DOI: https://doi.org/10.32782/mer.2025.105.07

УДК 004.056:330.43

МОДЕЛЮВАННЯ ВПЛИВУ СИСТЕМ ВИЯВЛЕННЯ КІБЕРЗАГРОЗ НА РІВЕНЬ КІБЕРЗБИТКІВ: DID-АНАЛІЗ ТА ПІДХІД ДОСЛІДЖЕННЯ ПОДІЙ

Койбічук Віталія Василівна
кандидат економічних наук, доцент,
завідувач кафедри економічної кібернетики,
Сумський державний університет
ORCID: https://orcid.org/0000-0002-3540-7922

Єфіменко Аліна Юріївна
доктор філософії, асистент кафедри економічної кібернетики,
Сумський державний університет
ORCID: https://orcid.org/0000-0002-2810-0965
E-mail: a.yefimenko@uabs.sumdu.edu.ua

Стрімке поширення технологій штучного інтелекту у сфері кібербезпеки супроводжується парадоксальним ефектом: системи AI-based Detection підвищують ефективність виявлення кіберінцидентів, однак не гарантують повного запобігання фінансовим збиткам від кібератак. У зв'язку з цим виникає методологічна проблема коректної інтерпретації динаміки кіберзбитків після впровадження інтелектуальних систем захисту. Метою дослідження є економетричне оцінювання каузального впливу використання систем виявлення кіберзагроз на основі штучного інтелекту на рівень кіберзбитків. Емпіричний аналіз базується на наборі даних, що охоплює кіберінциденти у дев'яти країнах світу (Австралії, Бразилії, Китаю, Франції, Німеччини, Індії, Японії, Великої Британії, Сполучених Штатів Америки) за період 2017–2024 роки. У роботі запропоновано авторський індекс кіберзбитків, сформований на основі інтеграції фінансових втрат від кібератак та кількості постраждалих користувачів. Для підвищення статистичної стабільності розподілів застосовано логарифмування показників і стандартизацію за Z-score. Для оцінювання причинно-наслідкового ефекту використано економетричний підхід Difference-in-Differences, а також його динамічне розширення у вигляді event-study аналізу, що дозволяє дослідити часову траєкторію впливу впровадження систем штучного інтелекту AI-based Detection. Отримані результати свідчать, що використання систем штучного інтелекту має статистично значущий вплив на динаміку кіберзбитків. У короткостроковому періоді після впровадження AI-based Detection спостерігається зростання індексу кіберзбитків, що пояснюється підвищенням рівня виявлення та фіксації кіберінцидентів, які раніше могли залишатися непоміченими. Водночас у середньостроковій перспективі (через 3–4 роки після впровадження) формується стійка тенденція до зниження кіберзбитків, що відображає поступове підвищення ефективності AI-систем, їх адаптацію та інтеграцію у кіберзахисну інфраструктуру. Отримані результати демонструють, що впровадження технологій штучного інтелекту у сфері кібербезпеки змінює не лише рівень захисту, але й сам механізм вимірювання кіберризиків. Запропонований підхід дозволяє більш коректно оцінювати ефективність AI-based систем кіберзахисту та рекомендований для аналізу політик цифрової безпеки та управління кіберризиками.

Ключові слова: *аналіз «різниця-в-різницях», дослідження подій, паралельні тренди, технології штучного інтелекту, індекс кіберзбитків.*

Постановка проблеми. В умовах сьогодення стрімка цифровізація економічних та суспільних процесів почала супроводжуватися зростанням масштабів, частоти та складності кіберінцидентів, що провокують значні фінансові втрати, порушення безперервності бізнес-процесів та зниження довіри до цифрових сервісів. У відповідь на ці виклики держави та великі організації дедалі активніше впроваджують системи кіберзахисту, що базуються на технологіях штучного інтелекту, позиціонуючи їх як інструмент підвищення ефективності виявлення кіберзагроз і мінімізації кіберзбитків. Варто зазначити, що станом на кінець 2025 року 20% підприємств країн Європейського Союзу використовували технології штучного інтелекту (Eurostat, 2025). Водночас зростання рівня проникнення ШІ супроводжується дискусіями щодо виникнення нових кібербезпекових ризиків, пов'язаних із підвищенням складності алгоритмічних систем, залежністю критичних бізнес-процесів від автоматизованих рішень та розширенням поверхні атаки.

Alblehai, 2025 зазначає, що сучасна парадигма кібербезпеки ґрунтується на активному використанні ШІ-технологій для побудови високорівневих і адаптивних моделей захисту інформаційних систем і мереж, зокрема в середовищах Інтернету речей. Водночас кібератаки, які також інтегрують ШІ-рішення, набувають здатності до самонавчання та динамічної адаптації, що сприяє еволюції більш складних і витончених загроз у кіберпросторі.

Паралельно із використанням ШІ-технологій для оцінки кіберризиків виникають кібератаки, що застосовують зазначені технології. Ramirez, 2025 зазначала,

що глобальні витрати, пов'язані з кібератаками, які використовують штучний інтелект, станом на кінець 2025 року перевищать $193 млрд. Середня вартість одного інциденту, пов'язаного з AI, оцінюється в $5,72 млн, що на 13 % більше, ніж у 2024 році.

За даними Глобального звіту DevSecOps за 2024 рік, організації з широким використанням AI та автоматизації мають значно нижчі середні витрати на наслідки витоку даних – 3,76 млн дол., тоді як ті, що не використовують таких інструментів, мають витрати 5,98 млн дол., тобто різниця складає близько 45,6 % (Gitlab, 2024). Це демонструє потенційний позитивний вплив AI-захисту на зменшення фінансових втрат.

Ключова проблема полягає в тому, що зростання зафіксованих кіберзбитків після впровадження ШІ-рішень відображає не реальне погіршення кібербезпекової ситуації, а підвищення спроможності до виявлення, обліку та оцінювання кіберінцидентів і їхніх наслідків за допомогою ШІ-технологій. Це ускладнює інтерпретацію динаміки індексів кіберзбитків та створює методологічні ризики помилкових висновків щодо ефективності політик та технологічних рішень у сфері кібербезпеки.

Таким чином, проблематика обраної теми полягає у відсутності чітко ідентифікованих каузальних оцінок впливу впровадження систем кіберзахисту на основі штучного інтелекту на динаміку кіберзбитків у просторово-часовому вимірі з урахуванням адаптаційних ефектів і нелінійного характеру післявпроваджувальної динаміки. У зв'язку з цим актуалізується потреба у застосуванні чітких економетричних підходів каузальної ідентифікації, які дозволяють відокремити ефект технологічного втручання від загальносвітових трендів та структурних зрушень, а також простежити динамічний характер впливу ШІ-систем кіберзахисту у доперіоді та після їх впровадження.

Огляд літератури. Не дивлячись на велику кількість публікацій, що використовують методологію «різниця в різницях» (DiD-аналіз), зазначена методологія є актуальною та використовуються в різних сферах дослідження – від оцінювання економічної політики до аналізу технологічних та інституційних змін. Популярність підходу демонструє динаміка публікацій (рис.1), що індексуються базою даних Scopus за останні п'ять років. Так, кількість публікацій з 2020 року з 243 документів зросла до 544 у 2025, а їх загальна кількість за зазначений період (з 2020 по 2025 рік) склала 2,194 одиниці.

Оскільки штучний інтелект (ШІ) змінює бізнес-процеси, фірми повинні адаптувати свої стратегії навчання для формування кваліфікованої робочої сили. Таке бачення в якості політики втручання висвітлює в своїй праці Muehlemann Samuel (2025). Автор на підгрунті панельних даних німецьких установ за 2019 по 2023 рік за допомогою DiD аналізу досліджує як впровадження штучного інтелекту (ШІ) впливає на стратегії навчання та професійної підготовки працівників у фірмах. Його результати показали, що після впровадження ШІ компанії, що вже проводили підготовку, збільшили кількість нових практикантів приблизно на 14%. Тобто такі фірми частіше інвестують у довгострокову підготовку молодих працівників. Також дослідження показало, що суттєвого впливу на рішення про тренінги загалом не спостерігалось, тобто впровадження ШІ не змінило сам факт того, чи компанія взагалі вирішує надавати тренінги або наймати нових працівників, однак замість низькокваліфікованого навчання, компанії з ШІ збільшили тренінги для висококваліфікованих працівників.

Не менш цікавою є праця (Polizzi et al., 2023), де науковці проводять DiD-аналіз за 2010-2019 роки на прикладі компаній, що входять до індексу Euro Stoxx 50 (відображає топ-50 компаній Єврозони) для вивчення впливу Паризької угоди (Конференція Організації Об'єднаних Націй зі зміни клімату, COP21, Paris Agreement (United Nations Climate Change Conference, COP21) та французького Закону 2015-992 (the French Law 2015-992) про енергетичний перехід для зеленого зростання. Їхні результати продемонстрували, що обидва регуляторні втручання сприяли покращенню розкриття інформації компаніями про навколишнє середовище. Також автори відмітили, що фірми, що належать до найбільш забруднюючих секторів, як правило, надають більше інформації з питань навколишнього середовища, ймовірно, намагаючись відвернути увагу зацікавлених сторін.

Мета дослідження полягає у розробленні моделей щодо каузального впливу систем виявлення кіберзагроз, що використовують технології штучного інтелекту на рівень кіберзбитків, за допомогою методології «різниця в різницях» та її динамічного розширення у формі event-study аналізу.

Методологія дослідження. Зміст методології «різниця в різницях» (Difference-in-Differences, DiD) полягає у економетричному оцінюванні причинно-наслідкового впливу певного втручання (політики чи програми), що ґрунтується на зіставленні змін показника результату в групі, яка зазнала втручання, зі змінами відповідного показника в контрольній групі за той самий часовий проміжок. Такий підхід дає змогу визначити, наскільки динаміка показників у групі впливу після реалізації втручання відхилилася від динаміки контрольної групи. Перша «різниця» відображає зміни показника до та після втручання в межах кожної групи, тоді як «різниця в різницях» порівнює ці зміни між групами, що дозволяє виокремити чистий ефект втручання, усуваючи вплив спільних часових тенденцій.

Підхід event-study являє собою динамічну DiD-специфікацію, у якій ефект взаємодії між втручанням (treatment) та часом оцінюється окремо для кожного року відносно моменту події (Callaway et al., 2021).

Запропоновано провести DiD-аналіз для дев'яти країн світу: Австралії, Бразилії, Китаю, Франції, Німеччини, Індії, Японії, Великої Британії, Сполучених Штатів Америки, з подальшим його розширенням у вигляді event-study аналізу.

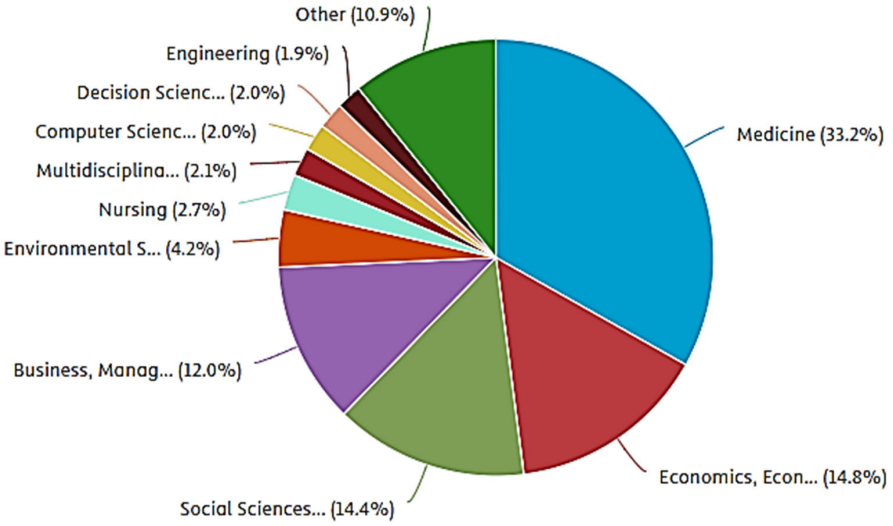

Рисунок 1 – Розподіл публікацій за галузями, що використовують DiD-аналіз та індексуються базою даних Scopus

В якості вхідних даних використано офіційний датасет Global Cybersecurity Threats з відкритої бази даних DeepDataLake про відстеження інцидентів кібербезпеки, векторів атак та загроз за період з 2017 по 2024 рік, що містить й інші (25+) високоякісні набори даних, призначені для машинного навчання, аналізу даних та дослідницьких проєктів. Усі набори даних дозволено використовувати безкоштовно за відкритими ліцензіями. База даних Global Cybersecurity Threats містить інформацію про виявлені типи кібератак (K1 – Attack_Type), зокрема phishing, DDoS, Malware, Ransomwarev, Man-in-the-Middle, SQL Injection; рік виявлення (Year); цільові галузі (K2 – Target_Industry); фінансові збитки від кібератак (млн $ US) (K3 – Financial_Loss); кількість уражених користувачів (K4 – Affected_Users); джерело кібератаки (K5 – Attack_Source); тип вразливості безпеки (K6 – Security_Vulnerability_Type); технології, що використовувались для виявлення кібератаки (K7 – Defense), час вирішення інциденту (K8 – Incident_Resolution_Time). Фрагмент вхідної вибірки даних наведено в таблиці 1.

У якості treatment-групи визначено кіберінциденти в дев'яти країнах, для яких застосовувався механізм виявлення з використанням технологій, що використовують штучний інтелект (AI-based Detection), але вплив цих кіберінцидентів настав. Всі інші механізми визначення впливу кіберінцидентів формують контрольну групу.

Ключовим припущенням DiD-аналізу є те, що за відсутності впровадження AI-based Detection динаміка індексу кіберзбитків у treatment- та control-групах була б однаковою.

Моментом втручання в даному дослідженні вважається 2020 рік, що відповідає періоду активного поширення та практичного впровадження AI-рішень у сфері кібербезпеки. Варто зазначити, що саме у 2020 році було опубліковано рейтинг, який оцінював готовність урядів різних країн до впровадження штучного інтелекту (Oxford Insights, 2020).

Отже, першим етапом для проведення DiD-аналізу було здійснено розрахунок індексу кіберзбитків на підґрунті інформації, що подана в DeepDataLake, 2025. Наразі не існує загальноусталеної методології щодо визначення індексу кіберзбитків. Так, наприклад, науковці Callaway et al., 2021 виділяють ключові напрямки для його визначення. Зокрема, індекс кіберзбитків (зокрема PERILS / CyberAcuView US Cyber Industry Loss Index) визначають як узагальнений фінансовий індикатор, призначений для моніторингу масштабних втрат від кіберінцидентів на страховому ринку, що перевищують 500 млн доларів США. Його основною функцією є формування інформаційної бази для інструментів трансферу ризиків, зокрема страхових цінних паперів (Insurance-Linked Securities, ILS) та галузевих гарантій збитків (Industry Loss Warranties, ILW), що дозволяє кількісно оцінювати системний кіберризик, який не обмежується окремими випадками порушень безпеки. Водночас альтернативні індекси, такі як Chubb Cyber Index, фокусуються на відображенні галузевих тенденцій і витрат підприємств, пов'язаних із кіберстраховими вимогами та претензіями.

Для визначення індексу кіберзбитків, запропонуємо власне авторське бачення. Індикатор K3 (Financial_Loss) відображає грошові втрати від кіберінциденту, а K4 (Affected_Users) в таблиці 1 – масштаб атаки (скільки користувачів постраждало). Обидва індикатори характеризують, наскільки серьозною була атака, тому композитний показник кіберзбитків запропоновано побудувати саме з використанням індикаторів K3 та K4. Їхню побудову здійснено на основі логарифмованих показників K3 та K4 за допомогою формули (1):

$$CyberLossIndex_i = \frac{z\_fin\_loss_i + z\_affected_i}{2} \quad (1)$$

Таблиця 1 – Глобальні кіберзагрози (фрагмент вхідної вибірки даних)

Country	Year	K1	K2	K3	K4	K5	K6	K7	K8
1	2	3	4	5	6	7	8	9	10
UK	2024	Ransomware	Telecommunications	41.44	659320	Nation-state	Social Engineering	AI-based Detection	7
France	2024	DDoS	IT	48.01	922258	Unknown	Social Engineering	Encryption	64
Australia	2024	Phishing	Education	93.32	93185	Unknown	Unpatched Software	VPN	14
India	2024	DDoS	Education	73.37	475719	Nation-state	Social Engineering	Firewall	52
China	2024	Malware	Education	72.35	132726	Hacker Group	Zero-day	Firewall	55
France	2024	Phishing	Government	5.8	508649	Hacker Group	Unpatched Software	Firewall	66
Australia	2024	Malware	Healthcare	62.08	969504	Hacker Group	Social Engineering	AI-based Detection	5
UK	2024	DDoS	Healthcare	58.31	487694	Hacker Group	Weak Passwords	Antivirus	41
Germany	2024	Phishing	Healthcare	84.53	189430	Nation-state	Social Engineering	Firewall	1
UK	2024	Phishing	Government	81.26	993903	Unknown	Weak Passwords	Firewall	25
China	2024	Ransomware	Telecommunications	7.32	179156	Nation-state	Social Engineering	AI-based Detection	34
Germany	2024	Phishing	IT	3.93	845785	Unknown	Unpatched Software	VPN	69
Australia	2024	Ransomware	Education	85.86	461012	Insider	Unpatched Software	AI-based Detection	38
France	2024	Ransomware	Retail	99.03	349008	Nation-state	Zero-day	Encryption	20
Australia	2024	Ransomware	Government	39.92	594424	Hacker Group	Social Engineering	Antivirus	54
China	2024	Ransomware	Telecommunications	70.47	549337	Unknown	Social Engineering	Antivirus	63
Germany	2024	Malware	Banking	22.84	797768	Insider	Zero-day	Firewall	58
France	2024	Ransomware	Banking	78.82	759804	Unknown	Unpatched Software	AI-based Detection	47
Australia	2024	Phishing	Retail	73.43	897569	Hacker Group	Social Engineering	AI-based Detection	66
China	2024	Phishing	Education	4.01	841054	Insider	Zero-day	Antivirus	49
Germany	2024	Ransomware	Education	11.81	418893	Insider	Social Engineering	Firewall	31
Brazil	2024	DDoS	Banking	55.73	750979	Unknown	Social Engineering	AI-based Detection	30
India	2024	Malware	Retail	56.19	114357	Insider	Zero-day	AI-based Detection	58
Japan	2024	Man-in-the-Middle	Healthcare	72.08	261230	Insider	Social Engineering	VPN	19
…	…	…	…	…	…	…	…	…	…
India	2017	Man-in-the-Middle	IT	38.65	605895	Hacker Group	Weak Passwords	VPN	20
Germany	2017	Man-in-the-Middle	Retail	98.24	285201	Unknown	Social Engineering	Antivirus	25
India	2017	DDoS	Government	76.71	246205	Nation-state	Unpatched Software	Firewall	30
USA	2017	DDoS	IT	74.12	403108	Hacker Group	Zero-day	VPN	64
Japan	2017	Phishing	Telecommunications	75.84	383395	Hacker Group	Zero-day	VPN	31
France	2017	Phishing	Banking	17.88	16585	Hacker Group	Unpatched Software	Encryption	12
Australia	2017	SQL Injection	Healthcare	68.91	634048	Hacker Group	Unpatched Software	Firewall	66
UK	2017	SQL Injection	IT	53.39	576110	Insider	Unpatched Software	Antivirus	37
UK	2017	Man-in-the-Middle	Government	48.9	177154	Unknown	Zero-day	Antivirus	68

Продовження Таблиці 1

1	2	3	4	5	6	7	8	9	10
France	2017	SQL Injection	Healthcare	60.77	212177	Hacker Group	Social Engineering	Encryption	2
Japan	2017	SQL Injection	Healthcare	21.66	277555	Unknown	Social Engineering	AI-based Detection	45
UK	2017	DDoS	IT	9.01	617543	Unknown	Zero-day	Antivirus	43
USA	2017	SQL Injection	Banking	9.08	895507	Unknown	Social Engineering	AI-based Detection	52
Australia	2017	Malware	Healthcare	77.32	563706	Nation-state	Unpatched Software	AI-based Detection	35
Australia	2017	Man-in-the-Middle	Telecommunications	79.72	799971	Insider	Zero-day	Firewall	63
India	2017	Man-in-the-Middle	Government	69.69	713419	Hacker Group	Unpatched Software	AI-based Detection	70
UK	2017	Phishing	Telecommunications	27.08	416833	Insider	Social Engineering	VPN	9
Brazil	2017	Malware	Banking	99.22	660182	Insider	Weak Passwords	AI-based Detection	21
Japan	2017	Phishing	IT	65.62	286208	Hacker Group	Zero-day	Firewall	57
USA	2017	Malware	IT	75.53	992377	Nation-state	Unpatched Software	Antivirus	33
Japan	2017	Man-in-the-Middle	Retail	74.12	559889	Nation-state	Zero-day	VPN	10

Джерело: побудовано авторами на основі DeepDataLake, 2025

де $z\_fin\_loss_i$ та $z\_affected_i$ стандартизовані значення логарифмованих фінансових збитків та логарифмованих значень кількості уражених користувачів. Процедуру стандартизації проведено з використанням класичної Z-score формули (2) (Tyagi and Tyagi, 2026):

$$z = \frac{x - \bar{x}}{sd(x)} \tag{2}$$

Логарифмування дозволило зменшити вплив екстремальних інцидентів та зробити розподіли ближчими до нормальних У первинних (нелогарифмованих) даних спостерігалися високі значення коефіцієнта асиметрії та ексцесу, що свідчило про наявність довгого правого хвоста та надмірної концентрації маси розподілу. Після застосування логарифмічного перетворення значення коефіцієнта асиметрії зменшилося за абсолютною величиною, а коефіцієнт ексцесу наблизився до нуля, що підтверджує зниження відхилення від нормального розподілу та коректність подальшого економетричного аналізу.

Класичний Difference-in-Differences аналіз оцінює один середній ефект «до–після» втручання:

$$Cyber_{Lossit} = \alpha + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_t + \\ + \beta_3 (\text{Treat}_i \times \text{Post}_t) + \varepsilon_{it} \tag{3}$$

де $Cyber\_Loss_{it}$ – індекс кіберзбитків (Cyber Loss Index) для спостереження i (окремий кіберінцидент) у році t (2017–2024); α – середній рівень індексу кіберзбитків у контрольній групі до впровадження виявлення кібератаки на основі штучного інтелекту (AI-based Detection); тобто контрольна група (treat = 0), pre-treatment період (post = 0);

β_1 – базова (початкова) різниця між групами країн, що показує, наскільки індекс кіберзбитків у treatment-групі відрізнявся від контрольної групи до втручання (до використаня AI-detection);

β_2 – загальний часовий ефект, відображає зміну індексу кіберзбитків після 2020 року, що спільна для всіх механізмів захисту (зростання складності кібератак, зміна регуляторного середовища);

ε_{it} – випадкова похибка, (різні фактори, що не спостерігались та впливають на кіберзбитки і не включені явно в модель. Передбачається, що

$$E(\varepsilon_{it} | \text{Treat}_i, \text{Post}_t) = 0 \tag{4}$$

Індикатор втручання (Treat_i) відображає структурну відмінність між кіберінцидентами з AI-детекцією та без неї до 2020 року та визначається як:

$$\text{Treat}_i = \begin{cases} 1, \textit{якщо використовується AI – based Detection} \\ 0, \quad \textit{не використовується AI – based Detection} \end{cases} \tag{5}$$

Індикатор пост-періоду (Post_t) фіксує загальну зміну у кіберсередовищі після 2020 року, що однаково впливає на всі механізми захисту:

$$\text{Post}_i = \begin{cases} 1, t \geq 2020 \\ 0, \ t < 2020 \end{cases} \tag{6}$$

Отже DiD-взаємодія визначається добутком індикаторів $\text{Treat}_i \times \text{Post}_t$, та відображає середній ефект обробки виявлення кібератак на основі штучного інтелекту за припущення паралельних трендів, а коефіцієнт β_3 в формулі (1) інтерпретується як середній каузальний ефект втручання:

$$\beta_3 = \left(Y^{T,post} - Y^{T,pre}\right) - \left(Y^{C,post} - Y^{C,pre}\right) \tag{7}$$

де $Y^{T,post} - Y^{T,pre}$ – зміна у treatment (у групі країн, в яких для виявлення кібератак та визначення кіберзбитків використовувались технології ШІ), $Y^{C,post} - Y^{C,pre}$ – зміна у control (у групі країн, в яких для виявлення кіберінцидентів та визначення кіберзбитків використовувались інші технології).

З математичної точки зору event-study модель подано формулою (8):

$$Y_{it} = \alpha + \sum_{k \neq -1} \beta_k \left(\text{Treat}_i \times 1\{t - T_0 = k\} \right) + \gamma_t + \delta_i + \varepsilon_{it} \quad (8)$$

де T_0 – рік події (2020); k – кількість років до або після події ($k = -1$ виключається як базовий), β_k – динамічний treatment-ефект у році k.

Результати. Практичну реалізацію моделей (3, 8) здійснено у програмному середовищі RStudio з використанням пакетів «readxl» (для завантаження вхідного датасету Global Cybersecurity Threats), «dplyr» (для підготовки та трансформації даних перед оцінюванням моделі (8), а саме для фільтрації спостережень (pre/post, treted/control), створення змінних події (event time), групування даних (за країнами/роками)), обчислення агрегованих показників), «lmtest» (для перевірки статитичної значущості DiD та event-study моделей), «sandwich» (для обчислення робастних (стійких) стандартних помилок), «ggplot2» (для візуалізації динамічного ефекту використання ШІ-технологій для виявлянні кіберінцидентів й визначення значення індексу кіберзбитків та наглядного представлення паралельних трендів до втручання в обох групах), «broom» (для структурування результатів регресії та візуалізації динамічних ефектів event-study).

Отримані результати DiD-аналізу щодо оцінювання ефекту використання ШІ-технологій на виявлення кіберінцидентів та визначення кіберзбитків засвідчили про наявність ефекту на рівні граничної статистичної значущості, що вказало на потенційну часову неоднорідність впливу (таблиця 2).

До впровадження технологій, що використовують ШІ для виявлення кіберінцидентів та визначення кіберзбитків (табл. 2) treatment-група вже мала нижчий рівень кіберзбитків, ніж контрольна (treat = -0.190, p-value < 0.01). У контрольній групі після 2020 року суттєвих змін у кіберзбитках не відбулося (post не значущий), тобто зовнішні фактори, такі як пандемія, загальне зростання розвитку цифрових процесів не пояснюють ефект втручаня. Зміни зосереджені саме у treatment-групі.

Середня зміна індексу кіберзбитків після впровадження AI-based Detection у treatment-групі відносно контрольної є позитивною – 0.229 (усереднений ефект по всіх post-періодах) та свідчить про наявність відносної інтенсивності вимірюваних кіберінцидентів у групі об'єктів, які впровадили AI-based detection, після моменту втручання.

Отримані результати event-study моделі наведено в таблиці 3.

Отже, як видно з таблиці 3, базовий ефект AI-based Detection (treat) є статистично значущим (p-value < 5%) та в середньому (поза динамікою за роками) використання ШІ-технологій для виявлення кібератак та визначення кіберзбитків свідчить про зниження індексу кіберзбитків приблизно на 0.34 одиниці. Тобто країни та організації, що застосовують AI-based Detection, зазнають суттєво менших кіберзбитків, навіть без урахування конкретного року впровадження. Також на основі значень коефіцієнтів pre-event взаємодії до втручання (до впровадження технологій ШІ для виявлення кібератак), бачимо, що до впровадження AI-based Detection тренди кіберзбитків у treatment та control групах не відрізнялися, оскільки жоден із коефіцієнтів не є статистично значущим (year-3, year-2, AI × Year -3, AI × Year -2). Отже, припущення паралельних трендів виконується, а DiD-оцінки є валідними. У рік переходу на AI-рішення та наступний рік (2021-й) бачимо короткостроковий ефект, спостерігається структурний злам у динаміці кіберзбитків AI × Year 0 = = 0.537 (p-value < 0.01), AI × Year +1 = 0.437 (p-value < 0.01), що пов'язано зі зростанням кількості виявлених кіберінцидентів та перехідним періодом (витрати на імплементацію ШІ). В 2021 році ефект від впровадження є дещо меншим, проте статистично значущим, відповідно зробимо висновки, що ефективність AI-based Detection не є миттєвою, а формується поступово. У середньостроковій перспективі ефект стабілізується, через 3–4 роки (2023, 2024 роки) після впровадження AI-based Detection спостерігається стійкий та значущий вплив на зменшення кіберзбитків, що підтверджує якість навчання ШІ-систем, інституційну адаптацію та кращу інтеграцію ШІ у кіберзахисну інфраструктуру.

Отже, розроблена модель (8) є статистично значущою, що підтверджують також результати F-критерію (p-value < 0.01), незважаючи не низьке значення коефіцієнту детермінації. Зауважимо, що у DiD-

Таблиця 2 – Результати DiD-аналізу щодо оцінювання індексу кіберзбитків

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.00394	0.0276	0.1426	0.8866
treat	-0.1901	0.06757	-2.8141	0.0049**
post	0.0079	0.0349	0.2268	0.8206
treat:post	0.2286	0.08298	2.7544	0.0054**

Джерело: побудовано авторами з використанням RStudio

Примітки: significance of p-value *p<0.1; **p<0.05; ***p<0.01

Таблиця 3 – Дослідження подій для залежної змінної індексу кіберзбитків

	ESTIMATE	STD. ERROR	T VALUE	PR(>|T|)
(Intercept)	0.0108	0.0477	0.2270	0.8205
AI-based Detection	-0.3366	0.1324	-2.5424	0.0111**
event_time-3 (2017)	0.0590	0.065681	0.8949	0.3709
event_time-2 (2018)	-0.0710	0.0689	-1.0590	0.2897
event_time0 (2020)	-0.0124	0.0648	-0.1917	0.8480
event_time+1 (2021)	0.0152	0.0699	0.2171	0.8281
event_time+2 (2022)	0.0310	0.0668	0.4650	0.6420
event_time+3 (2023)	-0.0315	0.0686	-0.4597	0.6458
event_time+4 (2024)	0.0062	0.0678	0.0915	0.9271
AI × Year -3	0.1233	0.1680	0.7340	0.4630
AI × Year -2	0.2555	0.1775	1.4396	0.4630
AI × Year 0	0.5372	0.1667	3.2230	0.0012***
AI × Year +1	0.4372	0.1670	2.6189	0.0088***
AI × Year +2	0.2511	0.1706	1.4716	0.1413
AI × Year +3	0.3411	0.1682	2.0287	0.0426**
AI × Year +4	0.3425	0.1798	1.9054	0.05*
Observations	2,197			
R^2	0.011			
Adjusted R^2	0.004			
Residual Std. Error	0.716 (df = 2181)			
F Statistic	1.561* (df = 15; 2181)			

Джерело: побудовано авторами

Примітки: Рівень значущості (p-value) *p<0.1; **p<0.05; ***p<0.01

дослідженнях низький R^2 не є проблемою, оскільки фокус саме на каузальному ефекті коефіцієнтів взаємодії, а не на прогнозуванні.

Візуалізацію моделі (8) подано event-study графіком (рис. 2).

Графік дослідження подій (even-study plot, рис. 2), відображає динамічний ефект використання ШІ-технологій щодо визначення індексу кіберзбитків відносно базового 2020 року. Значення -3 відповідає 2017 року, а значення +4 – 2024 року. Верти-

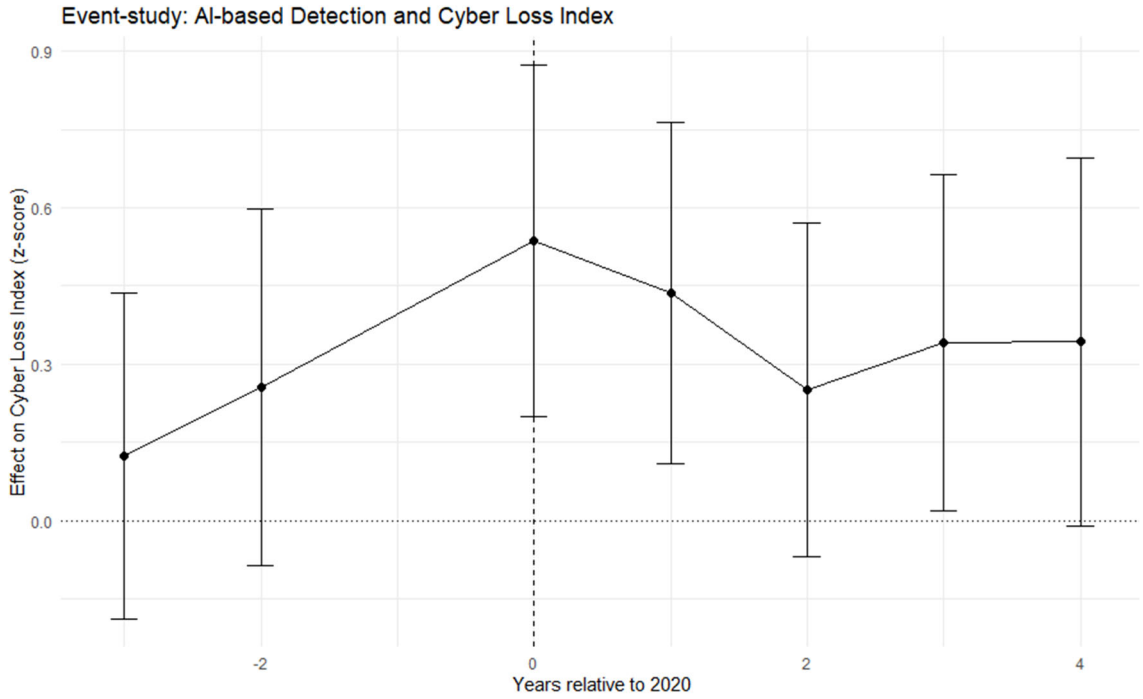

Рисунок 2 – Оцінки виявлення кіберзбитків з використанням ШІ-технологій

Джерело: побудовано авторами у програмному забезпеченні RStudio

кальні «вуса» відображають значення ефекту впливу технологій, що використовують штучний інтелект, на індекс кібезбитків (Cyber Loss Index) на 95% довірчих інтервалах (Borusyak et al., 2024). До 2020 року (event < 0) коефіцієнти для років -3, -2, -1 близькі до нуля, що підтверджує припущення про паралельні тенденції, а довірчі інтервали перетинають 0, тобто для цих років не спостерігається систематичного тренду. Отже, до впровадження ШІ-технологій не спостерігалось статистично значущих відмінностей між групами країн, які зазнали втручання (treated) та контрольною групами.

У момент впровадження AI-based Detection, 2020 рік (event = 0) помітний стрибок ефекту вгору, й довірчий інтервал не перетинає 0, тобто у рік впровадження AI-based Detection спостерігається статистично значуща зміна індексу кіберзбитків.

Після 2020 року (event > 0, тобто +1 – це 2021 рік, +4 – 2024 рік) оцінені ефекти зберігаються позитивними, однак у 2022 року є короткочасне послаблення, а далі – відновлення й стабілізація, що є цілком логічним та типовим патерном для AI-based Detection технологій, адже потрібен час на адаптацію й ефект є нелінійним. При цьому частина довірчих інтервалів повністю вище нуля або майже торкаються нуля (гранична значущість). Таким чином, після 2020 року ефект AI-based Detection не є одноразовим, а демонструє стійку динаміку у часі, а не тимчасовий шок.

Крім того, ключовим припущенням DiD-дизайну є паралельність трендів між treatment- та control-групами до втручання (Card et al, 1994).

Проведений F-тест на паралельність трендів до втручання підтвердив статистичну значущість event-study моделі, що грунтується на методології DiD-дизайну.

F- тест для періоду до втручання (до впровадження ШІ-технологій для виявлення кібершахрайств та значення кіберзбитків) перевіряє гіпотезу H_0, чи одночасно всі коефіцієнти до 2020 року дорівнюють нулю, тобто формально здійснює перевірку щодо припущення про паралельність трендів:

$$H_0 : \beta_{-3} = \beta_{-2} = 0 \quad (9)$$

Фрагмент коду для F-тесту такий:

linearHypothesis (event_model, c("treat:event_time-3 = 0", "treat:event_time-2 = 0"), vcov = vcovHC(event_model, type = "HC1")

Результати тесту зі значеннями F = 1.0458 та p-value = 0.3516 показили, що нульова гіпотеза (9) не може бути відхиленою, отже тренди до втручання є паралельними, тим самим, проведений івент-аналіз з використанням дизайну «різниці в різницях» є статистично значущим.

Висновки. У дослідженні проведено економетричний аналіз впливу використання технологій штучного інтелекту у системах виявлення кіберзагроз на рівень кіберзбитків у різних країнах. На відміну від значної частини існуючих робіт, які розглядають технології штучного інтелекту як однозначно ефективний інструмент підвищення кібербезпеки, у даному дослідженні здійснено кількісну оцінку їх фактичного впливу на результати протидії кіберзагрозам, зокрема з урахуванням випадків, коли атаки не були своєчасно виявлені або були пропущені системами захисту.

Методологічною основою дослідження став економетричний підхід Difference-in-Differences, доповнений динамічним event-study аналізом, що дозволило оцінити причинно-наслідковий ефект використання AI-based Detection на динаміку кіберзбитків. Емпіричний аналіз базувався на міжнародному наборі даних кіберінцидентів, який охоплює період 2017–2024 років. Для узагальнення інформації про масштаб кіберінцидентів було сформовано композитний індекс кіберзбитків (Cyber Loss Index), що інтегрує фінансові втрати від кібератак та кількість постраждалих користувачів. Логарифмування показників і їх подальша стандартизація за процедурою Z-score дозволили зменшити вплив екстремальних значень і забезпечити статистичну коректність економетричного моделювання.

Результати моделювання свідчать, що використання систем виявлення кіберзагроз на основі штучного інтелекту має статистично значущий вплив на динаміку кіберзбитків, однак цей вплив не є однозначно позитивним. Отримані результати демонструють, що навіть у країнах, де використовуються сучасні AI-технології кіберзахисту, частина кібератак залишається невиявленою або виявляється із запізненням. У таких випадках атаки встигають спричинити фінансові втрати та інші негативні наслідки, що відображається у зростанні показників кіберзбитків.

Динамічний аналіз показав, що після впровадження AI-based Detection у короткостроковому періоді спостерігається зростання зафіксованих кіберзбитків. Це пояснюється двома взаємопов'язаними факторами. По-перше, системи штучного інтелекту підвищують рівень виявлення кіберінцидентів, що призводить до більш повної фіксації атак у статистичних даних. По-друге, отримані результати свідчать, що наявність інтелектуальних систем захисту не гарантує повного блокування кіберзагроз, оскільки складність сучасних атак, використання нових технік обходу систем безпеки та обмеження алгоритмів детекції можуть призводити до пропуску частини інцидентів.

Водночас у середньостроковій перспективі результати event-study аналізу демонструють поступове зниження рівня кіберзбитків, що може бути пов'язано з адаптацією систем штучного інтелекту, накопиченням даних для навчання алгоритмів та вдосконаленням механізмів кіберзахисту. Це свідчить про те, що ефективність AI-систем у сфері кібербезпеки формується поступово і залежить від процесів їх інтеграції, навчання та вдосконалення.

Таким чином, результати дослідження підтверджують, що технології штучного інтелекту є важливим інструментом підвищення ефективності виявлення кіберзагроз, однак їх використання не усуває повністю ризик виникнення фінансових збитків від кібератак.

Навпаки, на початкових етапах впровадження таких систем може спостерігатися збільшення зафіксованих кіберінцидентів і збитків, що відображає як підвищення рівня їх виявлення, так і наявні обмеження алгоритмів детекції.

Практична значущість отриманих результатів полягає у тому, що вони підкреслюють необхідність комплексного підходу до забезпечення кібербезпеки, який поєднує використання технологій штучного інтелекту з організаційними, аналітичними та превентивними заходами захисту. Запропонований економетричний підхід та розроблений індекс кіберзбитків можуть бути використані для подальших досліджень ефективності технологій кіберзахисту, а також для оцінювання результативності інвестицій у системи штучного інтелекту у сфері кібербезпеки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Eurostat (2025). Digitalisation dashboard. URL: https://ec.europa.eu/eurostat/cache/dashboard/digitalisation/
2. Alblehai, F. (2025). Artificial intelligence-driven cybersecurity system for internet of things using self-attention deep learning and metaheuristic algorithms. *Scientific Reports*, Vol. 15. DOI: https://doi.org/10.1038/s41598-025-98056-2
3. Ramirez, S. (2025). AI Cyber Attacks Statistics 2026: How Attacks, Deepfakes & Ransomware Have Escalated. *SQ Magazine*, URL: https://sqmagazine.co.uk/ai-cyber-attacks-statistics/
4. Gitlab (2024). 2024 Global DevSecOps Report. URL: https://about.gitlab.com/developer-survey/
5. Muehlemann, Samuel (2025). Artificial intelligence adoption and workplace training. *Journal of Economic Behavior & Organization*. Vol. 238, Article No 107206. DOI: https://doi.org/10.1016/j.jebo.2025.107206
6. Salvatore Polizzi & Enzo Scannella (2023). Corporate environmental disclosure in Europe: the effects of the regulatory environment. *Journal of Financial Reporting and Accounting.* 23 (6). pp. 2345–2368. DOI: https://doi.org/10.1108/JFRA-03-2023-0165
7. Callaway, Brantly & Sant'Anna, Pedro H.C., 2021. "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, Elsevier, vol. 225(2), pp. 200-230. DOI: http://doi.org/10.1016/j.jeconom.2020.12.001
8. DeepDataLake (2025). Global Cybersecurity Threats (2015-2024). URL: https://deepdatalake.com/details.php
9. Oxford Insights (2020). Government AI Readiness Index 2020. URL: https://oxfordinsights.com/wp-content/uploads/2023/11/AIReadinessReport.pdf

MODELING THE IMPACT OF CYBER THREAT DETECTION SYSTEMS ON THE LEVEL OF CYBER DAMAGE: DID ANALYSIS AND EVENT INVESTIGATION APPROACH

Vitaliia V. Koibichuk
Ph.D. (Economics), Associate Professor,
Head of the Department of Economic Cybernetics,
Sumy State University

Alina Yu. Yefimenko
Ph.D. (Economics),
Assistant of the Department of Economic Cybernetics,
Sumy State University

The rapid spread of artificial intelligence technologies in the field of cybersecurity is accompanied by a paradoxical effect: AI-based Detection systems increase the efficiency of detecting cyber incidents, but do not guarantee complete prevention of financial losses from cyber attacks. In this regard, a methodological problem arises of correctly interpreting the dynamics of cyber damage after the implementation of intelligent protection systems. The purpose of the study is to econometrically assess the causal impact of the use of artificial intelligence-based cyber threat detection systems on the level of cyber damage. The empirical analysis is based on a data set covering cyber incidents in nine countries around the world (Australia, Brazil, China, France, Germany, India, Japan, the United Kingdom, the United States of America) for the period 2017–2024. The paper proposes an author's cyber damage index, formed on the basis of the integration of financial losses from cyber attacks and the number of affected users. To increase the statistical stability of the distributions, logarithmization of indicators and standardization by Z-score were applied. To assess the causal effect, the Difference-in-Differences econometric approach was used, as well as its dynamic extension in the form of event-study analysis, which allows us to study the time trajectory of the impact of the implementation of AI-based Detection systems. The results obtained indicate that the use of artificial intelligence systems has a statistically significant impact on the dynamics of cyber damage. In the short term after the implementation of AI-based Detection, an increase in the cyber damage index is observed, which is explained by an increase in the level of detection and fixation of cyber incidents that could have previously gone unnoticed. At the same time, in the medium term (3–4 years after implementation), a stable trend towards a decrease in cyber damage is formed, which reflects a gradual increase in the efficiency of AI-systems, their adaptation and integration into the cyber defense infrastructure. The results obtained demonstrate that the implementation of artificial intelligence technologies in the field of cybersecurity changes not only the level of protection, but also the mechanism for measuring cyber risks. The proposed approach allows for a more correct assessment of the effectiveness of AI-based cyber protection systems and is recommended for the analysis of digital security policies and cyber risk management.

Keywords: *difference-in-differences analysis, event research, parallel trends, artificial intelligence technologies, cyber damage index.*

JEL Classification: *C21, C23, C58, O33, G32*

Дата надходження статті: 10.04.2025
Дата прийняття статті: 07.05.2025
Дата публікації статті: 30.05.2025