УДК 004.056.5:004.8 **DOI:10.30837/rt.2025.2.221.02**

Є.О. ЛОГАЧОВА, М.В. ЄСІНА, канд. техн. наук, Д.Ю. ГОЛУБНИЧИЙ, канд. техн. наук

ДОСЛІДЖЕННЯ ТА АНАЛІЗ МІЖНАРОДНИХ СТАНДАРТІВ ТА РЕГУЛЯТОРНИХ ВИМОГ ЩОДО БЕЗПЕКИ ШТУЧНОГО ІНТЕЛЕКТУ, РОЗРОБКА МОДЕЛІ БЕЗПЕКИ ДЛЯ УКРАЇНИ

Вступ

В епоху сучасних технологій надважливо забезпечити їх необхідним рівнем безпеки для збереження належного функціонування бізнесу, різноманітних структур тощо. Також необхідно запобігти порушенню прав та свобод звичайних громадян і користувачів. Штучний інтелект (ШІ) – відносно нова технологія на ринку, що завоювала прихильність як звичайних користувачів, так і великих корпорацій. Розумні машини на основі штучного інтелекту допомогли у стрімкому розвитку багатьох галузей, таких як медицина, освіта, транспорт, сільське господарство, фінанси тощо. Разом з тим ШІ почали використовувати і для покращення рівня кібершахрайств. Для того щоб не відмовлятись від даних прогресивних технологій та не уповільнювати прогрес, багато країн вже почали вводити різноманітні рішення щодо безпеки використання штучного інтелекту. Україна наразі знаходиться на початковій стадії у даному питання, тому розгляд міжнародного досвіду допоможе зробити нормативно-правові регулювання в Україні ефективними та безпечними для усіх ланок держави.

1. Кібербезпека в епоху штучного інтелекту

Штучний інтелект радикально трансформує цифровий світ: автоматизація процесів, аналіз великих даних, прогнозування загроз. Спершу ШІ використовували для маленьких повторюваних задач, на зараз штучний інтелект використовується для діагностики здоров'я, організації бізнесу, навчання. Тобто технологія розвивається стрімко і використовується майже у всіх сферах життя, чи неабияк полегшує життя своїх користувачів.

За даними IBM, середня вартість витоку даних у 2023 р. становила понад 4,45 млн доларів, і ШІ дедалі частіше використовується як захисний, так і атакуючий інструмент [1]. А вже у 2024 р., згідно з IBM Cost of a Data Breach Report 2024, середня вартість витоку даних зросла до 4,88 млн доларів США, що на 10 % більше порівняно з 2023 р. Та при цьому штучний інтелект та автоматизація допомогли компаніям знизити витрати на усунення наслідків витоку даних у середньому на 2,2 млн доларів [2]. ШІ допомагає вирішувати декілька основних проблем інформаційної безпеки: виявлення загроз у реальному часі, автоматичне реагування на інциденти, прогнозування атак та інші.

Але з розвитком ШІ зростають і виклики – зокрема в сфері кібербезпеки. Відтак зловмисники можуть використати технології штучного інтелекту для створення deepfake контенту, автоматизації фішингових атак, атак соціальної інженерія нового рівня тощо. Окрім цього дана технологія може бути задіяна і у автоматизації шкідливого програмного забезпечення (ПЗ). Один із показових прикладів – дослідницький експеримент фахівців з кібербезпеки компанії Hyas, які створили умовний вірус під назвою BlackMamba. Цей шкідливий код мав здатність динамічно змінювати свою поведінку за допомогою ChatGPT, генеруючи частини свого функціоналу в режимі реального часу. Під час випробувань BlackMamba демонстрував здатність адаптуватися до різних середовищ та залишатися непоміченим більшістю популярних антивірусних систем [3].

У лютому 2024 р. в Гонконзі стався один із наймасштабніших випадків шахрайства із використанням штучного інтелекту. Співробітник міжнародної компанії отримав завдання надіслати 25 млн доларів США від імені свого нібито директора з Великої Британії. Після сумнівного електронного листа із запрошенням на відеоконференцію чоловік почав підозрювати фішинг. Проте участь кількох «знайомих колег» у відеодзвінку знизила його настороженість. Протягом кількох транзакцій, що охоплювали п'ять різних банківських рахунків,

працівник переказав загалом понад 200 млн гонконзьких доларів (приблизно 25 млн). Як згодом з'ясувалося, усі інші учасники відеоконференції були глибоко реалістичними штучними двійниками, створеними зловмисниками з використанням технологій штучного інтелекту та deepfake [4].

Усе це ставить непросту задачу перед спеціалістами з інформаційної безпеки, адже дана технологія неабияк підвищує рівень безпеки та при цьому ж може бути використана кіберзлочинцями для здійснення атак. Хорошим рішенням є впровадження необхідних стандартів та регуляторних вимог для безпечного використання ШІ. Для України це питання стоїть доволі серйозно, адже в умовах війни кількість фейків та кібератак все збільшується.

2. Огляд міжнародних стандартів та регуляторних вимог безпеки ШІ

Існує багато підходів і думок щодо нормативно-правових регулювань штучного інтелекту. У одних країнах до цієї технології ставляться з обережністю, а у інших вбачають у ній великий прогресивний поштовх для розвитку цифровізації.

Варто почати з Європейського Союзу (ЄС) та введеного ними EU AI Act, який фактично став першим офіційним і повноцінним документом з регулювання ШІ. EU AI Act встановлює чіткі правила для використання штучного інтелекту, враховуючи рівень потенційного ризику, який можуть нести різні типи систем ШІ. Залежно від цього рівня, як провайдери, так і користувачі зобов'язані дотримуватися певних вимог. Навіть у випадках, якщо ризик від технології вважається низьким, вона має пройти оцінювання, щоб гарантувати прозорість та відповідність етичним стандартам [5].

До категорії неприйнятного ризику потрапляють застосунки ШІ, які заборонені в ЄС, оскільки становлять загрозу для базових прав і свобод людини. Це включає штучний інтелект, що маніпулює поведінкою людей – особливо вразливих груп, таких як діти. Наприклад, голосові іграшки, які можуть спонукати до небезпечної поведінки. Також заборонені системи соціальної оцінки (оцінювання людей за поведінкою чи статусом), біометрична ідентифікація в публічних просторах у реальному часі, а також категоризація людей за біометричними ознаками [5].

У деяких випадках, як-от діяльність правоохоронних органів, можливе обмежене використання систем з високим ризиком. Наприклад, дистанційна біометрична ідентифікація в реальному часі дозволяється лише у виняткових ситуаціях та після попереднього дозволу суду. Ідентифікація «постфактум», тобто із затримкою, може застосовуватися для розслідування тяжких злочинів, але також лише за наявності юридичного схвалення [5].

Системи штучного інтелекту, які становлять високий ризик, охоплюють сфери, що мають прямий вплив на безпеку чи фундаментальні права громадян. Вони поділяються на дві ключові категорії. Перша – це ШІ, інтегрований у продукти, що регулюються європейським законодавством щодо безпеки. Наприклад, медичне обладнання, автомобілі, іграшки, ліфти. Друга – системи, які функціонують у чутливих галузях: управління критичною інфраструктурою, освіта, зайнятість, доступ до державних послуг, правоохоронна діяльність, а також процеси міграції й правозастосування [5].

Всі ці системи підлягають обов'язковому контролю як до їхнього впровадження на ринку, так і протягом усього життєвого циклу. Крім того, громадяни матимуть право подавати скарги на ШІ-системи до національних наглядових органів, що зміцнює принцип прозорості та підзвітності в епоху технологічного прогресу.

Ще одним важливим принципом є прозорість. Так, наприклад ChatGPT не вважається технологією з високим рівнем ризику, проте має відповідати законодавству ЄС, зокрема законодавству щодо авторського права. Таким чином, увесь контент, згенерований ChatGPT, має бути маркований спеціальним знаком, це допомагає запобігати поширенню фейкових зображень, відео та іншого контенту з шахрайською метою. Більш досконала модель штучного інтелекту GPT-4 повинна проходити ретельну оцінку, а про будь-які серйозні інциденти потрібно повідомляти Європейську комісію [5].

Перші норми, зокрема заборона на використання систем із неприйнятним рівнем ризику, набули чинності вже 2 лютого 2025 р. Інші елементи, як-от кодекси практики та вимоги до прозорості для універсальних моделей ШІ, почнуть діяти через 9 та 12 місяців відповідно [5].

Системи з високим рівнем ризику, включно із тими, що використовуються у сфері охорони здоров'я, освіти, критичної інфраструктури або правоохоронній діяльності, матимуть більше часу на адаптацію. Для них вимоги набудуть чинності через 36 місяців, що дозволить розробникам і користувачам належним чином підготуватися до впровадження нових стандартів.

NIST AI Risk Management Framework – це стратегічний документ, розроблений Національним інститутом стандартів і технологій США (NIST) з метою допомогти організаціям ефективно управляти ризиками, пов'язаними з впровадженням та використанням штучного інтелекту. Він є добровільним, але широко рекомендованим інструментом, який сприяє розробці безпечних, надійних, прозорих і етично обґрунтованих систем ШІ. За цим документом прийнято розподілити можливу нанесену шкоду на три категорії: шкода людям, шкода організації та шкода екосистемі. Шкода людям включає у себе фізичну, психологічну, соціальну або економічну шкоду. Наприклад, фальшиві медичні поради, неправдиві новини, або контент, що провокує насильство. ШІ також може бути використаний для дискримінації, упередженого оцінювання, неправомірного спостереження або цензури – особливо у випадках автоматизованих рішень без належного контролю, що призводить до порушення прав та свободи громадян [6].

Друга категорія під назвою «шкода організації» окреслює такі наслідки, як витіснення творчих професій, або недобросовісну конкуренцію через масове створення фальшивого або дезінформаційного контенту, збої на ринку праці та інших операцій, шкода репутації та інші потенційні ризики безпеці [6].

Третя категорія включає збої у глобальних фінансових системах або системах ланцюга поставок і шкоду навколишньому середовищу та природним ресурсам. Дані категорії створені для запобігання шкодам та ризикам описаним у них, що допомагає сконцентрувати увагу на актуальних проблемах [6].

Документ включає сім основних характеристик для надійних систем штучного інтелекту, яким вони мають відповідати. Першими є валідність та надійність, які означають, що система виконує свою функцію точно, стабільно і в межах запланованого контексту. Вона має бути перевірена на відповідність заявленим цілям і демонструвати передбачувану поведінку навіть в умовах змін середовища чи вхідних даних. Безпечність доповнює цю характеристику, адже система не повинна створювати фізичної або психологічної шкоди користувачам, і має бути захищеною від зловмисного впливу [6].

Неупередженість системи також дуже важлива – це здатність системи уникати дискримінаційних рішень або упередженості. Модель повинна однаково справедливо взаємодіяти з усіма користувачами незалежно від статі, раси, віку чи соціального статусу. Наступною є прозорість – користувачі повинні мати доступ до інформації про те, як система працює, на яких даних вона навчалась, і які можливі обмеження її застосування [6].

Іншою критично важливою характеристикою є пояснюваність – здатність системи та її розробників надати чітке пояснення, чому було прийнято те чи інше рішення. Це особливо актуально у сферах з високим ступенем відповідальності, наприклад, у медицині, фінансах чи юридичній сфері. Також важливою є здатність до захисту конфіденційності, що передбачає дотримання принципів збору, зберігання та обробки персональних даних відповідно до етичних та правових стандартів [6].

Останньою є стійкість – здатність системи зберігати свою функціональність навіть за наявності помилок, атак чи непередбачуваних ситуацій. Надійні системи повинні мати вбудовані механізми відновлення та захисту від зовнішніх загроз [6].

Документ базується на п'яти ключових функціях: Govern, Map, Measure, Manage і Improve [6]. Ці функції утворюють узгоджену систему, яка допомагає організаціям забезпе-

чити надійність, безпечність і етичність ШІ-систем протягом усього їхнього життєвого циклу – від початкового проєктування до постійного оновлення.

Перший етап – Govern, він полягає у створенні чіткої організаційної структури та політик, що регулюють впровадження ШІ. Це включає визначення відповідальних осіб, розподіл повноважень, прозорість процесів і формування внутрішніх стандартів. Добре організоване управління дозволяє приймати обґрунтовані рішення, враховуючи етичні, правові та соціальні аспекти використання технологій [6].

Далі йде Map – функція, що допомагає оцінити контекст, у якому працює система. На цьому етапі організації ідентифікують можливі загрози, враховують потреби користувачів і визначають потенційний вплив ШІ на людей та навколишнє середовище. Важливо не лише знати, як працює система, а й розуміти, для кого і з якою метою вона створена [6].

Measure забезпечує вимірювання ризиків та надійності системи [6]. Це можуть бути кількісні та якісні показники. Наприклад, точність, стабільність, справедливість або захищеність моделі. Регулярна оцінка допомагає виявляти слабкі місця до того, як вони спричинять шкоду, і дає змогу приймати обґрунтовані рішення щодо подальших кроків.

Функція Manage передбачає реалізацію конкретних заходів для мінімізації виявлених ризиків [6]. Це може включати зміни в алгоритмах, обмеження доступу до деяких функцій, захист персональних даних або навіть перегляд стратегії впровадження. Важливо, що управління ризиками не є разовим завданням, а триває протягом усього періоду використання ШІ.

Завершує цей цикл функція Improve, яка зосереджена на постійному аналізі та вдосконаленні практик управління ризиками. Організації мають враховувати досвід, нові знання, технологічний розвиток та зміни в нормативному середовищі, щоб адаптувати свої ШІ-системи до нових викликів [6].

Ще одним документом є стандарт ISO/IEC 23894:2023. Це перший міжнародний стандарт, що надає рекомендації з управління ризиками, пов'язаними з використанням штучного інтелекту. Документ адаптує загальні принципи управління ризиками до специфіки ШІ-систем, враховуючи їхню складність, динамічну поведінку, етичні виклики та вплив на права людини. Головна мета стандарту – допомогти організаціям мінімізувати можливу шкоду, підвищити довіру до ШІ та зробити його застосування безпечнішим і більш передбачуваним [7].

У стандарті ISO/IEC 23894:2023 окреслюється структура ефективного управління ризиками штучного інтелекту, що охоплює весь життєвий цикл систем ШІ. Одним із перших і ключових етапів є ідентифікація ризиків, яка передбачає глибоке розуміння того, як система ШІ буде використовуватися на практиці. Тут організації повинні враховувати як передбачувані сценарії застосування, так і потенційні випадки неправильного або зловмисного використання. Важливо також ретельно проаналізувати дані, на яких навчалась модель, способи прийняття нею рішень та ймовірний вплив її функціонування на різні групи користувачів і суспільство загалом [7].

Після виявлення потенційних ризиків стандарт пропонує провести їх оцінку, як у кількісному, так і у якісному вимірі. Оцінювання має охоплювати не лише ймовірність виникнення проблеми, але й масштаб можливих наслідків [7]. Особливо підкреслюється необхідність врахування каскадних ефектів – тобто того, як одна проблема може спричинити інші, пов’язані ризики. Це дозволяє побудувати більш повну картину загроз, які може створювати система ШІ у взаємодії з іншими технологіями або соціальними процесами.

Наступним кроком є обробка ризиків, тобто розробка практичних стратегій для їх зменшення. Це може включати перегляд архітектури самої моделі, посилення технічного або організаційного контролю, страхування ризиків або ж свідоме прийняття певного рівня залишкової небезпеки. Головне, щоб вибрані підходи відповідали як характеру загрози, так і загальній стратегії організації щодо етики та безпеки [7].

Завершальним елементом системи управління ризиками є постійний моніторинг і перегляд. Оскільки системи ШІ розвиваються у часі та взаємодіють з динамічним середовищем,

ISO/IEC 23894 наголошує на необхідності регулярної переоцінки ризиків. Це включає встановлення ключових індикаторів ризику (KRI), перевірку ефективності раніше впроваджених заходів і оперативне оновлення стратегій у відповідь на зміни в технологічному або соціальному контексті [7].

Найбільш відкритими до технологій штучного інтелекту є Об'єднані Арабські Емірати (ОАЕ). ОАЕ демонструють швидкий і стратегічний підхід до регулювання штучного інтелекту. Країна прагне не лише використовувати ШІ у своїй економіці, але й стати глобальним лідером у сфері інноваційного управління. У 2017 р. ОАЕ першими у світі призначили міністра штучного інтелекту, підкреслюючи політичну волю до активного розвитку цієї галузі. Відтоді в країні діє Національна стратегія ШІ 2031, що ставить за мету зробити ШІ ключовим інструментом підвищення ефективності урядових послуг, охорони здоров'я, транспорту та освіти [8].

На відміну від багатьох інших юрисдикцій, ОАЕ не приймають детального законодавчого акту, подібного до європейського AI Act. Натомість, держава діє через галузеве регулювання та ініціативи публічно-приватного партнерства. Наприклад, у сфері фінансових послуг Центральний банк ОАЕ та інші регулятори вже впровадили політики використання ШІ, спрямовані на запобігання зловживанням, шахрайству та упередженості в автоматизованих рішеннях [8].

Цікаво, що в ОАЕ також активно діє Dubai International Financial Centre (DIFC) – спеціальна юрисдикція, яка самостійно впроваджує власні цифрові стандарти. У межах ініціативи DIFC AI and Data Protection Guidelines пропонується фреймворк із використанням ШІ, який поєднує етичні принципи, прозорість та відповідальність розробників. ОАЕ, таким чином, рухаються у напрямку «гнучкого регулювання», яке дозволяє адаптуватися до швидких технологічних змін без надмірної бюрократії [8].

Однак відсутність єдиного національного закону щодо ШІ створює ризик фрагментації правового середовища, особливо для міжнародних компаній, які працюють в різних еміратах або секторах. У перспективі ОАЕ можуть розглянути ухвалення більш цілісного законодавства, яке б забезпечило узгодженість підходів у всіх сферах і дало більше впевненості бізнесу щодо вимог до відповідального впровадження технологій штучного інтелекту.

3. Поточна ситуація регулювання штучного інтелекту в Україні

Україна розробила поетапну дорожню карту для впровадження регулювання штучного інтелекту, орієнтуючись на інтеграцію в європейський цифровий простір та імплементацію майбутніх стандартів ЄС, зокрема AI Act. Основною метою документа є створення такого підходу до ШІ, який поєднує захист прав людини, розвиток інноваційної економіки та підвищення міжнародної конкурентоспроможності українських компаній [9].

Документ пропонує bottom-up підхід, тобто рух від м'яких, позазаконодавчих механізмів до поступового впровадження законодавства. На першому етапі (планується два-три роки) передбачається створення регуляторного середовища: тестові механізми, добровільні кодекси поведінки, оцінка ризиків та публікація «Білої книги» – аналітичного документа з рекомендаціями для держави й бізнесу [9]. Це дозволить учасникам ринку адаптуватися до майбутніх вимог без надмірного тиску.

Другий етап передбачає поступову імплементацію положень європейського AI Act, зокрема в момент, коли це стане вимогою для подальшої інтеграції України до ЄС [9]. Планується адаптація найвимогливіших стандартів захисту прав людини, прозорості та етики у використанні ШІ, але з урахуванням специфіки національного ринку. Такий підхід дозволить Україні не лише забезпечити відповідність міжнародним вимогам, але й залишитись гнучкою та конкурентною на глобальному ринку.

У «Білій книзі» окреслено три стратегічні цілі: забезпечення прав людини, підтримка конкурентоспроможності бізнесу та євроінтеграція. Україна прагне гармонізувати майбутнє законодавство із нормами ЄС, аби надати українським ШІ-продуктам вільний доступ до рин-

ку Європи. Водночас у сфері оборони регулювання ШІ не передбачається, з огляду на військовий стан і потребу в інноваціях для захисту держави [10].

Особливу увагу в «Білій книзі» приділено балансу між правами людини та інноваційністю. Міністерство цифрових трансформацій України визнає: надмірне регулювання може загальмувати розвиток ШІ-індустрії, а його повна відсутність – створити загрози для прав і свобод громадян. Тому Україна орієнтується на сервісну модель держави, яка не тисне, а допомагає – через створення публічних інструментів, як-от веб-портал відповідального ШІ, платформа юридичної допомоги, інструмент добровільного маркування систем [10].

Методологія оцінки впливу ШІ на права людини стане базовим інструментом: вона допоможе визначити рівень ризику продукту і стане передумовою для участі в регуляторній пісочниці або отримання консультацій. Регуляторна пісочниця, своєю чергою, дозволить стартапам і компаніям тестувати свої рішення під наглядом держави, готуючись до майбутніх вимог. Також планується система добровільного маркування ШІ, подібна до етикеток на продуктах – аби користувачі знали, як саме працює система, і могли оцінити її надійність, безпеку та етичність.

4. Модель безпеки використання штучного інтелекту для України

Одним із ключових елементів регулювання штучного інтелекту є побудова системи безпеки, яка гарантує відповідальне, етичне та надійне використання ШІ. Для України така модель повинна враховувати як світові стандарти, так і національні виклики, пов’язані з війною, цифровою трансформацією та інтеграцією в європейський правовий простір. У цьому розділі подано узагальнений підхід до формування моделі безпеки використання ШІ в українських умовах, модель наведено на рис. 1.

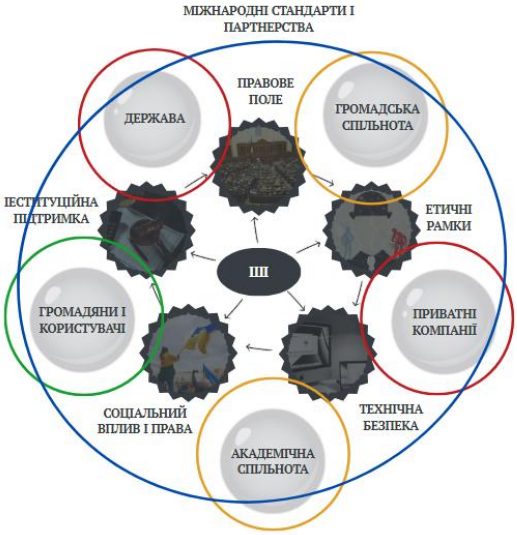

Рис. 1. Модель регулювання безпеки штучного інтелекту в Україні

Запропонована модель безпеки використання штучного інтелекту базується на системному підході, який передбачає взаємодію різних зацікавлених сторін і ключових сфер впливу. У центрі моделі знаходиться ШІ як технологічне ядро, довкола якого формуються міжсекторальні механізми регулювання, етики та технічної підтримки.

У моделі закладено взаємодію між державою, академічною спільнотою, громадянським суспільством, приватними компаніями, міжнародними партнерами та самими користувачами. Такий формат дозволяє не лише регулювати ШІ, а й формувати довіру до нього як до інструменту для розвитку, а не загрози. Україна – країна, що веде активну цифровізацію і не боїться впроваджувати нові технології. У цьому плані український громадський сектор та законодавча база є більш гнучкою і готовою до змін, аніж у сусідніх країнах Європейського Союзу.

Відповідно до рівня ризику, який може виникнути у тій чи іншій складовій її позначено відповідним кольором (де зелений – допустимий, жовтий – високий, червоний – підвищений). За концепцією, схожою з AI Act, цей поділ є необхідним для оцінки кожної моделі ШІ, яка використовується у тій чи іншій галузі. Україна могла б створити «відкритий етичний реєстр ШІ», куди кожна компанія добровільно додала б свою модель. І де кожна модель має власний «паспорт прозорості» – хто її створив, як навчав, які ризики враховані. Додатково кожна система проходила б оцінку на рівень ризику, відповідно до цього її власники мали б забезпечувати певний рівень безпеки, який відповідав би українському стандарту. Це не цензура, це – цифрова культура, адже користувачі мають знати, що саме використовують. Відповідно в уряді має бути створено спеціальний департамент з питань штучного інтелекту, який видавав би ліцензії безпеки та розглядав би і приймав усі потрібні нормативно-правові рішення. Також одним із важливих рішень має бути обов'язкове маркування продуктів та контенту, створеного штучним інтелектом, задля запобігання шахрайства та недоброчесності.

Окрему роль у цій моделі може відігравати громадська спільнота, залучена до моніторингу та аудиту ШІ-систем. Запровадження публічних механізмів контролю, таких як цифрові платформи для фіксації порушень або несправедливих рішень, дозволить забезпечити не лише технічну, а й соціальну безпеку. Водночас державні сервіси – зокрема платформа «Дія» – можуть стати прикладом прозорого, етичного використання ШІ в публічному управлінні, де кожен громадянин має доступ до зрозумілих пояснень рішень, прийнятих алгоритмом.

Ключову роль також відіграє академічна спільнота, яка здатна не лише досліджувати ризики, а й формувати освітню культуру довкола ШІ. Впровадження національного освітнього треку з етики та технологій, інтеграція відповідних курсів у навчальні програми та підтримка молодих дослідників допоможе створити критичну масу спеціалістів, здатних формувати відповідальне майбутнє ШІ в Україні.

В українському контексті приватні компанії не обмежуються впровадженням технологій – вони стають співавторами національної цифрової безпеки. Особливо у сфері штучного інтелекту бізнес має не лише комерційні, а й моральні зобов'язання: перед клієнтами, державою, суспільством і навіть перед власними працівниками. Українські компанії вже довели, що можуть бути лідерами в етичному програмуванні, відкритих кодах, благодійних проєктах та безпечних цифрових рішеннях.

Майбутнє моделі безпеки ШІ в Україні передбачає особливий соціальний договір між державою та бізнесом: не через штрафи чи контроль, а через довіру, сертифікацію, кооперацію та інноваційне партнерство. Компанії, які добровільно дотримуються стандартів прозорості, маркують свої моделі, відкривають алгоритми на ревізію, можуть отримати «цифрову довіру» – умовне етичне маркування, схоже на ISO, але у сфері штучного інтелекту. Водночас важливо підтримати компанії не тільки вимогами, а й ресурсами та середовищем. Наприклад, держава може створити інкубатори відповідального ШІ, де стартапи отримуватимуть доступ до відкритих даних, консультацій з етики, шаблонів політик – без тиску, без формальностей, з орієнтацією на співпрацю.

Ще однією важливою частиною запропонованої моделі безпеки використання ШІ є відкритість до міжнародної співпраці. Це включатиме обмін даними, впровадження міжнародних стандартизацій та регуляторних актів у інтегрованому для України форматі.

У контексті безпеки штучного інтелекту особливу увагу слід приділяти технічним практикам розробки та впровадження ШІ-рішень у приватному секторі, зокрема в компаніях, які створюють системи для використання в соціально чутливих сферах: оборона, фінанси, охорона здоров'я, освіта. Український бізнес сьогодні має справу не лише з комерційними викликами, а й із загрозами кібератак, викрадення моделей, отруєння даних та спробами маніпулювання алгоритмами. Ключовими напрямами технічної безпеки в межах української моделі ШІ можуть стати: безпечне навчання моделей, моніторинг поведінки моделей у

реальному часі, розмежування доступу до компонентів ШІ систем, впровадження explainable AI та контейнери для тестування.

Окремої уваги заслуговує питання інфраструктурної стійкості: більшість українських IT-компаній розміщують свої сервіси в хмарі, часто – за кордоном. В умовах воєнних ризиків важливо створити резервні дата-центри, енергонезалежні вузли або використовувати «розподілену відповідальність» за безпеку з партнерами по хмарним сервісам, із обов'язковим аудитом.

Таким чином, запропонована модель безпеки використання ШІ надаватиме простір для розвитку та новаторства, але при цьому увесь процес від створення моделей штучного інтелекту до їх подальшого безпечного використання буде контрольованим та безпечним. Відкритість до міжнародної співпраці тільки підкреслює намір України підтримувати європейські стандарти безпеки та при цьому дасть можливість ділитись власним досвідом з міжнародними партнерами.

Висновки

У результаті проведеного дослідження було проаналізовано міжнародні підходи до регулювання та управління безпекою штучного інтелекту, зокрема стандарти ЄС, США, ISO/IEC та моделі, що впроваджуються в Об'єднаних Арабських Еміратах. Встановлено, що кожна з цих систем має свої сильні сторони, але не є універсальною для застосування в українському контексті. Найбільш прийнятним для України є гібридний підхід, який поєднує сервісну, гнучку модель співрегулювання з орієнтацією на права людини та прозорість, характерну для європейського AI Act.

Особливої уваги потребує побудова власної національної моделі безпеки ШІ, яка враховуватиме специфіку воєнного часу, цифрової трансформації, високої ролі громадянського суспільства та унікального українського досвіду цифрового спротиву. Запропонована модель взаємодії між державою, бізнесом, академічною спільнотою, громадянами та міжнародними партнерами створює основу для формування довіри до ШІ та ефективного управління ризиками на всіх етапах життєвого циклу технологій.

Окрема роль у цій системі належить приватному сектору, який може не лише впроваджувати етичні стандарти, а й бути рушієм цифрової культури безпеки. Важливо, щоб держава підтримувала інноваційні ініціативи через інкубатори відповідального ШІ, сертифікування та механізми прозорості, що спростять адаптацію бізнесу до майбутнього регулювання.

Таким чином, Україна має реальну можливість не лише гармонізувати своє законодавство із міжнародними стандартами, а й запропонувати власний – інноваційний та адаптивний – підхід до безпеки штучного інтелекту. Це дозволить країні стати повноцінним учасником глобального цифрового ринку, водночас зберігаючи пріоритет прав людини, технологічну відкритість та національну стійкість.

Список літератури:

1. Cost of a Data Breach Report 2024. [Електронний ресурс]. Режим доступу: https://www.ibm.com/reports/data-breach.

2. IBM Report: Escalating Data Breach Disruption Pushes Costs to New Highs. [Електронний ресурс]. Режим доступу: https://newsroom.ibm.com/2024-07-30-ibm-report-escalating-data-breach-disruption-pushes-costs-to-new-highs?utm_source=chatgpt.com.

3. Атаки на основі штучного інтелекту: нові виклики для кібербезпеки. [Електронний ресурс]. Режим доступу: https://wezom.com.ua/ua/blog/ataki-na-osnovi-shtuchnogo-intelektu-novi-vikliki-dlya-kiberbezpeki.

4. 5 реальних прикладів хакерських атак за допомогою ШІ. [Електронний ресурс]. Режим доступу: https://dev.ua/news/5-prykladiv-khakerskykh-atak-ai.

5. EU AI Act: first regulation on artificial intelligence. [Електронний ресурс]. Режим доступу: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

6. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. [Електронний ресурс]. Режим доступу: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

7. Information technology – Artificial intelligence – Guidance on risk management [Електронний ресурс]. Режим доступу: https://cdn.standards.iteh.ai/samples/77304/cb803ee4e9624430a5db177459158b24/ISO-IEC-23894-2023.pdf.

8. AI Watch: Global regulatory tracker – United Arab Emirates. [Електронний ресурс]. Режим доступу: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-uae.

9. Дорожня карта з регулювання штучного інтелекту в Україні Bottom-Up Підхід. [Електронний ресурс]. Режим доступу: https://surli.cc/tyzbug.

10. Біла книга з регулювання ШІ в Україні: бачення Мінцифри. Режим доступу: https://thedigital.gov.ua/storage/uploads/files/page/community/docs/Регулювання%20ШІ.pdf.

Надійшла до редколегії 02.02.2025

Відомості про авторів:

Логачова Єлизавета Олегівна – Харківський національний університет імені В. Н. Каразіна, студентка кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту; Україна; e-mail: lohachova2020kb11@student.karazin.ua; ORCID: https://orcid.org/0000-0002-9815-466X

Єсіна Марина Віталіївна – канд. техн. наук, доцент, Харківський національний університет імені В. Н. Каразіна, в.о. завідувача кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту, АТ Інститут Інформаційних Технологій", науковий співробітник-консультант; Україна; e-mail: m.v.yesina@karazin.ua; ORCID: https://orcid.org/0000-0002-1252-7606

Голубничий Дмитро Юрійович – канд. техн. наук, доцент, АТ "Інститут Інформаційних Технологій", начальник наукового відділу; Україна; ORCID: https://orcid.org/0000-0002-6873-7004