УДК 004.75; 004.8
DOI: 10.31673/2409-7292.2025.018929

Марценюк Є. В., Партика А. І., Крет Т. Б.

ДОСЛІДЖЕННЯ ВРАЗЛИВОСТЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ТА ПОБУДОВА КОМПЛЕКСНОЇ МОДЕЛІ БЕЗПЕКИ ОРГАНІЗАЦІЇ

Стрімкий розвиток технологій штучного інтелекту супроводжується зростанням кількості кіберзагроз, що ставлять під загрозу конфіденційність, цілісність та безпеку ШІ-систем (Штучний інтегелект). Впровадження регуляторних вимог, зокрема AI Act Європейського Союзу**,** зобов'язує організації, які займаються розробкою та розгортанням моделей ШІ, дотримуватися високих стандартів кібербезпеки та ефективного управління ризиками. У цьому дослідженні здійснено класифікацію ключових вразливостей штучного інтелекту, оцінено їхній вплив на безпеку ШІ-систем, а також запропоновано багаторівневу архітектуру захисту ШІ-інфраструктури.

Аналіз загроз включає вивчення атак на навчальні дані, до яких належить data poisoning, коли зловмисники модифікують навчальний набір для зміни поведінки моделі. Також розглянуто атаки, спрямовані на саму модель, такі як Атаки противника (adversarial attacks), що дозволяють маніпулювати вихідними даними моделі шляхом введення спеціально підібраних значень. Дослідження охоплює атаки на входи користувачів, серед яких prompt injection та jailbreaking**,** що використовуються для обходу встановлених обмежень та отримання небажаної поведінки моделі. Крім того, розглянуто порушення конфіденційності, зокрема model inversion та membership inference attacks, що дозволяють зловмисникам відновити або виявити дані, використані під час навчання моделі. Окрему увагу приділено ризикам упередженості в алгоритмах штучного інтелекту, зокрема проявам упередженням (bias) в ШІ, які можуть призводити до дискримінаційних результатів через нерепрезентативні або викривлені навчальні вибірки.

На основі проведеного аналізу у статті запропоновано багаторівневу архітектуру безпеки, що сприяє зменшенню ризиків компрометації ШІ-моделей та інфраструктури. Зокрема, розглянуто механізми оцінки впливу ЄС ШІ Акт на безпеку організацій, включаючи аналіз потенційних штрафів, зобов'язань та заходів відповідності для ШІ-орієнтованих компаній. Окремий акцент зроблено на захисті ШІ-інфраструктури у хмарних середовищах (AWS, Azure, GCP) шляхом впровадження методів шифрування даних, ізоляції середовищ, обмеження доступу до моделей та протидії атакам на API. Для забезпечення надійності та безпеки запропоновано впровадження систем моніторингу та детекції загроз, зокрема використання таких інструментів, як Arize AI та Aporia для виявлення аномалій у поведінці моделей, LIME та SHAP для пояснюваності рішень ШІ, а також AWS GuardDuty, Azure Defender та Google SCC для моніторингу кіберзагроз у хмарній інфраструктурі.

Результати цього дослідження можуть бути використані для розробки ефективних методик захисту ШІ-систем, підвищення їхньої стійкості до атак, а також створення надійної та безпечної ШІ-інфраструктури, що відповідає сучасним викликам кібербезпеки та вимогам регуляторних стандартів.

Ключові слова: Штучний інтелект (ШІ/AI), AI Act EU, ШІ безпека, ШІ вразливості, автоматизоване розгортання інфраструктури для ШІ, захист данних, відповідність стандартам кібербезпеки, ШІ ризик менеджмент

Вступ

Сучасний розвиток технологій штучного інтелекту (ШІ) сприяє широкому впровадженню інтелектуальних систем у різні галузі, включаючи охорону здоров'я, фінанси, кібербезпеку та державне управління. Разом із цим зростає кількість потенційних загроз, пов'язаних із компрометацією моделей, маніпуляцією даними та порушенням конфіденційності. Відповідно до AI Act Європейського Союзу, компанії, що розробляють та розгортають AI-системи, зобов'язані впроваджувати механізми забезпечення безпеки та відповідності, щоб мінімізувати ризики, пов'язані із використанням ШІ. Вразливості, такі як атаки на навчальні дані (data poisoning), маніпуляції вхідними запитами (prompt injection, jailbreaking), атаки на модель (adversarial attacks), компрометація конфіденційності (model inversion, membership inference attacks) та ризики упередженості (bias in AI), можуть спричинити значні фінансові втрати, юридичні наслідки та репутаційні ризики для організацій.

Аналіз літературних джерел та формулювання проблеми

Незважаючи на значний прогрес у галузі машинного навчання, питання безпеки ШІ-систем залишається критичним через складність виявлення атак та відсутність єдиних стандартів захисту. Наприклад, adversarial attacks дозволяють зловмисникам вводити модифіковані дані, змушуючи модель видавати неправильні результати, що може призвести

до некоректних рішень у високоризикових сферах. У свою чергу, data poisoning атаки (Отруєння данних) можуть змінювати навчальні набори даних, що викривлює поведінку моделі та впливає на її точність.

Ще однією важливою проблемою є конфіденційність даних. Атаки на моделі, такі як model inversion (інверсія моделі) та membership inference (атака з висновком про членство), дозволяють відновлювати інформацію про вхідні дані, що є особливо небезпечним у випадках роботи з персональними або чутливими даними. Крім того, генеративні моделі ШІ можуть бути вразливими до атак типу prompt injection (промпт ін'єкції) або jailbreaking (джеіл брейк – «втеча з в'язниці»), які дозволяють користувачам змінювати поведінку моделі всупереч її обмеженням.

Сучасні дослідження у сфері безпеки штучного інтелекту підтверджують актуальність цієї проблематики та пропонують різні підходи до її вирішення. У роботі Trazzi & Yampolskiy (2018) [1] запропоновано концепцію "штучної дурості" (Artificial Stupidity) як метод запобігання неконтрольованій поведінці розвинених ШІ-моделей, що може бути використано для обмеження небезпечних сценаріїв використання ШІ. Щодо регуляторного аспекту безпеки ШІ, дослідження Soprana (2024) [2] аналізує сумісність AI Act із міжнародними торговельними угодами (GATS та TBT), що є важливим для глобальних компаній, які працюють на ринку ЄС. Дослідження Bangura (2024) [3] зосереджено на питаннях запобігання дискримінації у ШІ-системах, що є ключовим аспектом безпеки при використанні ШІ у фінансовому та юридичному секторі.

Комплексний аналіз європейських та міжнародних ініціатив щодо регулювання ШІ представлений у роботі Matai (2024), [4] де розглядається порівняння AI Act із іншими глобальними нормативними актами. У свою чергу, Molnar (2024) [5] досліджує процеси впровадження та адаптації AI Act в організаціях, наголошуючи на важливості управління ризиками та безпеки ШІ-моделей у продакшені.

Дослідження, проведені Arize AI та Aporia, демонструють ефективність методів моніторингу загроз та детекції аномальної поведінки ШІ-моделей у продакшені. Зокрема, Aporia спеціалізується на виявленні prompt injection, jailbreaking та інших маніпулятивних атак, тоді як Arize AI зосереджується на виявленні дрейфу моделей, data poisoning та adversarial attacks. У дослідженнях IBM AI Fairness 360 розглядаються механізми мінімізації упередженості в ШІ-системах, що є важливим фактором у забезпеченні справедливості алгоритмічних рішень. Дослідження безпеки ШІ також активно розглядається в контексті хмарних середовищ AWS, Azure, GCP. Інструменти, такі як AWS GuardDuty, Azure Defender та Google SCC, забезпечують виявлення та запобігання атакам на ШІ-системи, що працюють у хмарі. Водночас використання LIME, SHAP сприяє підвищенню пояснюваності моделей, що є важливим аспектом у забезпеченні прозорості та надійності ШІ.

Мета роботи та цілі дослідження

Метою цього дослідження є аналіз основних вразливостей ШІ-систем, моделювання сценаріїв атак та розробку багаторівневої архітектури безпеки для ШІ-орієнтованих організацій. Дослідження охоплює:

1) ідентифікацію та класифікацію основних загроз для ШІ-моделей;
2) тестування ефективності атак на ШІ-системи в контрольованих умовах;
3) розробку комплексного підходу до захисту ШІ-інфраструктури в хмарних середовищах AWS, Azure та GCP;
4) оцінку ефективності сучасних механізмів моніторингу загроз та виявлення аномалій у ШІ-моделях.

Очікується, що результати дослідження сприятимуть розробці ефективних методик захисту ШІ-систем, підвищенню стійкості хмарної і наземної інфраструктури до атак та створенню безпечної ШІ-інфраструктури, яка відповідатиме сучасним викликам кібербезпеки та регуляторним вимогам.

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Регуляторні виклики та ризики штучного інтелекту в контексті ШІ Акт Європейського Союзу. Розвиток штучного інтелекту відкриває нові можливості для автоматизації, підвищення продуктивності та покращення взаємодії між людиною і технологіями. Однак, поряд із перевагами, зростає й кількість ризиків, пов'язаних із безпекою, відповідальністю та потенційним зловживанням технологіями. У відповідь на ці виклики Європейський Союз розробив Закон про штучний інтелект (AI Act), який встановлює єдині правила для використання штучного інтелекту та регулює ризики, що можуть виникати під час його застосування.

ЄС ШІ Акт є першим у світі законодавчим документом, що визначає чіткі вимоги до використання ШІ та класифікує його за рівнем ризику. Організації, які працюють із високоризиковими ШІ-системами, повинні впроваджувати додаткові заходи безпеки та звітності, а порушення цих вимог може призвести до штрафів у розмірі до 35 мільйонів євро або 7% річного обороту компанії**.** [6]

Класифікація ризику в ЄС ШІ Акт. Закон передбачає розподіл ШІ-систем за рівнем ризику, залежно від їхнього впливу на людей та суспільство.

1. Системи з неприйнятним рівнем ризику включають технології, що можуть загрожувати основним правам людини. До них належать алгоритми соціального скорингу (оцінювання громадян на основі їхньої поведінки), маніпулятивні ШІ-системи**,** що впливають на вибір людини без її відома, а також масове біометричне спостереження без належних правових підстав. Використання таких систем заборонено в ЄС.

2. Високоризикові системи охоплюють ШІ-рішення, що застосовуються у критичних сферах**:** медицині, фінансах, правоохоронній діяльності та транспорті. Наприклад, це системи для автоматизованого прийняття рішень у наймі персоналу, оцінки кредитоспроможності чи прогнозування злочинності**.** Такі системи підлягають суворому контролю та повинні відповідати стандартам прозорості, точності та надійності.

3. Системи із середнім рівнем ризику включають ШІ-рішення, що взаємодіють із користувачами, наприклад чат-боти, генеративні моделі та інші інструменти персоналізованої комунікації. Закон вимагає, щоб користувачі знали, що вони взаємодіють з ШІ, а не з реальною людиною. [7]

4. Системи з мінімальним рівнем ризику охоплюють ШІ, що використовується для автоматизації повсякденних завдань, наприклад рекомендаційні алгоритми, системи фільтрації спаму чи пошукові механізми. Вони не підлягають жорсткому регулюванню.

Основні ризики для ШІ-орієнтованих компаній. Організації, що займаються розробкою та впровадженням ШІ, стикаються з низкою ризиків, пов'язаних із дотриманням ЄС ШІ Акт:

- Юридична відповідальність за рішення ШІ. Якщо система штучного інтелекту приймає упереджені або некоректні рішення, наприклад, дискримінує кандидатів під час найму або неправильно оцінює кредитоспроможність клієнтів, компанія може зазнати юридичних наслідків.
- Фінансові штрафи за недотримання регуляторних вимог. ЄС ШІ Акт передбачає серйозні фінансові санкції для компаній, які не забезпечують відповідність своїх ШІ-систем вимогам безпеки та прозорості.
- Ризик дискримінації та упередженості в ШІ. Дослідження Bangura (2024) підкреслює, що багато ШІ-моделей можуть містити приховані алгоритмічні упередження, що призводять до дискримінації за статтю, віком або етнічною приналежністю.
- Захист конфіденційних даних. Згідно з дослідженням Soprana (2024), ЄС ШІ Акт посилює вимоги до обробки персональних даних та зобов'язує компанії дотримуватися стандартів GDPR, щоб уникнути неправомірного використання інформації.
- Постійний моніторинг ШІ-систем. Компанії зобов'язані впроваджувати системи контролю та аудиту ШІ, щоб гарантувати, що алгоритми працюють належним чином. Для

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

цього застосовуються такі інструменти, як Arize AI, Aporia для виявлення аномалій та інструменти пояснюваності рішень, зокрема LIME та SHAP.

Регулювання штучного інтелекту в рамках ЄС ШІ Акт встановлює чіткі вимоги до розробників та постачальників ШІ-рішень, що працюють у сфері високоризикових технологій. Для організацій, що використовують штучний інтелект, це означає необхідність впровадження прозорих алгоритмів, проведення оцінки ризиків та дотримання стандартів безпеки.

Подальше дослідження буде зосереджене на розробці ефективних стратегій захисту ШІ-моделей, зокрема побудові безпечної ШІ архітектури, що допоможе організаціям мінімізувати потенційні загрози та забезпечити відповідність новим вимогам регулювання.

Методи зменшення ризиків та забезпечення безпеки ШІ-моделей. Для організацій, що розробляють або використовують штучний інтелект, питання безпеки ШІ-моделей є одним із ключових викликів. Потенційні ризики включають атаки на моделі, порушення конфіденційності, упередженість алгоритмів та невідповідність регуляторним вимогам. Впровадження багаторівневої стратегії безпеки дозволяє організаціям мінімізувати загрози та підвищити надійність ШІ-систем. Захист ШІ-моделей включає комплекс заходів, які охоплюють етапи розробки, тестування, розгортання та моніторингу. [8]

Розробка та навчання безпечних ШІ-моделей. Одним із ключових аспектів безпеки штучного інтелекту є впровадження захисних механізмів на всіх етапах створення та розгортання моделей. Для зменшення ризиків організація повинна дотримуватися принципу "безпека за замовчуванням" (Security by Design), що передбачає інтеграцію заходів безпеки безпосередньо в процес розробки, починаючи з етапу підготовки даних.

Якість та безпечність навчальних даних. Безпека ШІ-моделей значною мірою залежить від надійності навчального набору. Використання неперевірених або модифікованих даних може призвести до викривлення результатів роботи моделі, появи систематичних помилок або створення умов для реалізації атак. Для забезпечення надійності організація повинна застосовувати верифіковані джерела даних, що пройшли попередню перевірку на відповідність стандартам безпеки. Це дозволяє зменшити ризик атаки отруєння даних (Data Poisoning), коли зловмисники свідомо додають змінені або некоректні дані до навчального набору з метою впливу на роботу ШІ-системи. Окрім перевірки даних, важливим аспектом є виявлення аномалій, що можуть свідчити про потенційні ризики або відхилення в наборі. Для цього доцільно застосовувати автоматизовані інструменти, такі як Great Expectations або AI Fairness 360, що дозволяють ідентифікувати невідповідності, оцінювати якість вхідних даних та мінімізувати вплив небажаних факторів. [9]

Ще одним критичним моментом є аудит навчальних даних на предмет упередженості (bias). Штучний інтелект, що навчається на небалансованих даних, може демонструвати дискримінацію за віком, статтю, національністю чи іншими факторами. Регулярна перевірка навчального набору дозволяє зменшити ризики та покращити точність прогнозів.

Захист моделі від атак. Окрім контролю над даними, організація повинна впроваджувати заходи для захисту ШІ-моделей від маніпуляцій. Одним із ефективних методів є Adversarial Training – підхід, що передбачає навчання моделі на спеціально змінених даних, які можуть імітувати потенційні атаки. Це дозволяє підвищити стійкість системи до adversarial attacks, у яких зловмисники використовують незначні зміни у вхідних даних для того, щоб змусити модель видавати неправильні результати. [10]

Ще одним дієвим механізмом є Differential Privacy – метод, що забезпечує збереження конфіденційності навчальних даних. Ця технологія дозволяє замаскувати внесок окремих записів у вибірку, що робить неможливим визначення чи використання конкретних персональних даних у процесі навчання. Таким чином, навіть якщо модель стане об’єктом атаки membership inference attacks, зловмисники не зможуть отримати інформацію про окремі приклади з навчального набору.

Для запобігання несанкціонованому використанню ШІ-моделей організації можуть використовувати Model Watermarking – технологію цифрового водяного знаку для відстеження джерела та правомірності використання моделей. Це дозволяє захистити авторські права та виявити випадки несанкціонованого використання або копіювання алгоритмів.

Таким чином, ефективний захист ШІ-моделей починається з контролю навчальних даних, впровадження стійких методів тренування та захисту від атак. Використання спеціалізованих механізмів безпеки дозволяє мінімізувати ризики маніпуляцій, підвищити точність моделей та забезпечити їхню відповідність регуляторним стандартам.

Тестування та оцінка безпеки моделей. Перед впровадженням штучного інтелекту в реальні умови організація повинна провести всебічне тестування для виявлення потенційних вразливостей та оцінки ризиків, які можуть вплинути на функціонування AI-системи. Це включає перевірку стійкості моделей до атак, оцінку прозорості ухвалених рішень та тестування поведінки в нетипових сценаріях. [11]

Оцінка безпеки AI-моделей через проникнення. Одним із найважливіших етапів перевірки є тестування на проникнення (Penetration Testing), що дозволяє оцінити, наскільки модель стійка до потенційних атак. Для цього використовуються спеціалізовані фреймворки безпеки ШІ, наприклад Adversarial Robustness Toolbox (ART), який дозволяє виявляти вразливості та аналізувати ризики, пов’язані з маніпуляціями вхідними даними.

Окрім загальної оцінки стійкості, критично важливо перевіряти, чи можлива компрометація конфіденційних даних через атаку на модель. До таких загроз належать:

- Model Inversion – метод, за допомогою якого зловмисники можуть частково відновити дані, на яких навчалася модель, що становить ризик для приватної інформації.
- Membership Inference Attacks – атаки, що дозволяють визначити, чи входив конкретний зразок до навчального набору, що може використовуватися для отримання чутливої інформації.

Перевірка моделей на ці загрози є необхідною, особливо для ШІ-систем, що працюють із персональними, медичними або фінансовими даними.

Перевірка прозорості та пояснюваності рішень. Одним із важливих викликів при розгортанні ШІ-систем є пояснюваність та інтерпретованість їхніх рішень. У багатьох випадках алгоритми глибокого навчання працюють як "чорні скриньки", що ускладнює аналіз того, яким чином приймаються рішення. Це особливо критично для високоризикових ШІ-моделей, таких як автоматизовані кредитні скорингові системи, медичні діагностичні алгоритми або ШІ-рішення у правоохоронних органах.

Для покращення прозорості організації мають впроваджувати методи Explainable AI (XAI). До найбільш ефективних підходів належать:

- LIME (Local Interpretable Model-agnostic Explanations) – техніка, що дозволяє пояснювати, які характеристики вхідних даних найбільше вплинули на прогноз моделі.
- SHAP (Shapley Additive Explanations) – алгоритм, що розраховує, наскільки кожен параметр вплинув на кінцевий результат, допомагаючи ідентифікувати потенційні упередження або неточності в моделі. [12]

Окрім пояснюваності, важливо оцінювати поведінку ШІ в нестандартних або аномальних сценаріях. Для цього необхідно виконувати:

- Статичний аналіз моделі – оцінку архітектури та логіки алгоритму перед розгортанням.
- Тестування чорної скриньки (Black-box Testing) – метод перевірки, при якому система піддається різним типам аномальних вхідних даних, щоб оцінити її реакцію.

Комплексне тестування, що охоплює як безпеку, так і прозорість моделі**,** дозволяє організаціям уникнути потенційних загроз, підвищити надійність ШІ-рішень та забезпечити їхню відповідність регуляторним вимогам.

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Захист AI-інфраструктури під час розгортання. Безпека штучного інтелекту не обмежується лише захистом самої моделі, оскільки значну роль у стійкості системи відіграє середовище, в якому вона працює. Вразливості можуть виникати не лише на рівні алгоритмів машинного навчання, але й у хмарній інфраструктурі, мережевих інтерфейсах (API) та системах контролю доступу. Для забезпечення безпеки ШІ-інфраструктури необхідно впроваджувати багаторівневі механізми захисту, що охоплюють ізоляцію середовищ, шифрування даних, контроль доступу та моніторинг загроз. [13]

Безпечне розгортання в хмарних середовищах (AWS, Azure, GCP). Хмарні середовища є основним інфраструктурним рішенням для багатьох ШІ-систем, оскільки вони забезпечують гнучкість у розгортанні та масштабуванні моделей. Однак саме хмарні платформи є потенційною точкою атаки для зловмисників, тому організація повинна впроваджувати стандарти безпеки для обробки, зберігання та передачі даних.

Одним із ключових методів є ізоляція середовища (sandboxing), що дозволяє обмежити вплив зовнішніх загроз на ШІ-моделі. У хмарних середовищах AWS, Azure та GCP це реалізується через контейнери та віртуальні середовища, що забезпечують окреме виконання кожної моделі без ризику витоку даних або зовнішнього впливу.

Шифрування даних під час зберігання та передачі є ще одним критичним аспектом безпеки. Для цього використовуються спеціалізовані сервіси, такі як AWS Key Management Service (KMS), Azure Key Vault та Google Cloud KMS, що забезпечують автоматичне шифрування даних та контроль доступу на рівні ключів [Vault] [14]. Окрім цього, доцільним є впровадження архітектури Zero Trust, яка передбачає мінімізацію довірених зон та перевірку кожного доступу до AI-системи незалежно від його походження. Це дозволяє зменшити ризик несанкціонованого доступу навіть у разі компрометації внутрішніх мереж.

Захист API та контроль доступу. API є одним із основних каналів взаємодії між ШІ-системами та зовнішнім світом, тому вони є цільовим об'єктом атак, зокрема через API Exploitation та Prompt Injection. Для запобігання атакам на API необхідно впровадити сучасні механізми аутентифікації та авторизації, такі як OAuth 2.0, JWT (JSON Web Tokens) та mTLS (Mutual TLS). Вони дозволяють чітко визначати, хто може взаємодіяти з ШІ-системою та контролювати рівні доступу. Додатковим рівнем захисту є Web Application and API Protection (WAAP) – набір рішень, що дозволяють захистити API від зловмисних запитів. До таких рішень належать Cloudflare, AWS WAF, Azure Front Door, які забезпечують фільтрацію трафіку, виявлення підозрілих запитів та блокування несанкціонованих операцій.

Механізми безпеки для великих мовних моделей **(LLMs)**. Оскільки великі мовні моделі (LLMs) стають основою багатьох ШІ-систем, вони також потребують спеціалізованих механізмів безпеки. Одним із найпоширеніших ризиків є Prompt Injection, коли зловмисник вводить спеціально сформовані запити, що змушують ШІ модель виходити за межі допустимої поведінки [15]. Для протидії таким атакам організації можуть використовувати спеціалізовані інструменти моніторингу безпеки LLM, такі як Aporia, що дозволяють відстежувати маніпуляції з промптами, атаки jailbreaking та інші спроби обійти обмеження ШІ-систем.

Окрім захисту від атак на модель, важливим є контроль вихідних даних, що допомагає запобігати генерації небезпечного або некоректного контенту. Це можливо завдяки впровадженню Content Moderation ШІ – алгоритмів, що автоматично перевіряють вихідні відповіді моделі на наявність неприпустимого контенту. Таким чином, безпека ШІ-інфраструктури вимагає комплексного підходу, що включає ізоляцію середовища, шифрування даних, захист API та спеціалізовані механізми безпеки для мовних моделей. Впровадження цих заходів дозволяє захистити AI-системи від атак, підвищити їхню надійність та забезпечити відповідність регуляторним вимогам.

Постійний моніторинг та реагування на загрози. Після розгортання штучного інтелекту в реальному середовищі критично важливим є безперервний моніторинг AI-систем для своєчасного виявлення нових загроз, аналізу поведінки моделей та запобігання потенційним

атакам. Оскільки алгоритми машинного навчання можуть змінювати свої прогнози внаслідок зміни вхідних даних або цілеспрямованих атак, організація повинна забезпечити автоматизовані механізми моніторингу та реагування на інциденти. [16]

AI Security Operations Center (SOC-AI). Для централізованого керування безпекою AI-систем організація повинна впровадити AI Security Operations Center (SOC-AI) – платформу, що поєднує механізми моніторингу загроз, аналізу ризиків та автоматизованого реагування на інциденти. Основою SOC-AI є системи управління інформаційною безпекою та подіями (SIEM), які дозволяють агрегувати, аналізувати та виявляти аномальну активність у роботі моделей. Серед найефективніших рішень для цього можна виділити Splunk, Azure Sentinel та Google Chronicle, що забезпечують виявлення нестандартних поведінкових патернів моделей та можливих атак. Окрім традиційного моніторингу логів, для виявлення ризиків на рівні AI-моделі використовуються спеціалізовані інструменти, такі як Arize AI та Aporia. Вони дозволяють аналізувати аномальні зміни у прогнозах моделі, що можуть свідчити про атаки data poisoning, model inversion або adversarial attacks. [17]

Ще одним важливим компонентом SOC-AI є інтеграція Security Orchestration, Automation, and Response (SOAR) – технології автоматизації безпекових процесів. SOAR дозволяє створювати автоматичні плейбуки реагування, що запускаються у разі виявлення загроз та забезпечують оперативне усунення потенційних атак.

Аудит та перевірка відповідності регуляторним вимогам. Окрім оперативного моніторингу, організація повинна забезпечити регулярну перевірку відповідності AI-систем міжнародним стандартам безпеки та законодавчим нормам, таким як AI Act, GDPR та ISO/IEC 42001.

Для цього необхідно проводити періодичні тестування щодо дотримання принципів етичності, прозорості та безпеки у функціонуванні AI-моделей. Одним із підходів є автоматизація аудиту AI-інфраструктури, що дозволяє систематично оцінювати відповідність вимогам безпеки та миттєво виявляти потенційні відхилення.

Серед інструментів для автоматизації аудиту можна виділити:

- Terraform Sentinel – механізм контролю безпечного впровадження змін у хмарних середовищах.
- AWS Config – рішення для відстеження змін у конфігураціях AI-інфраструктури в AWS.
- Azure Policy – система управління політиками безпеки для AI-рішень у Microsoft Azure.

Поєднання постійного моніторингу, автоматизованого реагування на загрози та регулярного аудиту дозволяє забезпечити безперервний контроль за безпекою AI-моделей**,** підвищити їхню стійкість до атак та мінімізувати ризики невідповідності регуляторним вимогам. [18]

Результати впровадження та архітектура безпеки ШІ-моделей. Застосування комплексної архітектури безпеки для систем штучного інтелекту є необхідною умовою для організацій, що використовують машинне навчання у критично важливих процесах. Основними цілями такої архітектури є захист AI-моделей від атак, забезпечення прозорості їхньої роботи, зменшення ризиків упередженості та відповідність регуляторним вимогам, таким як ЄС ШІ Акт, GDPR та ISO/IEC 42001. Запропонована архітектура безпеки базується на трьох рівнях моніторингу та управління ризиками, які працюють у взаємодії, забезпечуючи автоматизоване виявлення загроз, аналіз ризиків та контроль відповідності (рис 1).

Контроль продуктивності та безпеки ШІ-моделей. Ефективність та безпека роботи штучного інтелекту значною мірою залежать від постійного моніторингу продуктивності, точності прогнозів та виявлення можливих загроз. Перший рівень архітектури безпеки спрямований на оцінку стабільності ШІ-моделей у реальному часі, що дозволяє запобігати можливим збоям та атакам. [19]

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

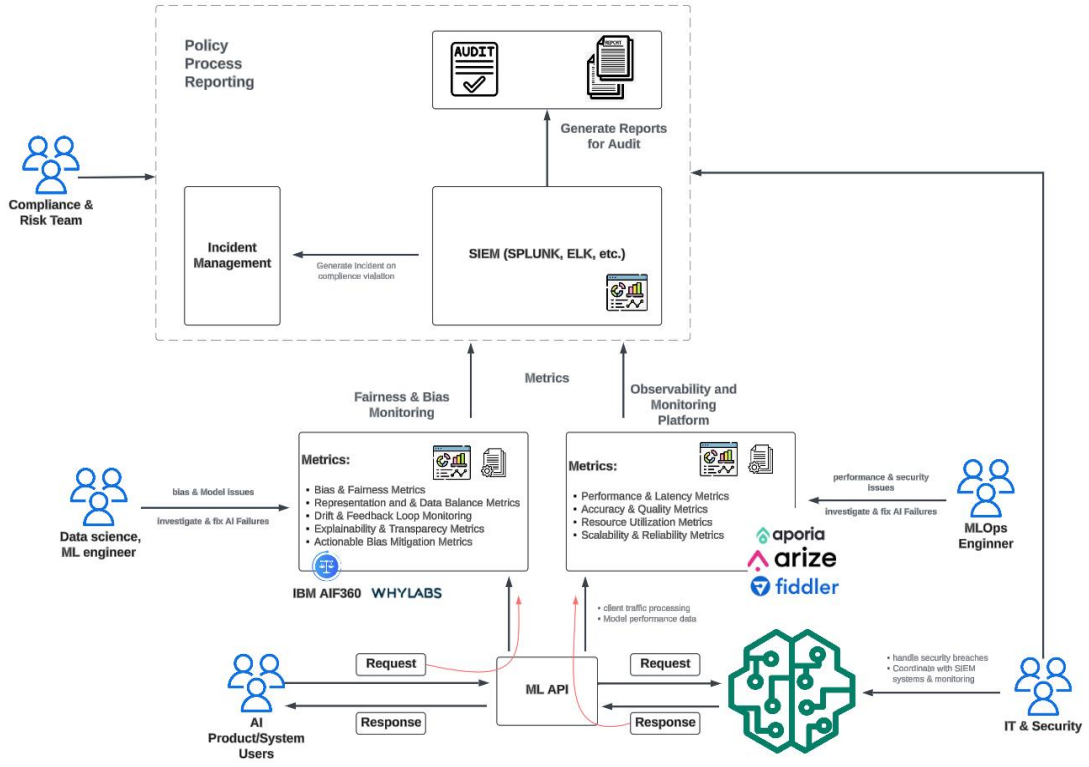

Рис. 1. Архітектура безпеки ШІ моделей

Серед основних загроз, які можуть вплинути на роботу моделей, виділяють:

- Атаки на навчальні дані (Data Poisoning**)** – коли зловмисники навмисно змінюють навчальний набір, щоб модель приймала неправильні рішення.
- Маніпуляція вхідними запитами (Prompt Injection) – ситуації, коли шкідливі або некоректні запити можуть змусити ШІ надати небажані відповіді.
- Зміни у вхідних даних (Data Drift) – поступове відхилення характеристик даних, на яких працює модель, що може призвести до зниження точності прогнозів.

Моніторинг продуктивності та якості прогнозів. Для підтримки стабільності ШІ-систем необхідно постійно оцінювати швидкодію моделі та її здатність генерувати коректні результати. Моніторинг охоплює:

- Вимірювання часу реакції на запити користувачів та аналіз затримок у процесі обробки даних.
- Перевірку точності прогнозів, шляхом порівняння результатів роботи моделі з очікуваними значеннями.
- Контроль за використанням ресурсів, щоб оптимізувати навантаження та запобігти збоям у роботі системи.

Виявлення аномалій у поведінці моделі. Моніторинг продуктивності доповнюється системами виявлення аномалій, які дозволяють ідентифікувати незвичайну поведінку ШІ-моделей та негайно реагувати на загрози.

Для цього використовуються спеціалізовані платформи, такі як Aporia, Arize AI, Fiddler, що аналізують вихідні дані моделі, ідентифікують відхилення у прогнозах та контролюють зміну характеристик вхідних даних. [20]

Якщо система виявляє аномальні зміни у прогнозах, автоматично надсилається сповіщення командам MLOps та кібербезпеки. Це дозволяє швидко реагувати на потенційні атаки, дрейф даних або інші загрози, які можуть вплинути на стабільність ШІ-системи.

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Інтеграція з процесами MLOps. Щоб підтримувати високу точність і безперервну роботу ШІ-моделей, важливо забезпечити гнучку інтеграцію із процесами MLOps. Це дозволяє:

- Періодично перевіряти роботу моделі та оновлювати її параметри у разі виявлення зниження точності прогнозів.
- Автоматично оновлювати навчальний набір або параметри моделі, якщо вхідні дані зазнають суттєвих змін, що може вплинути на її продуктивність.

Цей рівень архітектури дозволяє організаціям забезпечити контроль за технічною ефективністю ШІ-моделей та мінімізувати ризики, пов'язані зі зниженням точності прогнозів, атакою на систему або неконтрольованими змінами у вихідних даних. Завдяки впровадженню систем автоматизованого моніторингу та механізмів швидкого реагування компанії можуть підтримувати стабільність та безпеку своїх ШІ-рішень, що є критично важливим для їхньої надійної роботи. [21]

*Автоматизація розгортання та забезпечення безпеки хмарної інфраструктури для систем штучного інтелекту***.** З огляду на зростаючу складність та вимоги до безпеки моделей штучного інтелекту, організації стикаються з необхідністю не лише забезпечувати ефективний моніторинг та контроль відповідності, а й автоматизувати процеси розгортання. Це дозволяє мінімізувати ризики, пов'язані з неправильною конфігурацією, невідповідністю стандартам безпеки та впливом людського фактору. Автоматизація інфраструктури хмарного середовища є важливим кроком до стандартизації управління технологічними ресурсами, що у свою чергу забезпечує дотримання регуляторних вимог, таких як AI Act ЄС, GDPR та ISO/IEC 42001. [22]

Автоматизоване розгортання інфраструктури для ШІ-систем. Сучасні ШІ-моделі потребують гнучкої та безпечної архітектури, здатної до масштабування, швидкого оновлення та інтеграції з платформами моніторингу. Використання підходу Infrastructure as Code (IaC) забезпечує декларативне управління інфраструктурою, що зменшує ймовірність людських помилок та дозволяє стандартизувати процеси створення та розгортання ресурсів.

Автоматизоване розгортання передбачає застосування таких технологій, як AWS CloudFormation та Azure Resource Manager, які дозволяють організаціям централізовано контролювати конфігурації ресурсів у хмарних середовищах AWS, Azure та GCP. Це сприяє гнучкості інфраструктури та підвищує рівень безпеки за рахунок чітко визначених політик та правил.

Окрім управління інфраструктурою, критично важливим є належний захист хостингового середовища. Це досягається шляхом інтеграції з Prisma Cloud, AWS Config, Azure Policy та Google Security Command Center, які виконують функції аналізу відповідності безпеки. Додатково, для захисту конфіденційних даних використовується HashiCorp Vault, що забезпечує централізоване керування ключами доступу та безпечне зберігання секретної інформації. Дані, що передаються та зберігаються у хмарному середовищі, підлягають автоматичному шифруванню за допомогою AWS KMS, Azure Key Vault та Google Cloud KMS, що унеможливлює їхній несанкціонований доступ. [23]

Процес розгортання моделей ШІ також передбачає оркестрацію контейнеризованих середовищ. Використання Kubernetes (EKS, AKS, GKE) дозволяє масштабувати навантаження, ефективно розподіляти ресурси та гарантувати стабільну роботу моделей навіть під високими обчислювальними навантаженнями. Додатково, для автоматизації керування контейнерами застосовуються такі інструменти, як Rundeck, а CI/CD-процеси реалізуються через GitHub Actions, Jenkins або GitLab CI, що забезпечує швидке розгортання оновлень та виправлення можливих вразливостей.

Моніторинг та відповідність безпеки у ШІ-системах. Розгортання ШІ-моделей потребує постійного контролю їхньої продуктивності та безпеки, щоб уникнути загроз, пов'язаних із дрейфом даних, маніпуляцією вхідними запитами чи змінами у характеристиках моделі. Для цього інтегруються системи моніторингу продуктивності, які відстежують поведінку моделі в режимі реального часу. [24]

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Виявлення аномалій у роботі ШІ відбувається завдяки застосуванню спеціалізованих платформ, таких як Arize AI, Aporia та Fiddler. Вони дозволяють визначати випадки зниження точності прогнозів, зміни у розподілі вхідних даних та ідентифікувати потенційні загрози, зокрема Data Poisoning та Concept Drift. Для покращення пояснюваності рішень, що приймаються моделлю, використовується підхід Explainable AI (XAI), що допомагає розкривати внутрішню логіку функціонування алгоритму та оцінювати його коректність. Окрім контролю продуктивності, важливою складовою забезпечення безпеки є управління інцидентами та автоматизоване виявлення загроз. Для цього інтегруються SIEM-системи, такі як Splunk, ELK, Azure Sentinel та Google Chronicle, що дозволяють збирати, аналізувати та корелювати події безпеки. У разі виявлення аномальної активності або порушень безпеки система автоматично створює інцидент, який передається на розгляд відповідних команд для подальшого розслідування. [25]

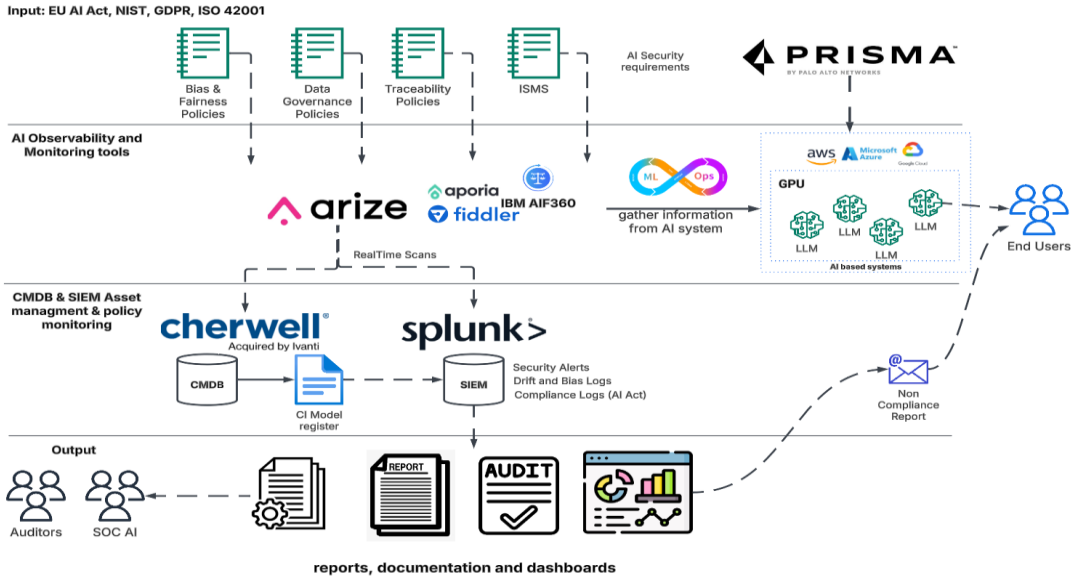

Рис. 2. Архітектура централізованого розгортання ШІ рішень в хмарних середовищах

Для дотримання регуляторних вимог, таких як AI Act та GDPR, організації впроваджують Prisma Cloud, та за допомогою цього інструменту проводять моніторинг хмарних хостінгів, що забезпечують автоматичне тестування конфігурацій на відповідність стандартам безпеки. Також проводиться постійне сканування середовищ за допомогою Prisma Cloud Defender, що дозволяє виявляти вразливості на рівні віртуальних GPU машин (Рис.2).

Централізоване управління та реагування на загрози. Щоб мінімізувати ризики, пов'язані з експлуатацією моделей ШІ, організації створюють єдиний центр управління безпекою та відповідністю. Це дозволяє не лише аналізувати загрози, а й здійснювати автоматизоване реагування на можливі інциденти, забезпечуючи максимальну безперервність роботи системи. [26]

Централізоване управління реалізується через AI Security Operations Center (SOC-AI), що поєднує можливості SIEM та SOAR-платформ для автоматизованого аналізу та усунення загроз. Використання AI-аналітики у процесах SOC-AI дозволяє прогнозувати потенційні ризики та розробляти стратегії запобігання атакам ще до їхнього виникнення.

Ще одним важливим аспектом є автоматичне виявлення та виправлення вразливостей. Це досягається шляхом інтеграції систем контролю безпеки, таких як Prisma Defender, які здійснюють безперервний аналіз конфігурацій інфраструктури. Якщо виявляється критична вразливість, система автоматично застосовує необхідні оновлення або змінює конфігурацію безпеки, щоб запобігти можливому використанню вразливості зловмисниками.

© Марценюк Є. В., Партика А. І., Крет Т. Б. (2025). Дослідження вразливостей штучного інтелекту та побудова комплексної моделі безпеки організації. Сучасний захист інформації, 1(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Контроль доступу до інтерфейсів API також відіграє важливу роль у захисті ШІ-систем. Використання OAuth 2.0, JWT та mTLS забезпечує надійну аутентифікацію користувачів, що взаємодіють із моделями, тоді як інтеграція Web Application & API Protection (WAAP) дозволяє виявляти та блокувати потенційно небезпечні запити. [27]

Автоматизація інфраструктури для розгортання та моніторингу моделей ШІ є ключовим фактором у забезпеченні їхньої безпеки, відповідності стандартам та ефективного управління ризиками. Використання декларативного підходу до конфігурації ресурсів, оркестрація контейнеризованих середовищ, впровадження SIEM/SOAR-систем та контроль аномалій за допомогою AI-аналітики дозволяють створити гнучке, масштабоване та безпечне середовище для роботи ШІ-систем. Це не лише зменшує ризики кіберзагроз, але й забезпечує довготривалу відповідність регуляторним вимогам та високий рівень надійності інфраструктури.

Висновки

Впровадження технологій штучного інтелекту в організаційні процеси значно підвищує ефективність аналізу даних, оптимізацію бізнес-процесів та прийняття рішень. Проте, водночас із цими перевагами виникає низка викликів, зокрема пов'язаних із забезпеченням безпеки ШІ-систем, їхньої прозорості та відповідності регуляторним вимогам. Основні загрози включають можливість маніпуляції навчальними даними, атаки через вхідні запити та порушення конфіденційності, що може спричинити неконтрольовані зміни у поведінці моделей.

Дослідження було спрямоване на:

1. ідентифікацію та класифікацію основних загроз для ШІ-моделей, що дало змогу виокремити ключові вразливості, пов'язані з атаками на навчальні дані, експлуатацією слабких місць архітектури моделей та впливом на результати прогнозування через зовнішні маніпуляції.

2. У межах дослідження також проведено тестування ефективності атак на ШІ-системи в контрольованих умовах, що дозволило оцінити стійкість моделей до таких загроз, як Data Poisoning, Prompt Injection та Model Inversion. Крім того, для забезпечення безперервного моніторингу хмарної інфраструктури інтегрується Prisma Cloud, що дозволяє виконувати автоматизоване тестування конфігурацій на відповідність стандартам безпеки. Система постійного сканування за допомогою Prisma Cloud Defender дає змогу ідентифікувати вразливості на рівні віртуальних машин (GPU) та контейнеризованих середовищ, що критично важливо для розподілених ШІ-рішень.

3. Для мінімізації цих ризиків у дослідженні запропоновано архітектуру безпеки ШІ та створення комплексного підходу до захисту ШІ інфраструктури в хмарних середовищах AWS, Azure та GCP, яка ґрунтується на трирівневій системі моніторингу та управління загрозами. Перший рівень відповідає за контроль продуктивності та оцінку стабільності ШІ-моделей, що дозволяє виявляти потенційні збої та аномальні відхилення у прогнозах, які можуть бути наслідком атак або змін у вхідних даних. Використання платформ Aporia, Arize AI та Fiddler сприяє автоматичному аналізу змін у поведінці моделей та запобіганню можливим порушенням їхньої роботи. Другий рівень сфокусований на оцінці прозорості та справедливості ШІ-рішень. Це дозволяє контролювати алгоритмічну упередженість, оцінювати рівномірність розподілу навчальних даних та мінімізувати ризики дискримінації. Впровадження методології Explainable AI (XAI) з використанням інструментів LIME та SHAP дає змогу пояснювати логіку ухвалення рішень моделлю, що підвищує рівень довіри користувачів до результатів її роботи.

Третій рівень передбачає централізоване управління інцидентами, моніторинг кіберзагроз та аудит відповідності. Інтеграція з SIEM-системами, такими як Splunk, ELK, Azure Sentinel та Google Chronicle, забезпечує всебічний аналіз подій безпеки та дозволяє автоматично реагувати на потенційні загрози. Використання Terraform Sentinel, AWS Config та Azure Policy сприяє підтриманню відповідності ШІ-систем міжнародним стандартам,

зокрема AI Act ЄС, GDPR та ISO/IEC 42001, та зменшенню ризиків невідповідності нормативним вимогам.

Розвиток архітектури безпеки неможливий без інтеграції з хмарними середовищами, у яких сьогодні працює більшість ШІ-систем. Для підвищення рівня захищеності інфраструктури було розроблено Запропонований підхід базується на використанні Infrastructure as Code (IaC) та політик безпеки для автоматизованого управління конфігураціями хмарних ресурсів. Важливим аспектом є захист конфіденційних даних, який реалізується через HashiCorp Vault та механізми шифрування AWS KMS, Azure Key Vault, Google Cloud KMS. Крім того, для забезпечення безперервного моніторингу хмарної інфраструктури інтегрується Prisma Cloud, що дозволяє виконувати автоматизоване тестування конфігурацій на відповідність стандартам безпеки. Використання Prisma Cloud Defender дає змогу виявляти вразливості на рівні віртуальних машин (GPU) та контейнеризованих середовищ, що є критично важливим для розподілених ШІ-рішень.

4. Централізоване управління безпекою та реагування на загрози реалізується через Security Operations Center для ШІ (SOC-AI). Це інтеграційна платформа, яка поєднує SIEM та SOAR-рішення, що дозволяє автоматизувати аналіз загроз, прогнозувати ризики та розробляти стратегії запобігання інцидентам. Застосування AI-аналітики в SOC-AI сприяє підвищенню рівня захисту, оскільки система здатна навчатися на історичних даних та вдосконалювати механізми виявлення аномалій. У межах дослідження також проведено оцінку ефективності сучасних механізмів моніторингу загроз та виявлення аномалій у ШІ-моделях, що дозволило визначити їхні сильні та слабкі сторони, а також розробити рекомендації щодо їхнього вдосконалення.

Запропонований підхід до захисту ШІ-систем дає змогу значно зменшити ризики атак, оптимізувати моніторинг безпеки та забезпечити відповідність міжнародним стандартам. Інтеграція сучасних технологій моніторингу та управління дозволяє своєчасно виявляти загрози та розробляти механізми для їхнього попередження. Впровадження автоматизованих систем аналізу даних та відповідності нормативним вимогам знижує навантаження на команди безпеки, що дозволяє їм зосередитися на стратегічному управлінні ризиками.

Серед основних переваг архітектури можна виділити:

- можливість автоматизованого виявлення та запобігання атакам на ШІ-моделі;
- підвищення прозорості прийняття рішень за рахунок використання Explainable AI;
- зменшення впливу людського фактору на процеси забезпечення безпеки;
- оптимізацію використання обчислювальних ресурсів та мінімізацію ризиків простою систем;
- спрощення процесів аудиту та відповідності за рахунок автоматизації перевірок на рівні хмарної інфраструктури.

Таким чином, запропонований комплексний підхід до безпеки ШІ-систем дозволяє оптимізувати управління ризиками, забезпечити прозорість функціонування моделей та відповідати сучасним вимогам кібербезпеки. Використання автоматизованих механізмів моніторингу та відповідності дозволяє організаціям не лише захищати власні ШІ-рішення, але й адаптуватися до швидко змінюваного ландшафту загроз у сфері штучного інтелекту.

Перелік посилань

1. Trazzi, Michaël & Yampolskiy, Roman. (2018). Building Safer AGI by introducing Artificial Stupidity. 10.48550/arXiv.1808.03644.

2. Soprana, Marta. (2024). Compatibility of emerging AI regulation with GATS and TBT: the EU Artificial Intelligence Act. Journal of International Economic Law. 27. 10.1093/jiel/jgae040.

3. Bangura, Gabriel. (2024). The ЄС ШІ Акт - Mitigating Discrimination In Artificial Intelligence Systems. 10.13140/RG.2.2.27020.63367.

4. Matai, Puneet. (2024). Comprehensive Guide to AI Regulations: Analyzing the ЄС ШІ Акт and Global Initiatives. International Journal of Computing and Engineering. 6. 45-54. 10.47941/ijce.2110.

5. Molnar, David. (2024). AI unleashed: mastering the maze of the ЄС ШІ Акт. 10.56461/iup_rlrc.2024.5.ch12.

6. EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. (n.d.). https://artificialintelligenceact.eu/

7. Скіцько, Олексій & Ширшов, Роман. (2024). НОРМАТИВНО-ПРАВОВЕ ЗАБЕЗПЕЧЕННЯ КІБЕРБЕЗПЕКИ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ. Науковий вісник Львівського державного університету внутрішніх справ (серія юридична). 73-79. 10.32782/2311-8040/2024-2-11.

8. Міжнародні стандарти регулювання штучного інтелекту: аналіз актів, розроблених за результатами Хіросімського процесу з ШІ. ЮРЛІГА. https://jurliga.ligazakon.net/. (2024, February 6).

9. Pochu, Sandeep & Nersu, Sai & Kathram, Srikanth. (2024). AI-Powered Monitoring: Next-Generation Observability Solutions for Cloud Infrastructure. Journal of AI-Powered Medical Innovations (International online ISSN 3078-1930). 2. 140-152. 10.60087/Japmi.Vol.02.Issue.01.Id.010.

10. Sign, Ghader. (2024). Data-Driven AI Models for Cybersecurity: Optimizing Data Pipelines and Infrastructure Protection in a Cloud-First World. 10.13140/RG.2.2.30481.44642.

11. Marinova, Miroslava. (2024). Balancing Innovation and Regulation: Evaluation of the CMA's Report on AI Foundation Models and their impact on competition and consumer protection.

12. Ruschemeier, Hannah. (2025). Generative AI and data protection. Cambridge Forum on AI: Law and Governance. 1. 10.1017/cfl.2024.2.

13. Nguyen, Phan & Quang, Nguyen. (2025). Copyright protection for AI-generated works: A comparative review of international and Vietnamese laws. Arts & Communication. 3745. 10.36922/ac.3745.

14. Pan, Qianqian & Dong, Mianxiong & Ota, Kaoru & Wu, Jun. (2022). Device-Bind Key-Storageless Hardware AI Model IP Protection: A PUF and Permute-Diffusion Encryption-Enabled Approach. 10.48550/arXiv.2212.11133.

15. Martseniuk Y., Partyka A., Harasymchuk O., Shevchenko*** S. Universal centralized secret data management for automated public cloud provisioning // CEUR Workshop Proceedings. – 2024. – Vol. 3826 : Proceedings of the workshop "Cybersecurity providing in information and telecommunication systems II", Kyiv, Ukraine, October 26, 2024 (online).. – P. 72–81.

16. Nagar, Mayura. (2025). From Data to Sustainability: AI Case Studies in Shaping Sustainable Landscapes. 10.4018/979-8-3693-3410-2.ch008.

17. Malik, Shoaib. (2024). The Future of AI in Biometric Security: Enhancing Authentication and Privacy Protection. 10.13140/RG.2.2.34306.59844.

18. Lee, Giljae. (2024). Personal Data Protection Issues in the Era of Artificial Intelligence. Journal of Medical Imaging. 7. 13-18. 10.31916/sjmi2024-01-03.

19. Aslam, Umair & Amelia, Oscar. (2022). Exploring the Influence of Data Protection Laws on AI Development and Ethical Use. 10.13140/RG.2.2.28372.41603.

20. Katrakazas, Panagiotis & Papastergiou, Spyros. (2024). A Stakeholder Needs Analysis in Cybersecurity: A Systemic Approach to Enhancing Digital Infrastructure Resilience. Businesses. 4. 225-240. 10.3390/businesses4020015.

21. Hindle, Andrew. (2020). Impact of GDPR on Identity and Access Management. IDPro Body of Knowledge. 1. 10.55621/idpro.24.

22. Martseniuk Y., Partyka A., Harasymchuk O., Korshun*** N. Automated conformity verification concept for cloud security // CEUR Workshop Proceedings. – 2024. – Vol. 3654 : Cybersecurity providing in information and telecommunication systems 2024. Proceedings of the workshop cybersecurity providing in information and telecommunication systems (CPITS 2024) Kyiv, Ukraine, February 28, 2024 (online).. – P. 25–37.–

23. Марценюк Є. В., Партика А. І. Аналіз впливу тіньових IT на інфраструктуру хмарних середовищ підприємства // Безпека інформації. – 2024. – Т. 30, № 2. – С. 270–278

24. Ashraf, Nadeem & Badi, Sadi. (2024). Integrating AI for Monitoring and Compliance: Tackling Climate Change in the Oil and Gas Industry give. 10.13140/RG.2.2.19022.78403.

25. Olasehinde, Tolamise & Jason, Frank. (2024). INTEGRATING AI AND MACHINE LEARNING FOR COMPLIANCE MONITORING IN DATA LAKES.

26. Khan, Umar & Aggarwal, D & Muslim, & Sohrab. (2024). Analyzing the Role of Artificial Intelligence (AI) in Monitoring Corporate Governance Practices and Ensuring Compliances in Improved Decision-Making Processes. 11. 3025-3031.

27. Dimitrijević, Nikola & Zdravković, Nemanja & Bogdanović, Milena & Mesterovic, Aleksandar. (2024). Advanced Security Mechanisms in the Spring Framework: JWT, OAuth, LDAP and Keycloak.

Надійшла 16.02.2025