https://doi.org/10.31891/2307-5732-2025-359-7
УДК 637.5.02

НАДТОЧИЙ МАКСИМ
Харківський національний університет радіоелектроніки
https://orcid.org/0009-0001-0530-9985
e-mail: maksym.nadtochyi@nure.ua
БАЛАГУРА ДМИТРО
Харківський національний університет радіоелектроніки
https://orcid.org/0009-0006-9839-3317
e-mail: dmytro.balagura@nure.ua
ШУЛІК ПАВЛО
Харківський національний університет радіоелектроніки
https://orcid.org/0009-0004-6200-2172
e-mail: pavlo.shulik@nure.ua
ПРОСОЛОВ ВЛАДИСЛАВ
Харківський національний університет радіоелектроніки
https://orcid.org/0009-0002-1276-4828
e-mail: vladyslav.prosolov@nure.ua

ПРОБЛЕМА НЕДОСТАТНЬОЇ АДАПТИВНОСТІ ТРАДИЦІЙНИХ СИСТЕМ МЕРЕЖЕВОЇ БЕЗПЕКИ ДО СУЧАСНИХ КІБЕРЗАГРОЗ: АНАЛІЗ ТА ШЛЯХИ ВИРІШЕННЯ

Традиційні системи мережевої безпеки, що базуються на статичних правилах та сигнатурах, виявляються неефективними проти сучасних адаптивних кіберзагроз, зокрема Advanced Persistent Threats (APT), zero-day експлойтів та AI-генерованих атак. В роботі проведено систематичний аналіз відкритих наукових публікацій 2020-2025 років з метою дослідження проблеми недостатньої адаптивності існуючих систем безпеки та аналізу можливостей застосування штучного інтелекту (AI) та великих мовних моделей (LLM) для її вирішення. Дослідження підтверджують суттєве підвищення швидкості виявлення загроз при використанні AI-рішень у порівнянні з класичними підходами. Великі мовні моделі продемонстрували потенціал у галузі аналізу розвідки загроз та організації безпеки, однак постають питання їх стійкості до атак та інтерпретованості рішень.

Ключові слова: *мережева безпека, адаптивні системи, штучний інтелект, великі мовні моделі, машинне навчання, системи виявлення вторгнень, пояснюваний штучний інтелект.*

NADTOCHYI MAKSYM, BALAGURA DMYTRO, SHULIK PAVLO, PROSOLOV VLADYSLAV
Kharkiv National University of Radio Electronics

THE PROBLEM OF INSUFFICIENT ADAPTABILITY OF TRADITIONAL NETWORK SECURITY SYSTEMS TO MODERN CYBER THREATS: ANALYSIS AND SOLUTIONS

Traditional network security systems based on static rules and signatures prove ineffective against modern adaptive cyber threats, particularly Advanced Persistent Threats (APTs), zero-day exploits, and AI-generated attacks. Contemporary digital transformation is accompanied by exponential growth of sophisticated cyberattacks characterized by increased complexity, autonomy, and adaptability that fundamentally challenge conventional security paradigms. This study conducts a systematic analysis of open scientific publications from 2020-2025 to investigate the problem of insufficient adaptability in existing security systems and analyse the potential applications of artificial intelligence (AI) and large language models (LLMs) for addressing this challenge.

The research methodology encompasses comprehensive evaluation of various AI-based approaches including ensemble methods, hybrid CNN-GRU architectures, Transformer-based models, and specialized large language models adapted for cybersecurity contexts. Particular attention is given to comparative analysis of effectiveness metrics across standardized datasets including UNSW-NB15, NSL-KDD, CIC-IDS2017, and CSE-CIC-IDS2018.

The research confirms a significant improvement in threat detection speed and accuracy when using AI-based solutions compared to classical approaches, with accuracy rates reaching 87-99.68% while maintaining controlled false positive rates. Ensemble approaches and genetically optimized hybrid CNN-GRU architectures demonstrate exceptional capability for dynamic adaptation to new attack types without manual intervention. Large language models have demonstrated substantial potential in threat intelligence analysis, automated security orchestration, and proactive threat prediction, with specialized models like CyBERT achieving 94.4% accuracy in security claim classification.

However, critical questions arise regarding their resilience to adversarial attacks, interpretability of decision-making processes, and computational overhead in real-world deployment scenarios. The study identifies key challenges including catastrophic forgetting, the need for explainable AI mechanisms, and requirements for continuous learning capabilities to address rapidly evolving threat landscapes. The findings emphasize the transition from reactive to proactive security paradigms through implementation of adaptive, self-learning systems capable of real-time threat prediction.

Keywords: *network security, adaptive systems, artificial intelligence, large language models, machine learning, intrusion detection systems, explainable artificial intelligence.*

Стаття надійшла до редакції / Received 03.10.2025
Прийнята до друку / Accepted 15.11.2025

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Сучасна цифрова трансформація суспільства супроводжується експоненціальним зростанням кіберзагроз, що характеризуються підвищенням складності, автономності та адаптивності атак. За даними звіту

WEF 2025 [10], кібератаки з використанням штучного інтелекту становлять одну з найсерйозніших загроз глобальній безпеці. Традиційні системи мережевої безпеки, базовані на статичних правилах та сигнатурах, виявляються неефективними проти Advanced Persistent Threats (APT), zero-day експлойтів та AI-генерованих атак.

Основна проблема полягає у фундаментальній неспроможності традиційних систем адаптуватися до динамічної природи сучасних кіберзагроз. Правило-орієнтовані системи безпеки, розроблені для виявлення відомих патернів атак, не здатні ефективно протистояти новим типам загроз, що еволюціонують швидше, ніж оновлюються сигнатурні бази [19].

Ця проблема безпосередньо пов'язана з критично важливими науковими задачами розвитку адаптивних систем кібербезпеки та практичними викликами захисту критичної інфраструктури. Відсутність ефективних AI-базованих рішень для виявлення та нейтралізації сучасних загроз створює системний ризик для економічної та національної безпеки.

Аналіз досліджень та публікацій

Останні наукові праці у сфері мережевої безпеки засвідчують активний розвиток систем виявлення вторгнень (IDS) на основі штучного інтелекту та глибинного навчання. Зокрема, дослідження Alshammari та Alsaleh [1] зосереджене на застосуванні інтелектуальних моделей для виявлення невідомих атак у транспортних системах, що підкреслює важливість використання гнучких алгоритмів у динамічних середовищах. У роботах Imrana та співавт. [2] та Zhang і Wang [3] обґрунтовано доцільність застосування гібридних архітектур CNN-GRU та GA-CNN, які забезпечують високу ефективність завдяки багаторівневому вилученню ознак.

Особливу увагу приділено ансамблевим методам. Так, Emanet та колеги [4] запропонували модель VEL-IDS на основі правила голосування, що демонструє покращену точність у складних сценаріях. Mohale та Obagbuwa [5] наголошують на важливості використання пояснюваного ШІ для підвищення прозорості, що є критичним для практичного застосування IDS у корпоративних мережах. Поряд із цим Chavan і Alone [6] показали високу результативність алгоритмів Random Forest на класичних наборах даних, зокрема CIC-IDS 2017.

Новим напрямом стали архітектури на основі трансформерів. Дослідження Long та ін. [7] підтверджує ефективність цього підходу для захисту хмарних обчислювальних середовищ. Водночас робота Mezzi та співавт. [8] акцентує на обмеженнях великих мовних моделей у контексті кіберрозвідки, вказуючи на ризики ненадійності та генерації помилкових даних. Нарешті, Chen із колегами [9] розробили багаторівневу модель VAE-GRU-XGBoost, яка поєднує методи глибинного навчання та бустингових алгоритмів, забезпечуючи підвищену точність класифікації атак.

Таким чином, аналіз досліджень демонструє поступовий перехід від традиційних статистичних методів до багаторівневих та адаптивних архітектур, що інтегрують штучний інтелект, пояснювані моделі та великі мовні моделі, але також виявляє виклики щодо їх надійності й інтерпретованості [1–9].

Формулювання цілей статті

Метою роботи є: аналіз проблеми недостатньої адаптивності традиційних систем мережевої безпеки до сучасних кіберзагроз, ідентифікація ключових обмежень систем, що базуються на правилах та сигнатурах, у контексті боротьби з динамічними та еволюціонуючими типами атак, а також окреслення шляхів вирішення цієї проблеми через впровадження адаптивних технологій штучного інтелекту та машинного навчання.

Виклад основного матеріалу

Природа проблеми недостатньої адаптивності традиційних систем

За останнє десятиліття мережна безпека еволюціонувала від стратегій захисту периметру до багатошарових архітектур, здатних протидіяти багатостадійним і високотехнологічним загрозам у мережевому середовищі. Традиційні засоби на основі сигнатур і статичних правил дедалі частіше виявляються недостатніми проти сучасних атак, зокрема APT, zero-day та AI-генерованих шкідливих матеріалів. Паралельно складність інфраструктур зростає через поширення хмарних сервісів, IoT та мобільного підключення, що створює нові площини атак, не повністю покриті класичними механізмами захисту.

Дослідження [20] підтверджує, що системи на основі сигнатур і статичних правил не здатні адаптуватися до нових патернів атак без ручного втручання експертів з безпеки. Це створює критичний розрив у часі між появою нової загрози та її включенням до систем захисту, що може становити від кількох днів до кількох місяців.

У дослідженні Security Operations Centers (SOC) [21] виявлені ключові проблеми signature-based систем виявлення вторгнень:

- низька специфічність: більшість правил генерують значну кількість хибнопозитивних спрацьовувань, що призводить до "втоми від сповіщень" серед аналітиків;
- обмежене покриття: значна частина правил ніколи не спрацьовує, а невелика частка відповідає за більшість сповіщень
- неможливість виявлення zero-day атак: системи можуть виявляти лише ті загрози, для яких попередньо створені сигнатури.

На цьому тлі прогрес у штучному інтелекті та машинному навчанні відкрив можливості для інтелектуального виявлення загроз, автоматизованого реагування й прогнозної аналітики, прискоривши перехід від статичних систем до адаптивних платформ.

Глибинне навчання (зокрема CNN, LSTM, GNN) довело ефективність у розпізнаванні складних векторів атак і аномалій, що не ідентифікуються сигнатурними методами, тоді як ансамблеві підходи

забезпечують високу точність і зниження хибнопозитивних спрацьовувань у гетерогенних середовищах. Великі мовні моделі (LLM) відкрили шлях до автоматизації аналізу розвідданих, екстракції індикаторів компрометації, генерації політик безпеки та покращеної людино-машинної взаємодії у SOC. Впровадження AI у практичні системи виявлення супроводжується зафіксованим підвищенням точності і помітним скороченням хибнопозитивних спрацьовувань порівняно з конвенційними засобами; інтеграція алгоритмів ML із моніторингом у реальному часі прискорює реагування організацій на інциденти.

Водночас зростає клас нових ризиків: атаки на самі AI-моделі, зловживання генеративними інструментами для масштабованого фішингу та соціальної інженерії, вразливості до маніпулятивних впливів на LLM, а також питання пояснюваності, стійкості та обчислювальної вартості. Це висуває вимоги до розвитку пояснюваного штучного інтелекту (XAI), підвищення стійкості систем та відповідального впровадження AI/LLM у процеси виявлення й реагування.

Таким чином, інтеграція AI та LLM у мережеву безпеку знаменує перехід від правило-орієнтованих систем до інтелектуальних, адаптивних і самонавчальних платформ, що поєднують поведінкову аналітику, виявлення аномалій і автоматизоване реагування. У цій роботі застосовано систематичний аналіз відкритих публікацій за 2020–2025 рр., узагальнено поточні можливості AI/LLM у мережевій безпеці й окреслено ключові виклики та напрямки подальшого розвитку.

Сучасні проблеми мережевої безпеки

Незважаючи на швидкий розвиток методик, технологій та механізмів захисту інформації у комп'ютерних системах, на сьогодні існує ряд проблем мережевої безпеки, як постійних, так і притаманних конкретному періоду. Розглянемо основні з них.

Еволюція кіберзагроз. Сучасний ландшафт кіберзагроз характеризується зростанням складності атак, використанням технік уникнення виявлення та широким застосуванням шифрування [1]. Advanced Persistent Threats (APT) та zero-day експлойти стали основними викликами для традиційних систем безпеки, що змушує організації переходити до більш інтелектуальних рішень.

Згідно з дослідженнями на наборі даних AWID, який містить різноманітні типи IoV-атак, навіть удосконалені системи виявлення досягають максимум 94% точності при збереженні прийнятного рівня хибних спрацьовувань у 3% [1]. Це підкреслює необхідність подальшого розвитку методів машинного навчання для кіберзахисту.

Вплив шифрування та IoT. Масове впровадження HTTPS, TLS 1.3 та інших протоколів шифрування обмежує можливості глибокого аналізу пакетів (DPI), примушуючи системи безпеки покладатися на метадані та поведінкові індикатори [2]. Водночас експоненціальне зростання IoT-пристроїв створює нові вектори атак і вимагає спеціалізованих підходів до виявлення аномалій.

Експерименти з наборами даних UNSW-NB15 та NSL-KDD показують, що гібридні CNN-GRU-FF моделі здатні досягати коефіцієнтів виявлення 98.22% та 99.68% відповідно, що робить їх перспективними для захисту IoT-середовищ [2].

Проблема хибнопозитивних спрацьовувань. Однією з найкритичніших проблем реальних IDS є високий рівень false positives, що призводить до "втоми від сповіщень" та зниження ефективності SOC-аналітиків [5]. Дослідження показують, що для промислового впровадження необхідний FPR нижче 0.1%, що досягається лише найбільш сучасними ансамблевими методами [3][4].

Таким чином, сучасні виклики мережевої безпеки чітко демонструють обмеженість традиційних підходів і підкреслюють потребу у впровадженні нових інтелектуальних рішень, здатних гнучко адаптуватися до мінливого ландшафту загроз.

Методи AI та LLM у мережевій безпеці

Саме на цьому етапі дедалі більшої актуальності набувають технології штучного інтелекту. Вони дозволяють не лише підвищувати точність виявлення атак, але й знижувати рівень хибних спрацьовувань та ефективніше аналізувати великі обсяги мережевого трафіку. Особливо перспективними вважаються ансамблеві моделі, гібридні нейромережеві архітектури та сучасні підходи на основі LLM, які суттєво розширюють можливості автоматизації у сфері кіберзахисту.

Ансамблеві підходи. Ансамблеве навчання демонструє значні переваги порівняно з окремими класифікаторами завдяки здатності комбінувати переваги різних алгоритмів [4][6]. Дослідження VEL-IDS показало, що поєднання Decision Tree, Extra Trees та Logistic Regression з ваговим голосуванням досягає точності 98.82% на наборі CSE-CIC-IDS2018 при скороченні часу навчання з 24 до 6.42 хвилин [4].

Експерименти з Random Forest на наборі CIC-IDS2017 підтверджують ефективність ансамблевих методів, демонструючи точність 97.20%, precision 96.80%, recall 96.50% та F1-score 96.70% з AUC 0.97 [6]. Ці результати значно перевищують показники окремих класифікаторів, таких як SVM (93.50%) або Decision Tree (91.80%).

Гібридні CNN/GRU архітектури. Комбінування Convolutional Neural Networks (CNN) для просторового аналізу та Gated Recurrent Units (GRU) для часових залежностей показує виняткові результати [2][3]. Модель CNN-GRU-FF з подвійним шаром виділення ознак та модифікованою focal loss функцією досягає коефіцієнтів виявлення 98.22% на UNSW-NB15 та 99.68% на NSL-KDD при низьких рівнях хибних спрацьовувань [2].

Оптимізація генетичними алгоритмами (GA) додатково покращує ефективність CNN-GRU мереж. Дослідження на промислових системах управління показало, що GA-CNN-GRU досягає коефіцієнта виявлення

96.97% при надзвичайно низькому FAR 0.0947% та часі навчання лише 8 секунд [3].

Transformer-архітектури. Застосування Transformer-моделей у мережевій безпеці відкриває нові можливості для аналізу послідовностей мережевого трафіку [7]. Transformer-based підходи досягають точності понад 93% на наборі CIC-IDS2018, що є порівняним з CNN-LSTM моделями при значно кращій здатності до захоплення глобальних залежностей [7].

Багатомасштабні Transformer-архітектури з Cross Feature Enrichment (CFE) демонструють подальше покращення результатів завдяки ефективному поєднанню ознак різних масштабів та використанню attention-механізмів для виявлення потенційних слідів вторгнень.

Великі мовні моделі в CTI. LLM показують перспективи у сфері Cyber Threat Intelligence, але з обмеженнями [8]. Дослідження CTIBench виявило, що попри заявлені високі показники точності (до 89% у деяких роботах), LLM демонструють проблеми з узгодженістю та надмірною впевненістю при аналізі реальних звітів розвідки загроз [8].

Основні застосування LLM у CTI включають:

- автоматичну анотацію та класифікацію звітів загроз;
- виділення індикаторів компрометації (IoC);
- генерацію структурованих даних з неструктурованих джерел;
- підтримку SOC-аналітиків у пріоритизації та початковому опрацюванні інцидентів;

Отже, сучасні AI та LLM-рішення відкривають широкі можливості для кіберзахисту, але їхня практична цінність має підтверджуватись конкретними експериментами, показниками точності, швидкодії та узагальнювальної здатності. Саме ці аспекти і стають предметом порівняльного аналізу.

Порівняльний аналіз методів

У цьому розділі здійснюється систематичне зіставлення результатів різних підходів, що дозволяє не лише оцінити рівень їхньої ефективності, але й виявити сильні та слабкі сторони кожного з них у контексті реальних умов використання. Особливу увагу буде приділено відмінностям між наборами даних, часовими характеристиками та здатністю моделей до узагальнення, що є критичними факторами для промислового впровадження. Також буде проведена оцінка вже існуючих рішень.

Різні набори даних демонструють відмінності у складності та репрезентативності сучасних загроз. UNSW-NB15, як один з найбільш реалістичних наборів, показує нижчі абсолютні значення точності порівняно з синтетичними наборами, але кращі результати узагальнення [2, 5].

Дослідження з XGBoost та CatBoost на UNSW-NB15 досягли точності 87% при FPR 0.07 та FNR 0.12, що підкреслює важливість вибору відповідного набору даних для оцінки [5]. Водночас експерименти з ensemble методами на більш збалансованих наборах показують значно вищі результати.

Таблиця 1

Основні категорії загроз та AI-рішення

Категорія загроз	AI-рішення	Результати	Набір даних
Мережеві вторгнення	ACIDS (CNN + SMOTE + Open-set)	Accuracy: 94%, DR: 100%, FAR: 3% [1]	AWID
Аномалії трафіку IoT	GA-CNN-GRU оптимізація	DR: 96.97%, FAR: 0.0947% [3]	CICIDS2017
Багатокласова детекція	CNN-GRU-FF з focal loss	DR: 98.22% (UNSW), 99.68% (NSL-KDD) [2]	UNSW-NB15, NSL-KDD
Промислові мережі	Ensemble VEL-IDS (voting)	Accuracy: 98.82%, час: 6.42 хв [4]	CSE-CIC-IDS2018
Хмарні середовища	Transformer + attention	Accuracy: >93%, порівняно з CNN-LSTM [7]	CIC-IDS2018

Таблиця 2

Порівняльна ефективність методів ШІ

Метод	Набір даних	Accuracy (%)	FPR/FAR (%)	Додаткові метрики	Джерело
Random Forest (ensemble)	CIC-IDS2017	97.20	-	Precision: 96.80%, AUC: 0.97 [6]	IJCA
GA-CNN-GRU	CICIDS2017	-	0.0947	DR: 96.97%, час навчання: 8с [3]	PLOS One
CNN-GRU-FF	NSL-KDD	-	Низький	DR: 99.68% [2]	Springer
CNN-GRU-FF	UNSW-NB15	-	Низький	DR: 98.22% [2]	Springer

Метод	Набір даних	Accuracy (%)	FPR/FAR (%)	Додаткові метрики	Джерело
VEL-IDS (DT+ET+LR)	CSE-CIC-IDS2018	98.82	1.18	F1: 97.65%, час: 6.42 хв [4]	PeerJ
XGBoost/CatBoost + XAI	UNSW-NB15	87.00	7.00	FNR: 12% [5]	Frontiers
VAE-GRU-XGBoost	KDD99	97.58	-	AUC: 97.48%, інференція: 0.003с [9]	PLOS One
ACIDS (CNN+SMOTE)	AWID	94.00	3.00	DR: 100% для IoV атак [1]	Nature

Часові характеристики. Швидкість навчання та інференції є критичними факторами для промислового впровадження. GA-оптимізовані CNN-GRU моделі демонструють час навчання 8 секунд порівняно з 10+ секундами для інших методів [3]. VAE-GRU-XGBoost показує час інференції 0.003 секунди при точності 97.58% [9].

Узагальнювальна здатність. Тестування моделей на різних наборах даних виявляє проблеми з узагальненням. CNN-GRU-FF модель показала незначне зниження точності при переході з NSL-KDD (99.68%) на UNSW-NB15 (98.22%), що свідчить про хорошу адаптивність [2].

Архітектурні основи та технологічні інновації LLM

Революція у сфері обробки природної мови, спричинена впровадженням трансформерних архітектур, знаходить своє продовження в галузі кібербезпеки. Механізм багатоголової уваги (multi-head self-attention), що лежить в основі цих архітектур, виявляється особливо ефективним для аналізу мережевого трафіку, де контекстуальні залежності між окремими пакетами можуть розтягуватися на значні часові інтервали [11]. На відміну від традиційних підходів, які аналізують мережеві події ізольовано, механізм self-attention дозволяє моделі одночасно враховувати важливість усіх частин вхідного простору, створюючи цілісне розуміння мережевої активності.

Адаптація мовних моделей до специфічних вимог мережевої безпеки потребувала кардинального переосмислення процесу токенізації. Традиційні NLP токенізатори, розроблені для обробки людської мови, виявилися неефективними при роботі з мережевими даними, що мають принципово іншу структуру та семантику. Відповідь на цей виклик прийшла у вигляді протокол-усвідомленої токенізації, яка враховує унікальні характеристики різних мережевих протоколів.

Модель ET-BERT [12] демонструє один із найбільш інноваційних підходів до цієї проблеми, розробляючи спеціальну схему токенізації для зашифрованого трафіку. Замість простого розбиття на байти, система використовує трирівневу ієрархію: байт-рівневе токенізування для детального аналізу, структурну сегментацію для розділення заголовків та корисного навантаження відповідно до специфіки протоколу, та темпоральне групування для об'єднання послідовних пакетів у логічні сесії на основі часових характеристик.

Ще більш досконалою є ієрархічна токенізація, впроваджена в рамках моделі netFound [13], де токени нижчого рівня представляють конкретні поля пакетів, а токени вищого рівня відображають агреговані характеристики цілих потоків. Такий підхід дозволяє моделі одночасно аналізувати як мікроскопічні деталі окремих пакетів, так і макроскопічні патерни поведінки цілих мережевих сесій.

Однією з найбільш складних проблем у сфері кібербезпеки є постійна еволюція тактик, технік та процедур зловмисників, що призводить до явища дрейфу даних. Традиційні системи машинного навчання страждають від катастрофічного забування - неспроможності зберігати знання про раніше вивчені типи атак при навчанні на нових даних. Системи на основі LLM пропонують елегантне рішення цієї проблеми через впровадження механізмів безперервного навчання [14].

Особливо ефективним виявилося використання гауссових сумішей моделей для кластеризації нових типів атак. Цей підхід дозволяє системі автоматично виявляти кластери невідомого трафіку та інтегрувати їх у існуючу модель знань без втрати інформації про раніше вивчені класи. Експериментальні дослідження демонструють, що використання восьми-дванадцяти компонентів забезпечує оптимальний баланс між точністю виявлення нових атак на рівні 95,6% та обчислювальною ефективністю [14].

В роботах [15, 16] розглянуті спеціалізовані моделі на базі архітектури BERT. На відміну від загальних мовних моделей, CyBERT був цілеспрямовано адаптований для роботи в специфічному контексті кібербезпеки, що включало розширення словника понад п'ятдесятьма тисячами технічних термінів, специфічних для цієї галузі.

Архітектура CyBERT передбачає двоетапне предтренування: спочатку модель навчається на загальних текстових корпусах для формування базового розуміння мови, а потім проходить додаткове тренування на спеціалізованих текстах з кібербезпеки. Модифікований механізм маскування враховує структурні особливості технічних документів, що дозволяє моделі краще розуміти контекст термінів та понять, специфічних для галузі

безпеки.

Результати впровадження CyBERT є вражаючими: точність класифікації заяв про безпеку зросла з 76% для стандартного BERT до 94,4% для доменно-адаптованої версії. Це покращення демонструє критичну важливість спеціалізації мовних моделей для конкретних галузей застосування [15].

Генеративні моделі та їх застосування

Розробка Hackphyr [17] представляє принципово новий підхід до автоматизації процедур тестування на проникнення. Ця модель з 7 мільярдами параметрів, побудована на основі GPT архітектури, була спеціально налаштована для проведення red-team операцій, що включають симуляцію реальних атак для оцінки стійкості систем безпеки.

Унікальність Hackphyr полягає в поєднанні декодер архітектури з спеціалізованим набором даних, що містить понад п'ятсот тисяч сценаріїв тестувань на проникнення. Модель була додатково покращена через застосування Reinforcement Learning from Human Feedback, що дозволило підвищити якість та релевантність її рекомендацій. Порівняльні тести показують, що Hackphyr перевершує GPT-3.5-turbo на 23% у складних невідомих сценаріях та досягає продуктивності, порівнянної з GPT-4, при роботі на одній графічній карті [17].

Особливо цінною є здатність моделі генерувати не лише технічні рекомендації, але й пояснення логіки атак, що робить її потужним інструментом для навчання та розвитку команд кібербезпеки. Локальне розгортання забезпечує конфіденційність операцій та дозволяє організаціям проводити тестування без ризику витоку чутливої інформації.

BARTPredict [18] представляє революційний підхід до проактивної кібербезпеки через впровадження двокомпонентної архітектури. Перший компонент, побудований на основі BART моделі, аналізує історичні дані мережевого трафіку для генерування прогнозів майбутньої активності. Другий компонент, що використовує BERT архітектуру, оцінює передбачений трафік на предмет потенційної зловмисності. Цей підхід дозволяє системі не просто реагувати на атаки після їх початку, але й передбачати потенційні загрози на основі попередніх індикаторів. Архітектура оптимізує комбіновану функцію втрат, що балансує точність передбачення майбутнього трафіку з якістю його класифікації, забезпечуючи високу ефективність обох компонентів системи. Експериментальні результати на наборі даних CICIoT2023 демонструють вражаючу загальну точність на рівні 98%, що підкреслює потенціал генеративних моделей для створення проактивних систем захисту. Особливо важливою є здатність системи адаптуватися до нових патернів атак через безперервне навчання на оновлених даних.

Висновки з даного дослідження
і перспективи подальших розвідок у даному напрямі

Проведений аналіз джерел демонструє критичну проблему недостатньої адаптивності традиційних систем мережевої безпеки до швидко еволюціонуючих кіберзагроз. Дослідження підтверджують фундаментальні обмеження правило-орієнтованих та signature-based систем у контексті сучасного ландшафту кіберзагроз.

Таким чином визначені ключові рішення проблеми адаптивності.

1. Впровадження адаптивних AI-систем безпеки.

Сучасні AI-методи досягають високих показників точності (87-99.68%) при контрольованих коефіцієнтах хибнопозитивних спрацьовувань (0.0947-7%). Особливо ефективними виявилися ансамблеві підходи та оптимізовані GA гібридні CNN-GRU архітектури, які здатні динамічно адаптуватися до нових типів атак без ручного втручання.

2. Розвиток систем безперервного навчання.

Впровадження LLM-базованих систем з механізмами безперервного навчання дозволяє вирішити проблему катастрофічного забування традиційних систем. Дослідження показують досягнення 95% точності виявлення нових атак при збереженні знань про попередні типи загроз [22].

3. Створення гібридних архітектур.

Комбінування різних AI-підходів (CNN-GRU, Transformer-based архітектури, ансамблеві методи) забезпечує подолання обмежень статичних систем через здатність до аналізу як просторових, так і часових патернів атак.

4. Впровадження проактивних систем захисту.

Генеративні моделі типу BARTPredict демонструють можливість передбачення потенційних загроз з точністю 98%, що кардинально змінює парадигму від реактивного до проактивного захисту.

5. Розвиток пояснюваних AI-систем (XAI)

Інтеграція механізмів пояснюваності дозволяє вирішити проблему "чорної скрині" AI-систем, забезпечуючи прозорість рішень при збереженні високої точності виявлення.

Методи з доведеною ефективністю на стандартизованих наборах даних (CICIDS2017, UNSW-NB15, NSL-KDD) готові для промислового впровадження за умови належної перевірки у специфічних середовищах організацій. Рекомендується поетапний підхід до впровадження AI-технологій з акцентом на гібридні рішення, що поєднують переваги різних архітектур та забезпечують баланс між точністю, швидкодією та здатністю до інтерпретації.

Розвиток федеративного навчання, квантових алгоритмів та автономних систем захисту відкриває нові можливості для створення більш ефективних та масштабованих рішень кіберзахисту, здатних динамічно адаптуватися до мінливого ландшафту кіберзагроз.

Основний висновок: проблема недостатньої адаптивності традиційних систем мережевої безпеки може

бути ефективно вирішена через впровадження інтелектуальних адаптивних систем на основі штучного інтелекту, здатних до самонавчання, динамічної модифікації стратегій захисту та проактивного виявлення загроз у реальному часі.

Література

1. Alshammari, F., Alsaleh, A. Smart intrusion detection model to identify unknown attacks for improved road safety and management. Sci Rep 15, 18584 (2025). https://doi.org/10.1038/s41598-025-03604-5
2. Imrana, Y., Xiang, Y., Ali, L. et al. CNN-GRU-FF: a double-layer feature fusion-based network intrusion detection system using convolutional neural network and gated recurrent units. Complex Intell. Syst. 10, 3353–3370 (2024). https://doi.org/10.1007/s40747-023-01313-y
3. Zhang J., Wang X., ICN intrusion detection method based on GA-CNN. PLOS One. 2025. DOI: https://doi.org/10.1371/journal.pone.0325367
4. Emanet S, Karatas Baydogmus G, Demir O. 2023. An ensemble learning based IDS using Voting rule: VEL-IDS. PeerJ Computer Science 9:e1553 https://doi.org/10.7717/peerj-cs.1553
5. Mohale VZ and Obagbuwa IC (2025) Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability. Front. Comput. Sci. 7:1520741. https://doi.org/10.3389/fcomp.2025.1520741
6. Chavan P.V., Alone N.V. Optimizing Intrusion Detection with Random Forest: A High-Accuracy Approach using CIC-IDS 2017. International Journal of Computer Applications. 2025. Vol. 187, No. 3.
7. Long, Z., Yan, H., Shen, G. et al. A Transformer-based network intrusion detection approach for cloud security. J Cloud Comp 13, 5 (2024). https://doi.org/10.1186/s13677-023-00574-9
8. Mezzi E. et al. Large Language Models are unreliable for Cyber Threat Intelligence. arXiv preprint. 2025. arXiv:2503.23175.
9. Chen Y, Zheng X, Wang N (2025) Construction of VAE-GRU-XGBoost intrusion detection model for network security. PLoS One 20(6): e0326205. https://doi.org/10.1371/journal.pone.0326205
10. World Economic Forum. (2025). Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards. URL: https://reports.weforum.org/docs/WEF_Artificial_Intelligence_and_Cybersecurity_Balancing_Risks_and_Rewards_2025.pdf
11. Daly Manocchio, L., Layeghy, S., Weng Lo, W., Kulatilleke, G. K., Sarhan, M., and Portmann, M., FlowTransformer: A Transformer Framework for Flow-based Network Intrusion Detection Systems, arXiv e-prints, Art. no. arXiv:2304.14746, 2023. doi: https://doi.org/10.48550/arXiv.2304.14746
12. Lin, X., Xiong, G., Gou, G., Li, Z., Shi, J., and Yu, J., ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification, arXiv e-prints, Art. no. arXiv:2202.06335, 2022. doi: https://doi.org/10.48550/arXiv.2202.06335
13. Guthula, S., Beltiukov, R., Battula, N., Guo, W., Gupta, A., and Monga, I., netFound: Foundation Model for Network Security, arXiv e-prints, Art. no. arXiv:2310.17025, 2023. doi: https://doi.org/10.48550/arXiv.2310.17025
14. Adjewa, F., Esseghir, M., and Merghem-Boulahia, L., LLM-based Continuous Intrusion Detection Framework for Next-Gen Networks, arXiv e-prints, Art. no. arXiv:2411.03354, 2024. doi: https://doi.org/10.48550/arXiv.2411.03354
15. Ameri, K.; Hempel, M.; Sharif, H.; Lopez Jr., J.; Perumalla, K. CyBERT: Cybersecurity Claim Classification by Fine-Tuning the BERT Language Model. J. Cybersecur. Priv. 2021, 1, 615-637. https://doi.org/10.3390/jcp1040031
16. Bayer, M., Kuehn, P., Shanehsaz, R., and Reuter, C., CySecBERT: A Domain-Adapted Language Model for the Cybersecurity Domain, arXiv e-prints, Art. no. arXiv:2212.02974, 2022. https://doi.org/10.48550/arXiv.2212.02974
17. Rigaki, M., Catania, C., and Garcia, S., Hackphyr: A Local Fine-Tuned LLM Agent for Network Security Environments, arXiv e-prints, Art. no. arXiv:2409.11276, 2024. https://doi.org/10.48550/arXiv.2409.11276
18. A. Diaf, A. A. Korba, N. Elislem Karabadji and Y. Ghamri-Doudane, "BARTPredict: Empowering IoT Security with LLM-Driven Cyber Threat Prediction," *GLOBECOM 2024 - 2024 IEEE Global Communications Conference*, Cape Town, South Africa, 2024, pp. 1239-1244, https://doi.org/10.1109/GLOBECOM52923.2024.10901770
19. Uddin, M. A., Aryal, S., Bouadjenek, M. R., Al-Hawawreh, M., and Talukder, M. A., “usfAD based effective unknown attack detection focused IDS framework”, Scientific Reports, vol. 14, no. 1, Art. no. 29103, 2024. https://doi.org/10.1038/s41598-024-80021-0
20. Dave, D., Sawhney, G., Aggarwal, P., Silswal, N., and Khut, D., The New Frontier of Cybersecurity: Emerging Threats and Innovations, arXiv e-prints, Art. no. arXiv:2311.02630, 2023. doi: https://doi.org/10.48550/arXiv.2311.02630
21. Teuwen, K. T. W., Mulders, T., Zambon, E., and Allodi, L., Ruling the Unruly: Designing Effective, Low-Noise Network Intrusion Detection Rules for Security Operations Centers, arXiv e-prints, Art. no. arXiv:2501.09808, 2025. doi: https://doi.org/10.48550/arXiv.2501.09808
22. Koca, M., & Avcı, İ. (2024). Enhancing Network Security: A Comprehensive Analysis of Intrusion Detection Systems. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 29(3), 927-938. https://doi.org/10.53433/yyufbed.1545033