КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (22), 2023
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2023.22.618
УДК 004.89:[004.056:007]

Скіцько Олексій
кандидат технічних наук, старший науковий співробітник
Національна академія Служби безпеки України, м. Київ, Україна
ORCID 0000-0003-4122-0889
oiskitsko@gmail.com

Складанний Павло
кандидат технічних наук, доцент,
завідувач кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
Київський університет імені Бориса Грінченка, м. Київ, Україна
ORCID 0000-0002-7775-6039
p.skladannyi@kubg.edu.ua

Ширшов Роман
Національна академія Служби безпеки України, м. Київ, Україна
ORCID 0000-0003-3534-8736
signorum@gmail.com

Гуменюк Михайло
кандидат юридичних наук
Національна академія Служби безпеки України, м. Київ, Україна
ORCID 0009-0006-3118-721X
ansysdempel@gmail.com

Ворохоб Максим
викладач кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка
Київський університет імені Бориса Грінченка, Київ, Україна
ORCID 0000-0001-5160-7134
m.vorokhob@kubg.edu.ua

ЗАГРОЗИ ТА РИЗИКИ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. В статті аналізуються переваги застосування Штучного Інтелекту (ШІ) в різних галузях та ризики впливу на виконання завдань забезпечення інформаційної безпеки та кібербезпеки, як невід'ємних складових національної безпеки. Визначено, що розвиток ШІ став ключовим пріоритетом для багатьох країн, та водночас, виникли питання щодо безпечності цієї технології та наслідків її використання. Поширення сфери застосування ШІ на об'єкти критичної інфраструктури, складність верифікації створених цими системами інформаційних ресурсів та рішень, загрози небезпечного впливу результатів їхнього функціонування на безпеку людини, суспільства та держави призводить до виникнення ризиків, пов'язаних з використанням ШІ. Відсутність прозорих методів перевірки запропонованих систем ШІ висновків та рекомендацій утворює джерело невизначеності щодо їх вірності і практичної цінності. Це фактично означає, що системи ШІ можуть бути частиною сукупності заходів інформаційної війни, які спрямовані на поширення сумнівних неперевірених відомостей та звичайних фейків. Застосування технології штучного інтелекту здатне покращити рівень комп'ютерної безпеки. В роботі розглядається механізм оцінки ризиків від використання ШІ в різних галузях та способи їх обробки. Запропоновані підходи до використання системи штучного інтелекту для ідентифікації та оцінки ризиків, які виникають як наслідок використання систем штучного інтелекту. Штучний інтелект відіграє ключову роль у забезпеченні національної безпеки, а його застосування в різних галузях сприяє покращенню ефективності, проте, існує нагальна потреба в розробці механізмів оцінки ризиків використання систем штучного інтелекту. Визначено, що одним з найважливіших заходів є створення системи управління ризиками штучного інтелекту, на якому має базуватися регуляторна політика держави у цій галузі.

Ключові слова: штучний інтелект; національна безпека; оцінка ризиків; управління ризиками.

ВСТУП

Постановка проблеми. Державна безпека, національна безпека і оборона, а також багатосторонній розвиток суспільства залежать від розвитку високих технологій. Специфічною галуззю, що грає важливу роль в цьому процесі, стає штучний інтелект (*Artificial Intelligence,* AI).

Принципи та алгоритми функціонування Систем Штучного Інтелекту (далі — СШІ) переважної більшості суспільства практично невідомі. За суттю СШІ сприймаються як деякі «чарівні чорні скриньки», які здатні розуміти природню мову людини, музичні опуси або графічні зображення та адекватно реагувати на запитання користувачів шляхом надання статистично вірної відповіді.

При цьому, звичайно, користувачі системи, що отримують результат відповідно до завдання, поставленого такій системі, не розуміють джерел формування відповіді і методів розв’язання завдання [1].

З одного боку, на відміну від звичайних обчислювальних систем, в випадку СШІ спостерігається ефект непередбачуваності (частково — не тривіальності) результатів її «роздумів», що в загальному випадку є однією з ознак творчості та інноваційної діяльності, яка притаманна людині.

З іншого боку, відсутність прозорих методів перевірки запропонованих СШІ висновків та рекомендацій утворює джерело невизначеності щодо їх вірності і практичної цінності. Це фактично означає, що СШІ можуть бути частиною сукупності заходів інформаційної війни, які спрямовані на поширення сумнівних неперевірених відомостей та звичайних фейків. СШІ може стати потужним інструментом в інформаційних війнах, створюючи більш переконливі та цільові фейкові новини, а також автоматизуючи їх поширення. Звернемо увагу, що платформа розповсюдження контенту, яка використовує алгоритми рекомендацій із підтримкою штучного інтелекту, була використана для визначення пріоритетності вмісту з метою маніпулювання емоціями, переконаннями та поведінкою [2].

У [3] відмічене, що потужні системи штучного інтелекту слід розробляти лише в тому випадку, якщо ми впевнені, що їхній ефект буде позитивним, а ризики керованими.

Таким чином, поширення сфери застосування СШІ на об’єкти критичної інфраструктури, складність верифікації створених цими системами інформаційних ресурсів та рішень, загрози небезпечного впливу результатів їхнього функціонування на безпеку людини, суспільства та держави призводить до виникнення ризиків, пов’язаних з використанням СШІ, а це висуває вимогу формування процедур виявлення та обробки таких ризиків, що може мати визначальне значення для майбутнього суспільства та забезпечення національної безпеки.

Аналіз останніх досліджень та публікацій. На поточний час серед поширених систем, що доступні широкому загалу суспільства та сприяють формуванню суспільної думки щодо можливостей штучного інтелекту, бачимо наступні СШІ генеративного типу:

- ChatGPT — чат-бот, що створений компанією OpenAI, підтримує діалог з користувачем з використанням природних мов та генерує тексти на задану тему;

– Midjourney (проміжний шлях) — сервіс від однойменної компанії, які виходячи з текстових описів — запитів щодо бажаного зображення генерує його.

На основі вивчення цих засобів вкрай складно зробити висновок щодо перспектив застосування в інтересах державних структур технологій штучного інтелекту, але за умов належного проєктування систем можливо прогнозувати можливість розвитку деяких напрямів, що вимагають від інформаційно-управляючих систем забезпечення певних якостей, притаманних штучному інтелекту.

На основі окремих повідомлень можливо зробити висновок, що системи штучного інтелекту включають різні компоненти, такі як алгоритми машинного навчання, глибокого навчання, нейронні мережі, обробка природної мови, комп'ютерний зір, робототехніка та інші технології. Вони також можуть включати комплекси автоматизації, аналітики даних, системи управління знаннями та інші інструменти для обробки інформації та прийняття рішень.

Які головні переваги від застосування СШІ можливо виділити у порівнянні зі звичайними програмними системами?

У багатьох СШІ застосовуються інтерпретатори природних мов, що суттєво підвищує їх ефективність порівняно із звичайними програмними системами у випадку обробки неструктурованих даних [4], [5].

Застосування в перспективі [6] в СШІ потужних комп’ютерних архітектур з швидкісними процесорами, що здатні підтримувати значну кількість глибоких нейронних мереж, дозволить вирішувати за оперативне придатний час розв’язання складних завдань, які потребують великої кількості обчислень.

Штучний інтелект забезпечує ефективну обробку великих обсягів даних (*big data*) завдяки застосуванню алгоритмів машинного навчання та аналізу даних, що забезпечує автоматизацію процесів їх обробки та аналізу, дає можливість виявляти приховані закономірності та прогнозувати тренди (тенденції) та патерни (зразки), оптимізувати процеси прийняття рішень та створювати інтелектуальні системи управління даними [7]. Таким чином, СШІ сприяє отриманню знань з великих обсягів даних та приймати обґрунтовані рішення на основі їх аналізу.

Застосування технології штучного інтелекту здатне покращити рівень комп’ютерної безпеки, зокрема, в [8] запропоноване використання штучного інтелекту та нейронних мереж для проектування системи захисту комп’ютерної мережі. При цьому, на основі експериментальних досліджень продемонстрований гарний ефект від цього в плані забезпечення безпеки комп’ютерної мережі.

Перелічені властивості технології штучного інтелекту дають підстави прогнозувати її ефективне застосування також для вирішення інших завдань забезпечення кібербезпеки. Зокрема, зважаючи на великі обсяги даних що обробляються ситуаційними центрами [9] та підвищені вимоги щодо їх гарантоздатності, кібербезпеки та оперативності прийняття рішень в кризових ситуаціях вказані центри є першочерговими об’єктами для впровадження технологій штучного інтелекту.

Аналогічна ситуація має у випадку протидії шифрувальним вірусам — вимагачам [10] при цьому успіху заходів протидії зазначеним загрозам сприяє головна особливість штучного інтелекту — здатність розв’язання складних завдань, які потребують великої кількості обчислень, без прямого втручання людини.

Водночас, вибуховий характер розвитку технології штучного інтелекту та її застосування, не дивлячись на позитивні властивості технології, несе великі потенційні ризики [11]. Важливо зазначити, що здатність керувати цими ризиками на думку авторів

відкритого листа [3] потребує спільної розробки та впровадження набору загальних протоколів безпеки для передового проектування та розробки штучного інтелекту, які ретельно перевіряються та контролюються незалежними зовнішніми експертами.

Необхідність визначати та регулювати ризики, пов'язані з використанням систем ШІ шляхом прийняття законодавчих актів та стандартів визнали Європейський Союз, Сполучені Штати Америки та інші країни світу.

Зокрема, Європейський парламент схвалив основні положення [12], що утворять основу майбутнього закону, якій визначатиме правила в сфері штучного інтелекту. Документом встановлюється високий рівень ризику застосування штучного інтелекту в галузі критичної інфраструктури, громадського порядку, освіті та управління міграцією. При цьому особливі вимогі висуваються до СШІ, що забезпечують генерацію аудіо, відео та іншого контенту.

В США Національний інститут стандартів та технологій (NIST) визначив рамковим документом AI RMF 1.0 [13] орієнтири для покращення здатності враховувати важливі аспекти щодо надійності під час проектування, розробки, використання та оцінки продуктів, послуг і систем, що засновані на технології штучного інтелекту.

Загалом, аналіз законодавчих записів 127 країн за індексом AI показує, що кількість законопроектів, які згадують «штучний інтелект» та були прийняті як закон, зросла з 1 у 2016 році до 37 у 2022 році. Аналіз парламентських записів щодо AI у 81 країні також показує, що згадки про штучний інтелект у глобальних законодавчих процедурах зросли майже в 6,5 рази з 2016 року [14].

Мета статті. Метою статті є визначення механізмів виявлення та обробки ризиків використання штучного інтелекту.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Для визначення можливого впливу загроз що пов'язані з СШІ та мінімізації ймовірності їх реалізації необхідно об'єднання зусиль науковців, дослідників та розробників наукових і освітніх установ, виробничих і промислових структур, державних і громадських організацій, законодавчих та виконавчих органів влади та міжнародної спільноти.

Реалізація таких заходів допоможе створити систему зниження ризиків, що дозволить швидше виявляти, готуватися і реагувати на виклики та загрози, які походитимуть від створення та використання систем штучного інтелекту.

Ризики, що породжені системами штучного інтелекту, багато в чому унікальні. Системи штучного інтелекту, наприклад, можуть бути навчені на даних, які можуть змінюватися з часом, іноді суттєво й несподівано, впливаючи на функціональність і надійність системи у спосіб, який важко зрозуміти. Системи штучного інтелекту та контексти, в яких вони розгортаються, часто складні, що ускладнює виявлення збоїв і реагування на них.

Системи штучного інтелекту за своєю природою мають соціально-технічний характер, тобто на них впливає суспільна динаміка та поведінка людей. Ризики та переваги штучного інтелекту можуть виникати через взаємодію технічних аспектів у поєднанні з суспільними факторами, пов'язаними з тим, як використовується система [13].

Розвиток систем штучного інтелекту створює передумови для посилення існуючих загроз національній безпеці в інформаційній сфері. Ці загрози включають:

1. Посилення кібератак. Штучний інтелект може підвищити ефективність процедур виявлення вразливостей в системах безпеки, виконання атак, маскування їх наслідків, імітацію поведінки людини в окремих фазах кібератаки [10], [15].

2. Утворення каналів витоку інформації з обмеженим доступом. Системи штучного інтелекту можуть бути використані для посилення комп'ютерної розвідки завдяки аналізу шляхом аналізу великих обсягів даних, визначення трендів і патернів, щоб виявити конфіденційні дані про об'єкти, які стосуються національної безпеки, критичну інфраструктуру тощо. Зокрема, завдяки інтерактивній карті, що опублікована в Інтернеті та показує місцезнаходження людей, які використовують такі фітнес-пристрої, як Fitbit, була продемонстрована можливість ідентифікації військових об'єктів США [16].

3. Атаки отруєння даних (*Data Poisoning*, DP) — це цільові атаки з метою модифікації або спотворення даних, що використовуються для машинного або глибокого навчання СШІ, внаслідок чого СШІ отримує небажані навички, які можуть завдати шкоди особі, суспільству та державі. Зазначимо, що процедури навчання СШІ зазвичай передбачають використання великої кількості даних для тренування моделі. Ці дані можуть бути зібрані з різних джерел і можуть містити помилки або неточності. DP — атака використовує ці неточності з метою введення помилкових чи зловмисних даних у навчальний набір [17]–[19].

Зважаючи на те, що СШІ можуть використовуватися для створення роботизованих збройних систем, які можуть самостійно визначати та атакувати цілі без участі оператора [20], DP атака може мати без перебільшення жахливі наслідки.

4. Містифікація даних. Спеціалістами компанії «Vulcan» [21] виявлена схильність генеративної СШІ ChatGPT до створення недостовірних (галюцінованих) фактів і даних (*hallucinated facts and figures*). А саме, ChatGPT в разі запиту рішень для кодування може пропонувати неіснуючі пакети (рис. 1, кроки 1, 2). Ці уявні пакети можуть бути використані зловмисниками для перетворення їх в замаскований шкідливий код, який завантажується в репозиторії кодів (крок 3). Якщо користувач, запитує у ChatGPT рекомендації щодо розробки (кроки 4, 5), в пропозиціях (кроки 6, 7) можуть з'явитися ці шкідливі пакети. Таким чином, ці недостовірні пакети СШІ потенційно перетворяться на канал витоку інформації користувач (крок 8).

Наведений вище перелік загроз не є вичерпним, але дає можливість оцінити складність, глибину та впливовість проблеми. Кожна з цих загроз вимагає глибокого розуміння технологій штучного інтелекту та формування ефективних стратегій для протидії їм.

Для блокування або нейтралізації визначених загроз необхідна реалізація низки заходів, спрямованих на мінімізацію (обробку) ризиків використання систем штучного інтелекту, і в першу чергу — на ідентифікацію ризиків використання СШІ. Як і ризики для інших типів технологій, ризики штучного інтелекту можуть виникати різними способами та можуть бути охарактеризовані як довготермінові чи короткострокові, з високою чи низькою ймовірністю, системні чи локалізовані, а також із сильним чи низьким впливом [13].

ISSN 2663 - 4023

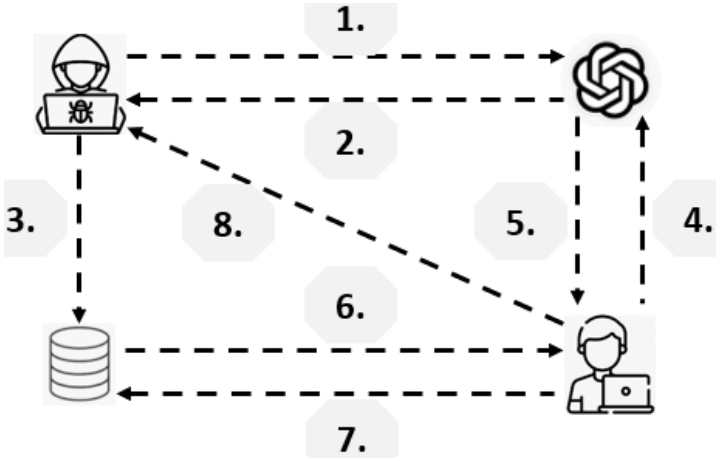

Рис. 1. Реалізація загрози витоку інформації через недійсні пакети СШІ

Загальні принципи та орієнтири для управління ризиками будь-якого типу, розміру чи природи визначені стандартом ISO 31000:2018 Risk management — Guidelines [22], [23]. Основна ідея полягає в тому, щоб створити систематичний і структурований підхід до ідентифікації, оцінки, керування та моніторингу ризиків. Ось деякі ключові етапи та аспекти управління ризиками за стандартом ISO 31000:

1. Встановлення контексту: Необхідно розуміти свій контекст і визначити фактори, що впливають на здатність до досягнення цілей;
2. Ідентифікація ризиків: Ризики ідентифікуються шляхом виявлення подій або ситуацій, які можуть впливати на досягнення цілей;
3. Оцінка ризиків: Визначення ймовірності та впливу ризиків для визначення їхньої значущості. Це допомагає визначити пріоритети для подальшого керування ризиками;
4. Обробка ризиків: В цьому етапі визначаються можливі стратегії обробки ризиків. Це може включати уникнення ризику, зменшення його впливу, передачу ризику і прийняття ризику;
5. Заходи управління ризиками: Розробляються та впроваджуються конкретні заходи для керування ризиками згідно з визначеними стратегіями;
6. Моніторинг та перегляд: Ризики та їхні заходи управління мають бути систематично переглянуті та оцінені для впевненості, що вони залишаються ефективними та актуальними;
7. Звітність та комунікація: Інформація про ризики та їхній стан повинна бути передана відповідним зацікавленим сторонам, включаючи керівництво та інші зацікавлені групи.

Загалом, управління ризиками за стандартом ISO 31000 вимагає системного підходу на всіх етапах життєвого циклу з області оцінювання.

Звернемо увагу, що в [3] для нашого випадку фактично визначена мета управління ризиками: сучасні системи штучного зробити більш точними, безпечними, інтерпретованими, прозорими, надійними, узгодженими, заслуговуючими на довіру і лояльними.

Використання системи штучного інтелекту для управління ризиками, які виникають внаслідок використання інших СШІ може бути ефективним засобом забезпечення безпеки та стабільності. Такий підхід може включати в себе ряд етапів (рис. 2) та функціональних можливостей:

ISSN 2663 - 4023

1. Ідентифікація ризиків: аналіз архітектури системи та алгоритмів, що використовуються, типи даних та інші параметри. Система може визначати слабкі місця, потенційні точки виникнення помилок або зони, де система може взаємодіяти з оточенням;
2. Моніторинг поведінки системи штучного інтелекту: система управління ризиками на базі системи штучного інтелекту може безперервно стежити за поведінкою інших систем управління, аналізуючи їх виходи, метрики та поведінку в реальному часі;
3. Автоматичне виявлення аномалій: використовуючи методи машинного навчання, система може виявляти аномалії або відхилення від норми в поведінці систем штучного інтелекту, що може свідчити про потенційні ризики;
4. Прогнозування ризиків: на основі історичних даних та актуального стану системи управління ризиками прогнозуються потенційні проблеми або непередбачувана поведінка системи штучного інтелекту у майбутньому;
5. Автоматична корекція: у випадках, коли виявлено ризик, система може автоматично вносити корективи в роботу іншої системи штучного інтелекту, наприклад, змінюючи її параметри або обмежуючи її дії;
6. Сценарії «чорної скриньки»: для вивчення і розуміння поведінки систем штучного інтелекту можна використовувати сценарії, де система штучного інтелекту піддається ряду тестів у контрольованому оточенні.
7. Аналіз причинно-наслідкових зв’язків: система може допомогти аналізувати причини певної поведінки системи штучного інтелекту, визначаючи, чи була ця поведінка результатом вхідних даних, алгоритмів, або інших факторів;
8. Зворотний зв’язок і навчання: на основі аналізу ризиків та інцидентів система може навчатися, вдосконалюючи свої методи виявлення та реагування на ризики.

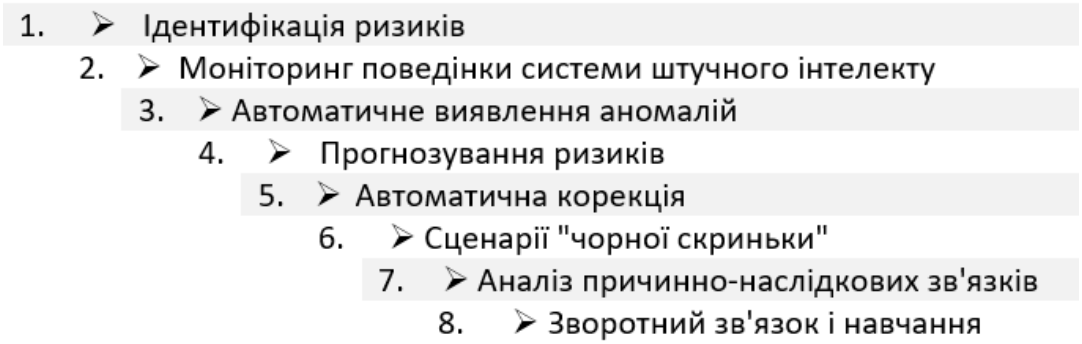

Рис. 2. Етапи використання СШІ для управління ризиками

Одним з найважливіших заходів є створення системи управління ризиками штучного інтелекту, на якому має базуватися регуляторна політика держави у цій галузі.

Система управління ризиками базується на використанні класичного підходу до оцінки ризиків, який наведено нижче:

1. Ідентифікація ризиків: Це початковий етап, де ідентифікуються потенційні ризики, пов’язані з ШІ. Це може включати розробку політик та алгоритмів, використання даних, вплив на користувачів і багато іншого;

2. Оцінка ризиків: Після ідентифікації ризиків вони повинні бути оцінені за їх потенційним впливом та ймовірністю виникнення. Це може включати аналіз чутливості, упередженості, моделювання ризиків або інші методики оцінки;
3. Прийняття рішень щодо ризиків: Після оцінки ризиків слід вирішити, як краще ними управляти. Це може включати прийняття рішень про вдосконалення процесів, модифікацію баз даних та баз знань ШІ або внесення змін в спосіб використання ШІ;
4. Управління ризиками: Це включає в себе виконання дій по управлінню ризиками, які було визначено на попередньому етапі. Це може включати виконання контрольних заходів, навчання персоналу, зміни в дизайні систем та інші заходи;
5. Моніторинг та перегляд ризиків: Ризики слід постійно моніторити та переглядати, щоб впевнитись, що вони залишаються під контролем та що вжиті заходи ефективні. Це може включати регулярний аудит, моніторинг впливу, збір зворотного зв'язку від користувачів та інші механізми моніторингу.

Всі ці етапи мають повторюватись циклічно, оскільки ризики можуть змінюватися з часом. Також, оцінку ризиків необхідно проводити на всіх етапах життєвого циклу системи ШІ.

Реалізація заходів з управління ризиками ШІ пропонується за трьома напрямками: нормативно-правовим, технічним та організаційним.

Нормативно-правові заходи:

1. Визначення та прийняття державної політики в галузі штучного інтелекту;
2. Законодавче регулювання: Створення чітких законодавчих норм, які регулюють розробку та використання ШІ, може бути ефективним способом відповіді на ці загрози. Держава може прийняти закони, які обмежують використання ШІ в автономних збройних системах або встановлюють стандарти безпеки для ШІ в кібернетичних системах;
3. Міжнародне співробітництво: Участь в міжнародних угодах та ініціативах, спрямованих на регулювання ШІ, створенні міжнародних норм і стандартів безпеки для ШІ та забезпечення їх використання в Україні.

Технічні заходи:

1. Розробка безпечних систем ШІ: сприяння на рівні держави розробці та впровадженню безпечних систем ШІ, що включають вбудовані заходи безпеки;
2. Обмеження доступу до даних: Встановлення технічних обмежень на доступ ШІ до даних, таких як персональні дані громадян, що запобігає несанкціонованому використанню цих даних;
3. Створення систем ШІ для проведення аудиту знань прикладних систем ШІ;
4. Розробка механізмів виявлення ознак роботи небезпечних ШІ;
5. Розробка заходів з активної протидії небезпечним ШІ.

Організаційні заходи:

1. Управління ризиками: Створення моделі управління ризиками ШІ, яка містить механізми визначення рівнів загроз та імовірності їх реалізації в різних областях діяльності людини, суспільства, держави;
2. Освіта: Проведення освітніх кампаній для збільшення обізнаності про потенційні ризики, пов'язані з ШІ;

3. Співпраця з приватним сектором: Держава співпрацює з приватним сектором для створення безпечних систем ШІ і розробки ефективних стратегій протидії потенційним загрозам;
4. Створення спеціалізованих державних органів: Створення спеціалізованих органів, які будуть відповідальні за моніторинг та реагування на загрози, пов'язані з ШІ.
5. Організаційне обмеження доступу до масивів даних та спеціалізованих баз знань створених державними установами для використання їх в моделях навчання штучного інтелекту.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Найважливішим етапом в системі управління ризиками, які виникають в наслідок використання систем штучного інтелекту є оцінка ландшафту можливих ризиків та їх ідентифікація. Це ітеративний процес пошуку нових типів ризиків та профілювання їх основних характеристик для подальшої інтерпретації, аналізу та обробки.

Завдання ідентифікації ризиків вирішується як завдання пошуку аномалій в масивах даних про діяльність, що стосується галузі застосування ризик-менеджменту. Аномальні спостереження в таких даних можуть пояснюватися наявністю взаємозв'язків та взаємодій між об'єктами та суб'єктами діяльності, що призводять до появи ще не ідентифікованих ризикових ситуацій та відповідних наслідків, або є потенційними джерелами виникнення таких ситуацій у майбутньому.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bagchi, S., & US, T. C. (2023). *Why We Need to See Inside AI's Black Box*. Scientific American. https://www.scientificamerican.com/article/why-we-need-to-see-inside-ais-black-box/
2. Auchard, E., & Ingram, D. (2018). *Cambridge Analytica CEO claims influence on U.S. election, Facebook* questioned. Reuters. https://www.reuters.com/article/us-facebook-cambridge-analytica-idUSKBN1GW1SG
3. Pause Giant AI Experiments: An Open Letter - Future of Life Institute. (2023). Future of Life Institute.https://futureoflife.org/wp-content/uploads/2023/05/FLI_Pause-Giant-AI-Experiments_An-Open-Letter.pdf.
4. Abdullah, M. F., & Ahmad, K. (2013). The Mapping Process of Unstructured Data to Structured Data. 3rd International Conference on Research and Innovation in Information Systems (ICRIIS), 151–155. https://doi.org/10.1109/ICRIIS.2013.6716700
5. Abdullah, M. F. & Ahmad, K. (2015). Business Intelligence Model for Unstructured Data Management. 5th International Conference on Electrical Engineering and Informatics, 473–477. https://doi.org/10.1109/ICEEI.2015.7352547
6. Venieris, S.; Bouganis, C., & Lane, N. (2023). Multiple-Deep Neural Network Accelerators for Next-Generation Artificial Intelligence Systems. *Computer, 56(3),* 70–79. https://doi.org/10.1109/MC.2022.3176845
7. Xing, J. (2019). The Application of Artificial Intelligence in Computer Network Technology in Big Data Era. 4th International Workshop on Materials Engineering and Computer Sciences, 211–215. https://doi.org/10.25236/iwmecs.2019.044
8. Bian, L. (2023). Design of Computer Network Security Defense System Based on Artificial Intelligence and Neural Network. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-023-10721-9
9. Grechaninov, V., et al. (2022). Formation of Dependability and Cyber Protection Model in Information Systems of Situational Center. *Emerging Technology Trends on the Smart Industry and the Internet of Things, 3149*, 107–117.

10. Hulak, H., et al. (2020). Cryptovirology: Security Threats to Guaranteed Information Systems and Measures to Combat Encryption Viruses. *Cybersecurity: Education, Science, Technique, 2(10),* 6–28. https://doi.org/10.28925/2663-4023.2020.10.628
11. Moskalenko, V.; Kharchenko, V.; Moskalenko A., &; Kuzikov, B. (2023). Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods. *Algorithms, 16(3) 165.* https://doi.org/10.3390/a16030165
12. EU Legislation in Progress. Artificial intelligence act (2023). https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf
13. Artificial Intelligence Risk Management Framework (2023). https://doi.org/10.6028/NIST.AI.100-1.
14. The Artificial Intelligence Index 2023 Annual Report: AI Index Steering Committee (2023). Institute for Human-Centered AI, Stanford University.
15. Satter, R. (2023). Exclusive: AI being used for hacking and misinformation, top Canadian cyber official says. Reuters. https://www.reuters.com/technology/ai-being-used-hacking-misinfo-top-canadian-cyber-official-says-2023-07-20
16. Sly, L. (2018). U.S. soldiers are revealing sensitive and dangerous information by jogging. Washington Post. https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world-see-where-us-soldiers-are-and-what-they-are-doing/2018/01/28/86915662-0441-11e8-aa61-f3391373867e_story.html
17. Rahman, M., et al. (2023). Security Risk and Attacks in AI: A Survey of Security and Privacy. 47th IEEE-Computer-Society Annual International Conference on Computers, Software, and Applications (COMPSAC), 1834–1839. https://doi.org/10.1109/COMPSAC57700.2023.00284
18. *Data Poisoning and Its Impact on the AI Ecosystem* (2023). https://themathcompany.com/blog/data-poisoning-and-its-impact-on-the-ai-ecosystem
19. Zhu, Y. (2023). Online data poisoning attack against edge AI paradigm for IoT-enabled smart city. Mathematical Biosciences And Engineering. 20(10),17726–17746. https://doi.org/10.3934/mbe.2023788
20. Knight W. (2023). The AI-Powered, Totally Autonomous Future of War Is Here. WIRED. https://www.wired.com/story/ai-powered-totally-autonomous-future-of-war-is-here/
21. Can you trust ChatGPT's package recommendations? (2023). Vulcan Cyber. https://vulcan.io/blog/ai-hallucinations-package-risk
22. ДП «Український науково-дослідний і навчальний центр проблем стандартизації, сертифікації та якості» (ДП «УкрНДНЦ») (2018). Менеджмент ризиків. Принципи та настанови (31000:2018).
23. Barafort, B.; Mesquida, A. & Mas, A. (2019). ISO 31000-based integrated risk management process assessment model for IT organizations. Journal Of Software-Evolution And Process, 31(1). https://doi.org/10.1002/smr.1984

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (22), 2023
ISSN 2663 - 4023

Oleksii Skitsko
PhD, senior researcher
National Academy of the Security Service of Ukraine, Kyiv, Ukraine
ORCID 0000-0003-4122-0889
oiskitsko@gmail.com

Pavlo Skladannyi
PhD, associate professor,
head of the Department of Information and Cyber Security named after Professor Volodymyr Buryachok
Borys Grinchenko Kyiv University, Kyiv, Ukraine
ORCID 0000-0002-7775-6039
p.skladannyi@kubg.edu.ua

Roman Shyrshov
National Academy of the Security Service of Ukraine, Kyiv, Ukraine
ORCID 0000-0003-3534-8736
signorum@gmail.com

Humeniuk Mykhailo
PhD
National Academy of the Security Service of Ukraine, Kyiv, Ukraine
ORCID 0009-0006-3118-721X
ansysdempel@gmail.com

Maksym Vorokhob
lecturer of the Department of Information and Cyber Security named after Professor Volodymyr Buriachok
Borys Grinchenko Kyiv University, Kyiv, Ukraine
ORCID 0000-0001-5160-7134
m.vorokhob@kubg.edu.ua

THREATS AND RISKS OF THE USE OF ARTIFICIAL INTELLIGENCE

Abstract. The article analyzes the advantages of using Artificial Intelligence (AI) in various fields and the risks of impact on the performance of information security and cyber security tasks, as integral components of national security. It was determined that the development of AI has become a key priority for many countries, and at the same time, questions have arisen regarding the safety of this technology and the consequences of its use. The expansion of the scope of application of AI to critical infrastructure objects, the difficulty of verifying the information resources and solutions created by these systems, the threat of a dangerous impact of the results of their operation on the safety of people, society and the state leads to the emergence of risks associated with the use of AI. The lack of transparent methods for checking the conclusions and recommendations of the proposed SSI is a source of uncertainty regarding their accuracy and practical value. This effectively means that SSI can be part of a set of information warfare measures aimed at spreading dubious unverified information and common fakes. The use of artificial intelligence technology can improve the level of computer security. The paper considers the mechanism of risk assessment from the use of AI in various industries and methods of their processing. Proposed approaches to the use of artificial intelligence systems for identification and assessment of risks that arise as a result of the use of artificial intelligence systems.
Artificial intelligence plays a key role in ensuring national security, and its application in various industries contributes to improving efficiency, however, there is an urgent need to develop risk assessment mechanisms for the use of artificial intelligence systems.

Keywords: artificial intelligence; national security; risk assessment; risk management.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1 Bagchi, S., & US, T. C. (2023). *Why We Need to See Inside AI's Black Box*. Scientific American. https://www.scientificamerican.com/article/why-we-need-to-see-inside-ais-black-box/

2 Auchard, E., & Ingram, D. (2018). *Cambridge Analytica CEO claims influence on U.S. election, Facebook questioned*. Reuters. https://www.reuters.com/article/us-facebook-cambridge-analytica-idUSKBN1GW1SG

3 Pause Giant AI Experiments: An Open Letter - Future of Life Institute. (2023). Future of Life Institute.https://futureoflife.org/wp-content/uploads/2023/05/FLI_Pause-Giant-AI-Experiments_An-Open-Letter.pdf.

4 Abdullah, M. F., & Ahmad, K. (2013). The Mapping Process of Unstructured Data to Structured Data. 3rd International Conference on Research and Innovation in Information Systems (ICRIIS), 151–155. https://doi.org/10.1109/ICRIIS.2013.6716700

5 Abdullah, M. F. & Ahmad, K. (2015). Business Intelligence Model for Unstructured Data Management. 5th International Conference on Electrical Engineering and Informatics, 473–477. https://doi.org/10.1109/ICEEI.2015.7352547

6 Venieris, S.; Bouganis, C., & Lane, N. (2023). Multiple-Deep Neural Network Accelerators for Next-Generation Artificial Intelligence Systems. *Computer, 56(3),* 70–79. https://doi.org/10.1109/MC.2022.3176845

7 Xing, J. (2019). The Application of Artificial Intelligence in Computer Network Technology in Big Data Era. 4th International Workshop on Materials Engineering and Computer Sciences, 211–215. https://doi.org/10.25236/iwmecs.2019.044

8 Bian, L. (2023). Design of Computer Network Security Defense System Based on Artificial Intelligence and Neural Network. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-023-10721-9

9 Grechaninov, V., et al. (2022). Formation of Dependability and Cyber Protection Model in Information Systems of Situational Center. *Emerging Technology Trends on the Smart Industry and the Internet of Things, 3149,* 107–117.

10 Hulak, H., et al. (2020). Cryptovirology: Security Threats to Guaranteed Information Systems and Measures to Combat Encryption Viruses. *Cybersecurity: Education, Science, Technique, 2(10),* 6–28. https://doi.org/10.28925/2663-4023.2020.10.628

11 Moskalenko, V.; Kharchenko, V.; Moskalenko A., &; Kuzikov, B. (2023). Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods. *Algorithms, 16(3) 165.* https://doi.org/10.3390/a16030165

12 EU Legislation in Progress. Artificial intelligence act (2023). https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

13 Artificial Intelligence Risk Management Framework (2023). https://doi.org/10.6028/NIST.AI.100-1.

14 The Artificial Intelligence Index 2023 Annual Report: AI Index Steering Committee (2023). Institute for Human-Centered AI, Stanford University.

15 Satter, R. (2023). Exclusive: AI being used for hacking and misinformation, top Canadian cyber official says. Reuters. https://www.reuters.com/technology/ai-being-used-hacking-misinfo-top-canadian-cyber-official-says-2023-07-20

16 Sly, L. (2018). U.S. soldiers are revealing sensitive and dangerous information by jogging. Washington Post. https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world-see-where-us-soldiers-are-and-what-they-are-doing/2018/01/28/86915662-0441-11e8-aa61-f3391373867e_story.html

17 Rahman, M., et al. (2023). Security Risk and Attacks in AI: A Survey of Security and Privacy. 47th IEEE-Computer-Society Annual International Conference on Computers, Software, and Applications (COMPSAC), 1834–1839. https://doi.org/10.1109/COMPSAC57700.2023.00284

18 Data Poisoning and Its Impact on the AI Ecosystem (2023). https://themathcompany.com/blog/data-poisoning-and-its-impact-on-the-ai-ecosystem

19 Zhu, Y. (2023). Online data poisoning attack against edge AI paradigm for IoT-enabled smart city. Mathematical Biosciences And Engineering. 20(10),17726–17746. https://doi.org/10.3934/mbe.2023788

20 Knight W. (2023). The AI-Powered, Totally Autonomous Future of War Is Here. WIRED. https://www.wired.com/story/ai-powered-totally-autonomous-future-of-war-is-here/

21 Can you trust ChatGPT's package recommendations? (2023). Vulcan Cyber. https://vulcan.io/blog/ai-hallucinations-package-risk

22 SE “Ukrainian research and training center for problems of standardization, certification and quality” (2018). Risk management. Principles and guidelines (31000:2018).

23 Barafort, B.; Mesquida, A. & Mas, A. (2019). ISO 31000-based integrated risk management process assessment model for IT organizations. *Journal Of Software-Evolution And Process, 31(1).* https://doi.org/10.1002/smr.1984