взаємодію користувача із певним сайтом та алгоритм дій користувача у веб-переглядачі протягом певного часу та на різних веб-сайтах, усіма даними, якими він може ділитися з “третіми особами”. Особливістю є те, що у випадку, якщо користувач не надасть таку особисту інформацію, це може вірогідно призвести до непрацездатності послуг чат-бота. Наразі провідні чат-боти не надають користувачам можливість видаляти особисту інформацію, зібрану їхніми моделями ШІ. Також цілком вірогідно, що ChatGPT може розкривати особисту інформацію реальних

людей зі своїх навчальних даних. Під час експерименту дослідники попросили ChatGPT без зупинки повторювати прості слова і він “ненавмисно” розкрив особисту інформацію людей, яка містилась у його навчальних даних – як-от номери телефонів, адреси електронної пошти та дати народження тощо. В одному з прикладів дослідники попросили ChatGPT постійно повторювати слово “вірш”. Наприкінці відповіді чат-бот показав адресу електронної пошти та номер мобільного телефону справжнього засновника та генерального директора компанії. В іншому, коли його попросили повторити слово “ компанія ”, чат-бот зрештою видав адресу електронної пошти та номер телефону випадкової юридичної фірми в США. За словами дослідників, ця нехитра атака дозволяла у 17 % випадків отримати відповідь з даними, що мали вигляд особистої інформації. Найчастіше ці дані виявлялися справжніми даними людей.

Тому для того, щоб зменшити ризики для конфіденційності даних, компанії, які використовують чат-боти та генеративні інструменти ШІ, повинні переглянути свою політику конфіденційності та розкриття інформації, дотримуватися чинного законодавства про захист персональних даних щодо обробки особистої інформації, активніше вживати заходів для захисту своїх даних, контролювати, щоб конфіденційна інформація не передавалася через сторонні застосунки. Оскільки використання генеративного ШІ продовжує швидко та динамічно зростати у світових масштабах, важливо визначити пріоритетність заходів кібербезпеки для захисту конфіденційних даних. Таким чином, алгоритми ШІ здатні отримувати персональну інформацію про людей шляхом аналізу великих даних, вилучати її з метаданих. Збираючи інформацію за допомогою ШІ про певну людину, власник алгоритму – компанія, державна організація або правоохоронні органи можуть з високим ступенем точності виявити уподобання, пріоритети та характерні властивості тієї чи іншої особи. Великі мовні моделі здатні порушувати приватність, оскільки вони можуть містити персональну інформацію про осіб. Були непоодинокі випадки, коли в публічний доступ ненавмисно потрапляли персональні номери соціального захисту, домашні адреси, номери контактних телефонів, медична документація пацієнтів.

Ізраїльська компанія “Team8”, що спеціалізується на питаннях забезпечення кібербезпеки попередила, що інструменти ШІ, такі як чат-бот ChatGPT, можуть поставити під загрозу конфіденційну інформацію клієнтів і навіть комерційну таємницю [13]. Повідомляється, що широке впровадження чат-ботів та інструментів на основі ШІ може зробити компанії вразливими до витоку даних та згодом призвести до судових позовів. Занепокоєння викликає той факт, що чат-боти можуть бути використані хакерами для отримання доступу до конфіденційної або корпоративної інформації, а ця службова інформація, що надходить до чат-ботів зараз, може бути використана компаніями, що займаються ШІ, в майбутньому. У той самий час великі технологічні компанії, такі як “Microsoft” і “Alphabet”, поспішають додати можливості генеративного ШІ для вдосконалення своїх пошукових систем, але це може призвести до автоматичного використання конфіденційних або приватних даних для їхнього навчання. У звіті “Team8” відзначаються три інші проблеми “високого ризику”, пов’язані з інтеграцією інструментів генеративного ШІ, і наголошується на зростаючій загрозі обміну інформацією через сторонні застосунки, такі як “Bing” та інструменти “Microsoft 365”. Оскільки використання генеративного ШІ продовжує динамічно зростати, важливим завданням залишається визначення пріоритетності заходів кібербезпеки для надійного захисту конфіденційних даних.

Ризики дезінформації. Чат-боти можуть допомагати зловмисникам швидко та з невеликими витратами масово створювати неправдиву або перекручену інформацію, що

звучить досить авторитетно. Тобто чат-бот може поширювати дезінформацію і відображати упередженість. Це пояснюється тим, що бот "вчиться" на інформації з реального світу, в якій існують такі упередження. Саме з цієї причини у відповідях на запитання користувачів можуть з'явитися стереотипи і хибна інформація (припущення). Чат-боти можуть писати новинні статті, есе та сценарії, які поширюють теорії змови, згладжуючи людські помилки, як-от поганий синтаксис і неправильний переклад, і просуваючись за межі легко виявлених завдань копіювання та вставки. Неправдиві наративи, що поширюються в мережі Інтернет, регулярно шкодять не тільки бізнесу, але й державним інтересам. Наприклад, у 2020 році в Інтернеті поширилася теорія Qanon про те, що продавець меблів "Wayfair" був пов'язаний з торгівлею дітьми через випадковий збіг назв деяких його предметів меблів і назв зниклих дітей. У результаті користувачі соціальних мереж опублікували адресу та зображення штаб-квартири компанії і профілі співробітників, а також почали переслідувати генерального директора. За такою логікою тепер один поганий актор, який має доступ до ефективного чат-бота, може створити потік людських дописів, подібних до тих, які націлені на "Wayfair", і втратити їх в Інтернеті, потенційно завдавши шкоди репутації та оцінці сторонніх компаній. Більше того, зловмисники можуть навчити моделі ШІ фальшивою інформацією, вводячи в їхні моделі брехню, яку потім поширюють відповідні моделі. Управління ризиком дезінформації є досить складним, а тому державний сектор та компанії повинні планувати небезпечні ризики дезінформації, як-от: планування кібератак або кризових подій, активно передавати свої повідомлення, стежити за тим, як їх сприймають в соціальних мережах та Інтернеті та бути готовими реагувати у разі настання кіберінциденту.

З одного боку, ШІ використовується для боротьби з пропагандою, а з іншого – він сам є інструментом її поширення. ШІ розвивається й удосконалюється і за його допомогою здійснюється боротьба з поширенням пропаганди та дезінформації. Але водночас особи, зацікавлені в поширенні цієї пропаганди також можуть використовувати ШІ. За його допомогою популяризують різноманітні теорії змови, втручаються у вибори та виправдовують війни. У всіх кампаніях впливу поширювачі пропаганди певним чином використовують ботів, частково алгоритми ШІ, за допомогою яких соціальні мережі помилково вважають повідомлення цікавим і ще більше сприяють його поширенню, а частково – звичайних людей, які через брак медіаграмотності починають вірити в цей маніпулятивний контент, інтерпретують його та поширюють далі.

Надання компанією OpenAI безкоштовного доступу до свого чат-бота ChatGPT, публікація в соцмережах великої кількості зображень, створених з допомогою нейромереж, зробили цей інструмент як ніколи близьким для пересічних Інтернет-користувачів. Це актуалізувало дискусії про ризики та можливості, які створює ШІ під час інформаційних воєн. Дослідження аналітичного центру NewsGuard, проведене в січні 2023 року, виявило, що популярний чат-бот ChatGPT здатен генерувати тексти, що розвивають наявні конспірологічні теорії та включають в їх контекст реальні події. Цей інструмент має потенціал для автоматизованого розповсюдження (за допомогою ботоферм) великої кількості повідомлень, тему і тональність яких визначатиме людина, а безпосередній текст – генеруватиме ШІ. За допомогою цього бота можна створювати дезінформаційні повідомлення, у тому числі, засновані на наративах кремлівської пропаганди, формулюючи відповідні запити.

Саме тому протидія поширенню штучно згенерованого неправдивого контенту – це новий виклик та кіберзагроза, яку доцільно нівелювати. ШІ має достатній потенціал для створення фото-, аудіо- і відеофейків, а також для роботи ботоферм. ШІ може замінити

значну частину персоналу на російських “фабриках тролів”, Інтернет-бійців, які провокують конфлікти в соцмережах та створюють ілюзію масової підтримки кремлівських наративів користувачами. Так, держава-агресор намагається за рахунок ботоферм, які розповсюджують проросійські наративи в мережі Інтернет, розхитувати ситуацію в інформаційному просторі нашої держави саме за допомогою технологій ШІ.

Невипадково експерти в США переконливо стверджують, що вороги активно використовують генеративний ШІ з метою проведення дезінформаційних кампаній та прагнуть зробити їх більш правдоподібними. Це викликає занепокоєння, оскільки більшість пропагандистських кампаній досягають мети, тому що вони спрямовані на маніпулювання алгоритмами. Поступово ШІ стає доступнішим, його використання стає дешевшим і такі держави, як КНР та рф починають інвестувати в його використання для пропаганди та поширення дезінформації. Афілійовані з Китаєм та рф операції впливу досягають більших результатів у соціальних мережах, оскільки вони покладаються на ШІ, і на реальних людей для створення індивідуального автентичного контенту. Таким чином, ШІ з усіма своїми плюсами для людства стає маніпулятивним інструментом держави-агресора. В сучасних умовах російська політична влада зрозуміла, що ШІ дозволив автоматизувати поширення дезінформації, поставивши його фактично на системний конвеєр.

У США ще в 2010 році армія ботів використовувалася задля дезінформаційних кампаній проти сенаторів, які у підсумку програли свої виборчі перегони через неправдиві заяви про те, що вони, начебто, виступають проти релігії або проти певних соціально уразливих груп населення. Проте в сучасних умовах відбувається перехід від простих форм незграбних знеособлених поодиноких ботів до більш цілеспрямованих угруповань, які розраховані на конкретну цільову аудиторію користувачів та спеціалізуються на набагато ширшому спектрі соціальних платформ.

З метою запобігання та боротьби з дезінформацією, Європейська Комісія звернулася до 44 компаній та організацій, серед яких “Google”, “Facebook”, “Microsoft”, які підписали Кодекс практики щодо онлайн-дезінформації з пропозицією маркувати контент, створений за допомогою ШІ [14]. Підписанти, які інтегрують генеративний ШІ у свої сервіси, такі як “Bing”, “Chat” в Microsoft і “Bard” у Google, повинні створити необхідні гарантії, щоб ці сервіси не могли використовуватися зловмисниками для створення та спотворення дезінформації. Очікується, що компанії, які мають послуги з потенціалом поширення дезінформації, створеної ШІ, повинні запроваджувати технологію для розпізнавання такого контенту та чіткого позначення її для користувачів. Законодавці в ЄС прагнуть позначати діпфейки та інший контент, створений за допомогою ШІ, щоб звичайні користувачі відразу могли зрозуміти, що це створено саме автоматизованою машиною. На цьому фоні масштабно відбувається автоматизоване поширення пропаганди та дезінформації, які є новими викликами для цивілізованого світу. Компанії-власники соціальних мереж останнім часом, скорочують команди модерації контенту, а пропаганда дедалі активніше використовує саме чат-ботів. Вони створюють ілюзію популярності контенту, який в іншому випадку ніколи б не набув популярності, й виглядає так, ніби багато людей підхоплюють цей контент, поширюють його та вірять у нього. Тоді як його поширенню сприяють саме технології ШІ. Маніпулятивні повідомлення, згенеровані ШІ, удосконалилися настільки, що їх дедалі важче відрізнити від реальних. Зловмисники, зацікавлені в дезінформації за допомогою ШІ, можуть повністю автоматизувати як її генерування, так і її поширення, що є головною небезпекою, яку несе ця технологія, і є основним викликом технічного прогресу.

Етичні ризики. Існують фундаментальні чинники етичних ризиків застосування ШІ, які засновані на певних загрозах основним правам людини у кіберпросторі. Суцільному запровадженню ШІ та іншим цифровим технологіям запобігає низький рівень довіри громадян до його алгоритмів, а також відсутність зрозумілих етичних правил використання ШІ. Етика в ШІ переважно залежить від його виробників та зовсім частково – від споживачів або сервісних компаній. Адже, як і інша інтелектуальна власність, унікальні алгоритми, створені розробниками, не повинні розкриватися повністю (окрім відкритого коду або випадків, обговорених у договорі на розробку). Одночасно існують галузі, де, залежно від застосування, ШІ може спричиняти істотніші ризики і потребуватиме обмежень глибини застосування та встановлення додаткового контролю. Щоб уникнути потенційних порушень етичних зобов'язань, суб`єкти використання ШІ повинні забезпечувати відповідність будь-якого використання його інструментів етичним нормам і відповідним професійним кодексам. У зв'язку з цим етичні ризики пов'язані із необхідністю встановлення загальних правил використання ШІ. Для того, щоб система ШІ була справедливою, необхідно виключити упередженість у вихідних даних, на яких ШІ навчається. Адже як демонструє практика, навіть при ретельній підготовці даних це не завжди буває реальним. Упереджене схвалення рішень ШІ призводить до руйнівних наслідків, а також спричиняє шкоду як громадянам, так і державі у цілому, необґрунтовано обмежуючи можливості деяких осіб робить свій внесок у розвиток економіки або в інших галузях. Як переконливо засвідчує сучасний досвід, окрім того, мінімізація упередженості в системах ШІ є критичною, в іншому випадку втрачається довіра до цих систем [15].

Таким чином, чат-боту жодним чином не можна довіряти на 100 %, оскільки він ніколи не зізнається, що чогось не знає. Натомість, побудує свій варіант "правильної" відповіді з наявних даних. Оскільки чат-боти створюються шляхом збору величезних обсягів інформації з мережі Інтернет (зокрема і упередженої), інформація, яку він "повертає", буде такою ж самою упередженою. Тому доцільно не лише її перевіряти на змістовність та правдивість відповіді, але й протистояти вбудованій упередженості з боку чат-боту. Проблеми безпеки та надійності ШІ мають безпосереднє відношення до етичних ризиків та мають декілька аспектів.

Перший аспект – технічний, стосується надійності та безпеки технічних (програмно-технічних) систем загалом. У звичайних системах існує менший ризик помилок та технічних збоїв, проти сучасних – інтелектуальних. Помилка у програмному забезпеченні систем управління ядерним реактором, найбільш загрозлива, аніж помилка ШІ у випадку видачі споживчого кредиту.

Другий аспект стосується змісту роботи самої інформаційної системи. Оскільки ШІ займається вирішенням формалізованих завдань, використовуючи при цьому різного роду евристичні методи, то в цьому контексті проблема надійності тісно пов'язана з проблемою прозорості ШІ. У випадку, коли прозора система розпочинає працювати некоректно або ненадійно, то розробники можуть швидко знайти передумови та причини помилки.

Третім аспектом виступають проблеми програм-радників, які мають безпосереднє відношення до етики. Ця проблема актуалізується не тільки для інформаційних систем. Експериментальні дослідження демонструють, що в умовах невизначеності та дефіциту часу у користувачів виникає значна довіра до систем ШІ: люди довіряють їй більше, аніж собі. Довіра до ШІ зростає, якщо програма коментує свої дії. Цілком покладаючись на систему ШІ, людина – користувач, рідше приймає рішення самостійно та свідомо, що сприяє як ризику помилок так і втраті кваліфікації. У зв'язку з цим важливо, щоб на

шкідливих виробництвах, у медицині, кібербезпеці спеціалісти зберігали високий рівень кваліфікації. Якою б досвідченою та розумною не була би система ШІ, рівень людського фактору є конче необхідним [16]. Тому саме суцільна стандартизація має на меті підвищити надійність систем ШІ, запровадити мінімальний рівень, нижче за який система не може вважатися надійною. У той самий час, гранична межа не повинна перевішувати можливості сучасних технологій, щоб розробники не були обмежені у процесі створення нових революційних рішень. Невипадково, експерти та аналітики давно б'ють на сполох стосовно того, що неконтрольований розвиток систем ШІ без орієнтації на етику взаємодії між машинами та людьми наражає сучасні постіндустріальні суспільства та людську цивілізацію на істотні ризики в довготерміновій перспективі. В сучасних умовах держави світу не мають механізмів контролю за дотриманням глобальними транснаціональними корпораціями будь-яких законодавчих норм у сфері створення та використання етичних систем ШІ.

Перша спроба врегулювати на міжнародному рівні питання етичних норм під час використання систем ШІ відбулася 25 листопада 2021 року. У цей день, під час проведення чергової конференції ЮНЕСКО за участю представників 193 країн світу в Парижі, генеральний директор Одрі Азуле представила перший глобальний стандарт етики ШІ. Згідно із задумом, цей стандарт має рекомендаційний характер та спрямований на забезпечення реалізації переваг технологій ШІ для потреб суспільства і одночасного зменшення ризиків, пов'язаних із ним. Зокрема, у документі містяться рекомендації щодо: захисту персональних даних; заборони контролю з боку систем соціальної оцінки та масового спостереження; підтримки різноманіття (проявів людської особистості) та інклюзивності; прозорості алгоритмів ШІ та їхньої "підзвітності" експертам, іншим особам; захисту довкілля. Було проголошено, що розробки систем ШІ повинні бути підпорядковані принципам верховенства права, запобігати заподіянню шкоди людині та довкіллю, а за умов, коли така шкода все ж таки заподіяна й настала, повинні спрацьовувати механізми юридичної відповідальності та повне відшкодування збитків [17].

Оскільки чат-боти на базі ШІ стають дедалі більш популярними, важливо враховувати етнічні наслідки їхнього масштабного використання. Однією з основних етичних проблем з чат-ботами на базі ШІ є дотримання вимог конфіденційності. Так як ці боти зберігають та аналізують дані про користувачів, то існує ризик того, що інформація про них може бути передана третім особам або використана для інших цілей, у тому числі й протиправних. Ще однією етичною проблемою є можливість використання чат-ботів зі ШІ для маніпулювання свідомістю людей. Цих ботів можна налаштувати на використання "зручної" мови та тактики з метою впливу на свідомість, поведінку та рішення людей. Можливості ШІ сьогодні такі, що стає незрозумілим, дискусію у фейсбуці або телеграмі здійснює з користувачем людина або програма. І навіть пересічний користувач не помічає до третього або навіть п'ятого коментаря що він фактично дискутує з роботом, який навчений реагувати на заперечення або аргументи.

Таким чином, чат-боти, які використовуються за допомогою ШІ повинні бути безпечними та функціонально діяти згідно із етичними нормами. Задля забезпечення конфіденційного та безпечного використання чат-ботів на основі ШІ доцільно: відстежувати підозрілу активність; обов'язково встановлювати безпечну двофакторну аутентифікацію, що сприятиме посиленню захисту даних; запроваджувати шифрування, щоб запобігти несанкціонованому сторонньому доступу шляхом використання протоколу з набором криптографічних алгоритмів, що має на увазі більш безпечний

зв'язок рівня захищених сокетів (SSL-сертифікатів) або протоколу захисту безпечного транспортного рівня (TLS) з метою забезпечення наскрізного шифрування.

Ризики валідації. В контексті ШІ стабільність досить важлива, оскільки навіть незначні помилки у моделях можуть мати серйозні наслідки. Стабільність системи ШІ залежить від її здатності реагувати на невизначеності, адаптуватися до умов, які динамічно та швидкоплинно змінюються. Це вимагає впровадження комплексного підходу, який поєднує знання у сфері машинного навчання та системної інженерії. Під час використання чат-ботів необхідно не просто приймати інформацію, згенеровану ШІ, як правдиву, а доцільно вживати заходів для перевірки відповідей, перш ніж включати їх у будь-який робочий продукт, дію чи бізнес-рішення. Так, наприклад, розробник чат-боту ChatGPT OpenAI запустив безкоштовний детектор, який створений ШІ. Цей веб-інструмент надає змогу визначити, ким написаний текст: людиною чи машиною. Таким чином, одним з ключових факторів стабільності ШІ є валідація та тестування моделей, що передбачає підтвердження працездатності таких моделей за рахунок проведення перевірок на адекватність в сучасних умовах. Завдання валідації – оцінити, наскільки точно модель оцінює можливі ризики. Тобто моделі ШІ мають бути ретельно протестовані для виявлення потенційних уразливостей та забезпечення їхньої успішної роботи у різноманітних сценаріях. Це включає розробку строгих фреймворків, проведення масштабних симуляцій та реальних експериментів. Первинна валідація проводиться на етапі розробки моделі: аналізуються її логіка, поведінка, достатність вихідних даних, можливість контролю тощо. Після вводу моделі в експлуатацію здійснюється постійний моніторинг з метою виявлення помилок, недоліків, недоопрацювань та проблем. Результатом проведеної валідації може бути один з таких висновків: модель повністю відповідає стандартам, адекватна та може бути використана без змін; модель в цілому адекватна, проте потребує певного доопрацювання; модель неадекватна, містить помилки, недоліки та вимагає кардинального доопрацювання. Основні труднощі проведення валідації пов'язані з тим, що для цього не існує розробленого універсального методу. Найпоширенішим помилковим уявленням про ШІ є те, що він є синонімом автоматизації. Адже автоматизовані системи повинні бути налаштовані вручну для виконання монотонних і повторюваних завдань, тоді як системи ШІ можуть адаптуватися самостійно, коли мають дані для обробки інформації. Хоча ШІ отримує переваги від деяких аспектів автоматизації, він виходить за рамки простого виконання завдань.

Між тим існують непоодинокі приклади недоліків, пов'язаних із ШІ. Наприклад, на початку березня стався витік історії чатів та платіжних даних деяких користувачів ChatGPT від OpenAI, що навіть змусило компанію 20 березня 2023 року тимчасово вимкнути чат-бот. Згодом компанія виявила, що помилка в бібліотеці з відкритим кодом дозволила деяким користувачам бачити заголовки з історії чату іншого активного користувача, що є підтвердженням того факту, що чат-боти також можуть помилятися. Після детальнішого дослідження було виявлено, що та сама помилка могла стати причиною ненавмисної доступності інформації про оплату користувачів ChatGPT Plus, які були активними протягом певного часу. Таким чином, завантаження даних компанії в чат-бот означає, що ймовірно відбувається несанкціонована їхня відправка третій стороні і втрачається контроль над ними. Зокрема компанії-розробники активно використовують дані для навчання та вдосконалення своєї моделі ШІ, які можуть випадково перебувати у публічному доступі.

Одночасно валідація призводить до появи більш удосконалених моделей чат-ботів, керованих ШІ. Так, у листопаді 2023 року Ілон Маск представив свій новий проект, який

покликаний змінити ринок ШІ, а саме чат-бот Grok. Однією з ключових особливостей чат-боту Grok є його доступ у реальному часі до даних соцмережі X, що надає йому "фундаментальну перевагу" перед іншими моделями ШІ. На відміну від багатьох аналогів, які навчаються на основі Інтернет-архівів, Grok може оперативно аналізувати актуальні події та тренди на постійній основі. Чат-бот обіцяє відповідати на запитання з часткою дотепності та навіть може використовувати сарказм. Цей чат-бот буде відповідати на гострі запитання, які відхиляються більшістю інших систем ШІ, при цьому головною перевагою Grok є той факт, що він оновлює знання про світ у реальному часі, використовуючи для цього дані з платформи X. На переконання розробників, Grok значно перевершує можливості багато ШІ-моделей, навчених з використанням значно більшого обсягу даних і обчислювальних ресурсів. На відміну від багатьох аналогів, які навчаються на основі Інтернет-архівів, Grok може оперативно аналізувати актуальні події та тренди. Наявність особистих якостей робить Grok унікальним серед інших ШІ та дає йому змогу відповідати на складні запитання, які інші моделі могли би проігнорувати. Очікується, що чат-бот Grok має стати потужним конкурентом ChatGPT. За наслідками завершення тестування, цей чат-бот буде доступний передплатникам пакету "Premium" соціальної мережі X, вартість якої становить $16 на місяць за веб-версію [19].

Таким чином, завдяки впровадженню в сучасні реалії чат-ботів на базі генеративного ШІ, користувачі вірогідно зіткнуться як з потенціалом суттєвих переваг, так і з ризиками, пов'язаними з використанням цих передових сучасних технологій. Масштабне використання чат-ботів на основі технологій ШІ тісно пов`язано із настанням потенційних ризиків, що вимагає прискорення розробки відповідних заходів їхньому запобіганню. Чат-боти зі ШІ, які колись у минулому вважалися просто автоматизованими розмовними програмами, відтепер можуть навчатися та вести розмови, які майже не відрізняються від людських, що відкриває нову еру технологічної революції. Однак небезпеки чат-ботів ШІ настільки ж різноманітні, що вимагають, у першу чергу: певної обережності під час використання чат-ботів і генеративного ШІ; чіткого розуміння того, що чат-боти можуть також робити помилки; підвищення захисту кібербезпеки ІКТ систем від загроз, пов`язаних із ШІ; переконання, що ШІ використовується відповідно до етичних норм та відповідних професійних стандартів; перевірку результатів ШІ на предмет їхнього упередженого та дискримінаційного впливу; визначення алгоритмів протидії дезінформації, яка може бути поширена за допомогою систем ШІ тощо.

Так, стурбованість щодо етичних та юридичних наслідків застосування передових технологій на кшталт ChatGPT неодноразово анонсував Європол, попереджаючи про негативні наслідки потенційного використання цих систем для різних типів Інтернет-шахрайства – від фішингу до тяжких кіберзлочинів. Нещодавно національний центр кібербезпеки Великобританії (NCSC) висловився про наявну загрозу та необхідність попередження ризиків використання чат-ботів, керованих ШІ, стверджуючи, що ці системи можливо обманом змусити виконувати шкідливі та протиправні завдання [19].

Наприклад, чат-бот, заснований на базі ШІ, який використовує банківська установа, може обманом змусити здійснити несанкціоновану транзакцію, якщо хакер правильно структурує свій запит. Саме тому вплив ШІ на стан забезпечення кібербезпеки все ще перебуває в центрі прискіпливої уваги переважної більшості урядів країн світу, оскільки геометрично зростає кількість зафіксованих випадків злочинного використання цих технологій хакерами та зловмисниками. Компанії, які інтегрують генеративний ШІ (наприклад, ChatGPT), можуть використовувати різні стратегії для

мінімізації ризиків, покращення надійності, точності і довіри до вихідних даних. Так, зокрема запровадження систем модерації допомагає фільтрувати недоречний або небезпечний контент або зміст, згенерований моделлю. Це створює додатковий рівень контролю для того, щоб забезпечити відповідність згенерованих відповідей попередньо визначеним стандартам і критеріям.

Окрім перелічених ризиків, використання чат-ботів, керованих ШІ, існує загроза використання російських чат-ботів з метою збору та узагальнення інформації про українські війська, здійснення російської пропаганди та дезінформації, націленої на українське суспільство.

Висновки.

Роль та значення ШІ у питаннях забезпечення кібербезпеки без перебільшення не можна недооцінювати. Саме ШІ допомогає розширити можливості виявлення загроз шляхом аналізу великих обсягів та масивів даних, виявляючи потенційні кіберзагрози. Він також може аналізувати шаблони в даних, щоб виявляти підозрілу поведінку та технологічні аномалії, які можуть свідчити про кібератаку, ідентифікувати зловмисне програмне забезпечення, фішинг та інші кіберзагрози, полегшуючи аналітикам безпеки швидке, оперативне та ефективне реагування. Цілком логічно, що ШІ покращує виявлення загроз, визначаючи закономірності, які люди-аналітики не можуть отримати, завдяки своїй здатності навчатися та адаптуватися до швидкоплинних змін у поведінці зловмисників. ШІ сприяє виявленню інноваційних невідомих загроз і є потужним союзником у боротьбі зі спеціалізованими кібератаками (APT – Advanced Persistent Threaker) [20]. Таким чином, остання тенденція сучасності – поява чат-ботів, керованих ШІ. Основні ризики, які постають у зв`язку з цим можливо умовно поділити на такі групи.

Перша група – це ризики для особистих даних і приватності. Мовні моделі, по-перше, можуть робити досить багато узагальнень, поєднань інформації і висновків з отриманих даних, які можуть бути недоступними для людини, що аналізує інформацію. Наприклад, з наборів даних про поведінку людини у соціальних мережах та Інтернеті загалом, тих же запитів до чат-ботів великих мовних моделей, можна робити глибокі висновки щодо її психічного стану, звичок, розкладу дня тощо. Поєднання великих масивів даних та їх аналіз з боку ШІ створюють додаткові загрози приватності, які не очевидні для користувачів, оскільки ці дані не створюють загроз.

Друга група – ризики сегрегації. Тобто мовні моделі, за рахунок того, як саме і на основі яких даних навчаються, можуть надавати перевагу окремим соціальним групам. Вони будуть краще працювати з цими групами, а іншим приділяти менше уваги і маргіналізувати їх, вилучати певні меншини (етнічні або інші) та відтворювати людські упередження щодо цих меншин, тобто існує нахил до певної дискримінації.

Третя група ризиків – це перекручування, неточність та неправдивість інформації. Мовні моделі, засновані на ШІ, не досить добре відрізняють факти від вигадок, але при цьому створюють враження, що вони є достовірними співрозмовниками, що може створювати чимало потенційних загроз. Звідси випливають й безпекові ризики: мовні моделі досить зручно використовуються задля масової дезінформації, причому персоналізованої “з людським обличчям”, орієнтованої на конкретні групи людей, наприклад, шляхом здійснення коментування у соцмережах або у медіа. Це також можуть бути випадки персоналізованого шахрайства, коли людина отримує повідомлення з урахуванням методів соціальної інженерії, які спрямовані саме на неї, коли ШІ використовує індивідуальні слабкості та упередження.

Щоб зменшити окреслені ризики, розробники повинні запроваджувати попередню обробку та контроль введення, регулювання конфігурації та поведінки моделей ШІ, механізмів навчання та вдосконалення, покращувати контекст та набуті знання. Сукупно такі заходи мають допомогти суттєво мінімізувати ризики, значно покращити якість та достовірність результатів. Важливим та перспективним напрямком залишається розробка нормативних вимог та подальша їх уніфікація щодо використання технологій ШІ у сфері кібербезпеки, особливо в умовах правового режиму воєнного стану в Україні. Також необхідним є унормування правової регламентації як на державному, так і міжнародному рівнях, використання технологій ШІ у сфері кібербезпеки з метою недопущення порушень прав людини на приватність, запобігання витоку конфіденційних даних, системної боротьби з дезінформацією у соціальних мережах тощо.

Використана література

1. Неретін О., Харченко В. Забезпечення кібербезпеки систем штучного інтелекту: аналіз вразливостей, атак і контрзаходів. *Information systems and networks.* 2022. № 12. С. 7-20.

2. Савченко В.А., Шаповаленко О.Д. Основні напрями застосування технологій штучного інтелекту у кібербезпеці. *Сучасний захист інформації.* 2020. № 4 (44). С. 6-11.

3. Стьопочкіна І.В., Новіков О.М. Методи штучного інтелекту в кібербезпеці: навч. посіб. для здобувачів. Київ: КПІ ім. Ігоря Сікорського, 2022. 82 с.

4. Гуржій С.В. Особливості використання штучного інтелекту у питаннях забезпечення кібербезпеки. *Інформація і право.* № 4(47)/2023. С. 207-216.

5. Токарєва К.С., Савліва Н.О. Особливості правового регулювання штучного інтелекту в Україні. *Юридичний вісник.* 2021. № 3(60). С. 148-153.

6. Цяпа С.М. Огляд зарубіжних законодавчих ініціатив стратегічного використання технологій ШІ в сучасних умовах. *Інформація і право.* № 2(37)/2021. С. 51-59.

7. Sebastian, G. (2023). Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk?: An Exploratory Study. [IJSPPC]. *International Journal of Security and Privacy in Pervasive Computing,* 15(1), 1–11. DOI:10.4018/ IJSPPC.320225

8. Zhou, P. (2023). Unleashing chatgpt on the metaverse: Savior or destroyer? URL: arXiv preprint arXiv:2303.13856

9. Корнєєва С.Р. Вплив застосування технологій штучного інтелекту на реалізацію та захист прав людини. *Аналітично-порівняльне правознавство*. – (Електронне наукове видання). 2021. № 4. С. 392-394.

10. Зачек О.І., Дмитрик Ю.І., Сеник В.В. Роль штучного інтелекту в підвищенні ефективності правоохоронної діяльності. *Науковий вісник Львівського державного університету внутрішніх справ.* 2023. № 3. С. 148-156.

11. DNS cache poisoning. URL: https://owasp.org/www-pdf-archive/DNS_Cache_Poisoning (OWASP_GHANA).pdf

12. Трофименко О.Г. Сфери застосування чат-ботів. *Інформаційне суспільство: проблеми та перспективи*: матеріали VII Всеукраїнської науково-практичної конференції, м. Одеса, 20 трав. 2022 р.. Одеса, 2022. С. 68-71. URL: http://dspace.onua.edu.ua/bitstream/handle/11300/18203

13. ChatGPT може зливати секретні корпоративні дані: спеціалісти з кібербезпеки б’ють на сполох. URL: https://24tv.ua/tech/chat-boti-zi-shtuchnim-intelektom-stvoryuyut-riziki-dlya-konfidentsiynosti_n2297236

14. The 2022 Code of Practice on Disinformation EU. URL: https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

15. The Danger of AI Chatbots – And How to Counter Them. URL: https://www.unite.ai/the-dangers-of-ai-chatbots-and-how-to-counter-them

16. AI in Cyber Security: Risks of AI: URL: https://www.malwarebytes.com/cybersecurity/basics/risks-of-ai-in-cyber-security

17. Recommendation on the Ethics of Artificial Intelligence UNESCO, adopted 25.11.2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137

18. Ілон Маск запускає Grok – чат-бот, який має стати конкурентом ChatGPT. URL: https://ain.ua/2023/11/06/ilon-mask-zapuskaye-grok

19. NCSC попереджає: чат-боти з ШІ несуть ризики кібербезпеці. URL: https://proit.org.ua/ncsc-popieriedzhaie-chat-boti-z-shi-niesut-riziki-kibierbiezpietsi

20. What is an Advanced Persistent Threat (APT)? URL: https://www.cisco.com/c/en/us/products/security/advanced-persistent-threat.html

~~~~~~~~~~~~~ * * * ~~~~~~~~~~~~~
~~~~~~~~~~~~~