КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.29.941
УДК 004.896

Чернігівський Іван Андрійович
аспірант
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0009-0003-4568-3212
i.chernihivskiy@gmail.com

Крючкова Лариса Петрівна
доктор технічних наук, професор
професор кафедри інформаційної та кібернетичної безпеки
імені професора Володимира Бурячка
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0002-8509-6659
l.kriuchkova@kubg.edu.ua

ТЕСТУВАННЯ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ ВИЯВЛЕННЯ ЗАРАЖЕНИХ ПК НА БАЗІ ЦИФРОВИХ СЛІДІВ

Анотація. Розвиток штучного інтелекту досягнув великого прогресу і вже сьогодні має значний вплив на велику кількість галузей, а з розвитком LLM буде мати ще більший вплив у майбутньому, особливо на кібербезпеку. ШІ може як допомогти зберегти дані, шляхом раннього виявлення кібератак, так і зашкодити кібербезпеці полегшуючи написання переконливих фішингових листів, відтворюючи фрагменти шкідливого коду, допомагаючи виявляти слабкі місця у мережі, та знаходити ще невідомі виробникам програмного забезпечення вразливості в операційній системі, програмах тощо (zero day vulnerability). Тому, щоб не відставати у цій «гонці озброєнь», потрібно вже зараз впроваджувати ШІ у якості одного із компонентів кіберзахисту на підприємстві. Актуальність роботи полягає у необхідності знаходження таких моделей штучного інтелекту, які вже зараз можливо залучити до вирішення задач захисту інфокомунікаційних мереж. Метою статті є тестування нейромережевих моделей формату GGUF для оцінки можливості їх застосування при вирішенні задачі виявлення заражених ПК на базі цифрових слідів. У роботі розглянуто типи і технології штучного інтелекту та їх вплив на кібербезпеку і в якості захисту від кібератак, і в якості одного із компонентів для атак на інформаційну інфраструктуру. Щоб оцінити можливості застосування існуючих моделей ШІ для вирішення актуальних задач кіберзахисту, зокрема виявлення заражених ПК на базі цифрових слідів із застосуванням ШІ, визначено критерії для моделі ШІ які будуть прийнятними для використання у корпоративному середовищі та проведено тестування 135 моделей формату GGUF на предмет виявлення або невиявлення ними ознак вірусної активності та індикаторів компрометації у промпті, що надавався користувачем. Оскільки виявлено, що при запуску однієї і тієї ж нейромережевої моделі з однаковими промптами але різними програмами, що можуть запускати локальні моделі на ПК, її відповідь кардинально змінюється, підготовлено низку зведених таблиць де є назва моделі і варіанти відповідей під кожну програму для запуску ШІ моделей, без урахування тих, що надали неправильну відповідь, витратили надто багато часу на відповідь або завершилися з помилкою. Визначено перелік доцільних для використання моделей ШІ у форматі GGUF для вирішення задач кібербезпеки, зокрема для виявлення заражених ПК на базі цифрових слідів. Проте, так як кожна модель краще проявляє себе у специфічних умовах із різними сценаріями запуску, вибір моделі буде залежати від актуальних задач та наявних ресурсів. Подальші дослідження можуть бути зосереджені на вдосконаленні методики дослідження моделей для обробки цифрових слідів, перетворенні цифрових слідів з ПК у зрозумілий для ШІ промт та автоматизацію аналізу відповіді ШІ.

Ключові слова: штучний інтелект; нейромережеві моделі; LLM; комп’ютерні віруси; цифрові сліди; тестування; промпт; кібербезпека.

ВСТУП

Штучний інтелект (ШІ) в інформаційній безпеці використовується з 1980-х років, а останні досягнення зробили його набагато ефективнішим. ШІ трансформує кібербезпеку, включаючи безпеку даних, систему керування ідентичністю та доступом, керування ІТ, безпеку хмарних технологій, а також виявлення кіберзагроз і спрощення реагування на них спеціалістами із безпеки. Майбутні досягнення в галузі штучного інтелекту, що стрімко розвивається, стимулюватимуть розробку продуктів, які будуть покладатися на співпрацю між людьми та системами на основі ШІ [1].

Проте, ШІ може як допомогти зберегти дані, шляхом раннього виявлення кібератак, так і зашкодити кібербезпеці полегшуючи написання переконливих фішингових листів, відтворюючи фрагменти шкідливого коду, допомагаючи виявляти слабкі місця у мережі, та знаходити ще невідомі виробникам програмного забезпечення вразливості в операційній системі, програмах тощо (zero day vulnerability).

Наприклад, вже зараз кіберзлочинці впроваджують генеративні моделі штучного інтелекту для підвищення ефективності та масштабів добре відомих форм атаки, таких як програми-вимагачі та компрометація ділової електронної пошти (BEC) за рахунок переконливого відтворення стилів спілкування керівників компанії за допомогою тексту або відео чи аудіо (deepfake) та написання шкідливого коду із залученням ШІ. Інструменти GenAI знижують вартість фішингових кампаній з використанням соціальної інженерії, які надають зловмисникам початковий доступ до організацій. Ці інструменти зазвичай налаштовуються під конкретну компанію шляхом використання даних отриманих в результаті OSINT для донавчання моделі ШІ. Джерелом таких даних можуть слугувати соціальні мережі, публічні заяви чи документи, витоки з різних джерел (наприклад, служби таксі та доставки речей) та багато іншого, що робить спроби соціальної інженерії набагато складнішими для ідентифікації звичайними користувачами. GenAI також допомагає зловмисникам застосовувати соціальну інженерію ширшим колом мов, що розширює поверхню атаки на більшу кількість людей у більшій кількості країн із меншими витратами, що в свою чергу дозволяє здійснювати більш складні та масштабовані кібератаки [2]–[5].

Платформи «кіберзлочинність як послуга» (CaaS) продовжують бути домінуючою та швидкозростаючою бізнес-моделлю у кримінальному ландшафті, оскільки усувають бар’єри для входу в кіберзлочинну діяльність, дозволяючи окремим особам або групам без технічних знань займатися незаконною онлайн-діяльністю, купуючи необхідні інструменти та підтримку. Ця модель, яка вже добре зарекомендувала себе серед злочинних груп, поступово застосовується в інших сферах кіберзлочинності, таких як фішингові атаки, посилені штучним інтелектом. Оскільки персонал компаній є основною мішенню атак дипфейків, а також фішингових кампаній загалом, організаціям потрібно буде переглянути те, як вони навчають та захищають усіх, від співробітників до керівництва та ради директорів, від нових моделей кіберзлочинності [5].

Тому, організації, що приділяють додаткову увагу захисту від розроблених з технологією deepfake фішингових кампаній, надають роз’яснення персоналу, пріоритезують впровадження ШІ у своїй роботі та інвестують у захисні рішення із залученням ШІ, можуть суттєво зменшити або нівелювати вплив кіберзагроз на свою інфокомунікаційну систему.

Постановка проблеми. Розвиток штучного інтелекту сьогодні досягнув великого прогресу і має значний вплив на велику кількість галузей, а особливо на кібербезпеку.

Оскільки такі компанії, як OpenAI з ChatGPT та Meta з відкритим вихідним кодом Llama, зробили технологію ШІ доступнішою для кінцевого користувача, деякі зловмисники почали експлуатувати ці інновації, що призвело до здешевлення та підвищення витонченості кібератак. Зловживання цими технологіями є значним ризиком, оскільки тепер зловмисники можуть легко використовувати вразливості в системах, створюючи проблеми для команд безпеки [6].

Наприклад у [7] прогнозується, що з початку 2025 року набори інструментів для кібератак будуть адаптовані та автоматизовані за допомогою ШІ, у тому числі і сканери вразливостей, оскільки ШІ може знаходити вразливості про які ще не відомо компаніям-розробникам програмного забезпечення (zero day). А також, що новим вектором атак із залученням ШІ стануть самі ШІ моделі які використовуються в компанії, у тому числі і у захисних рішеннях [7].

Згідно з даними AV-TEST, у 2024 році загальна кількість вірусів вже перевалила за 1,5 млрд. екземплярів а кожну годину з'являється більше ніж 12 тис. вірусів [8]. Спроби використовувати ШІ для захисту від вірусів були ще раніше, наприклад, в [9] досліджується можливість класифікації вірусних файлів на сімейства, використовуючи представлення бінарних даних у вигляді картинки на основі згорткових нейронних мереж (CNN). Проте лише з появою великих мовних моделей (LLM) стало можливо інтегрувати ШІ моделі у безпеку.

Варто зазначити, що розподіленість інфокомунікаційних мереж збільшує час на виявлення та нейтралізацію загроз за рахунок збільшення кількості цифрових артефактів для дослідження. Узагальнену структуру інфокомунікаційної мережі наведено на рис.1. в публікації [10].

Тому, щоб не відставати у цій "гонці озброєнь", потрібно вже зараз впроваджувати ШІ у якості одного із компонентів кіберзахисту на підприємстві.

Зазвичай, дискусії щодо якості та ефективності моделей машинного навчання (ML) часто точаться навколо розміру моделі, кількості параметрів і продуктивності на встановлених тестових даних. Водночас до уваги не береться час — найважливіший результат, який може забезпечити якісний ШІ. У певних сферах, таких як обробка текстової інформації, завдання категоризації та ідентифікації об'єктів, оцінювати час не критично важливо. Однак у кібербезпеці час має вирішальне значення для виявлення загроз у контексті захисту від шкідливого програмного забезпечення до його виконання. Саме тут моделі виявляють і блокують шкідливе програмне забезпечення до того, як воно розгортається і виконується [11].

Аналіз останніх досліджень і публікацій. У працях науковців Hakan T. Otal, M. Abdullah Canbaz [12], Yazi Gholami [13], Luigi Coppolino та ін. [14], D. Ucci, L. Aniello, R. Baldoni [15] досліджено великі мовні моделі та їх можливе застосування в кібербезпеці, оскільки з їх стрімким розвитком, вони відіграватимуть дедалі важливішу роль у захисті інформаційних систем від нових кіберзагроз. Проте у цих працях не приводяться конкретні дані по моделям ШІ (назва і розмір моделі, час відповіді, релевантність відповіді), що доступні для завантаження користувачами і можуть залучатися до будування ешелонованого захисту інфокомунікаційної мережі на підприємстві.

Актуальність роботи полягає у необхідності знаходження таких моделей штучного інтелекту, які вже зараз можливо залучити до вирішення задач захисту інфокомунікаційних мереж.

Метою статті є тестування нейромережевих моделей формату GGUF для оцінки можливості їх застосування при вирішенні задачі виявлення заражених ПК на базі цифрових слідів.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

У погоні за трендами впровадження ШІ, більшість компаній, що пропонують захисні рішення для кінцевих точок, такі як CrowdStrike, SentinelOne, Cybereason, Symantec тощо почали додавати до своїх продуктів власні моделі ШІ [16]. Ще не можна остаточно сказати, наскільки впровадження ШІ допомогло при захисті кінцевих точок, оскільки порівняльні тести захисних рішень «до» та «після» впровадження ШІ не проводилися. Для оцінки ефективності захисних рішень із залученням ШІ у корпоративному середовищі можна використовувати матеріал викладений у [17]. Оскільки немає можливості використовувати зазначені моделі локально (агентам ШІ потрібен доступ до інтернету та хмарного сховища виробника) та відокремити їх від основного функціоналу захисного рішення кінцевих точок, а також надавати на обробку власні запити – у цій статті вони не розглядаються. Розглядаються локальні моделі ШІ, сконвертовані у формат GGUF (GPT Generated Unified Format) [18,19].

Існують наступні типи штучного інтелекту [6]:

- *Генеративний ШІ.* Використовує глибокі нейронні мережі для створення контенту на основі заданих даних та попереднього контексту.
- *Розподілений ШІ.* Узгоджено працює над розв'язанням складних завдань.
- *Спеціалізований ШІ.* Вирішує вузькі, специфічні завдання.

Технології штучного інтелекту [6]:

1. ***Машинне навчання (ML).*** Фундаментальна технологія ШІ, яка дозволяє системам навчатися на основі даних та робити прогнози без необхідності явного програмування.
2. ***Глибоке навчання (DL).*** Технологія машинного навчання, що ґрунтується на використанні багатошарових нейронних мереж. Кожна нейронна мережа складається з кількох шарів, де дані проходять через кожен шар, навчаючись знаходити зв'язки та закономірності.
3. ***Рекурентні нейронні мережі (RNN).*** Тип нейронних мереж, який особливо ефективний для роботи з послідовними даними, такими як текст. RNN мають пам'ять, яка дозволяє їм враховувати попередні елементи послідовності при обробці поточного елемента.
4. ***Конволюційні нейронні мережі (CNN).*** Нейронні мережі, призначені для аналізу візуальних даних, які використовують конволюційні фільтри для отримання ключових характеристик, таких як краї, текстури та об'єкти. Ці фільтри сканують зображення частинами, витягуючи важливу інформацію на різних рівнях.
5. ***Трансформери.*** Використовуються для роботи з послідовностями даних, дозволяючи вирішувати такі завдання, як генерація та аналіз контенту та виявлення загроз. На відміну від рекурентних нейронних мереж (RNN), трансформери здатні обробляти весь набір вхідних даних одночасно, що робить їх ефективнішими для завдань прогнозування та генерації. Трансформери використовують механізми уваги, які допомагають моделям зосередитися на найважливіших частинах даних. Саме ця технологія лягла в основу багатьох сучасних мовних моделей, таких як GPT та BERT, які здатні розуміти контекст та генерувати осмислений текст.
6. ***Обробка природної мови (NLP).*** Область штучного інтелекту, що дозволяє машинам розуміти, інтерпретувати та генерувати людську мову. NLP включає кілька ключових технологій, таких як синтаксис, семантика і аналіз лінгвістичних моделей. NLP використовується для виконання таких завдань, як переклад тексту, автоматичне

резюмування, відповіді на запитання та аналіз тональності тексту. Застосування NLP стало можливим завдяки таким технологіям, як трансформери та машинне навчання, які дозволяють моделям розуміти контекст та зміст тексту, а не просто виконувати поверхневий аналіз.

7. ***Автокодувальники.*** Тип нейронних мереж, які навчаються стискати (кодувати) дані в більш компактний вигляд, та потім відновлювати (декодувати) в вихідну форму. Це робиться для виявлення прихованих закономірностей у даних чи виконання завдань зі стискання інформації.

Недоліки використання ШІ у кібербезпеці [6]:

- Потреба у великих обсягах даних;
- Завжди будуть хибні спрацювання (false-positive);
- Обмежені обчислювальні ресурси;
- Питання ліцензування та контролю ШІ;
- Складність адаптації до нормативних вимог регуляторів.

Тестування обраних для вирішення поставленої задачі виявлення заражених ПК на базі цифрових слідів здійснювалось за схемою, поданою на рис. 1.

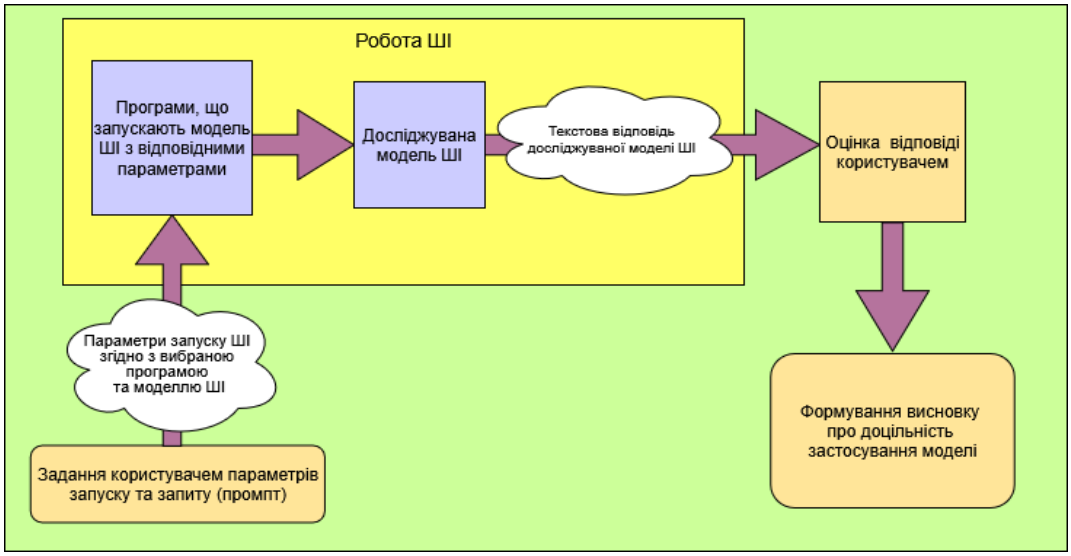


Рис. 1. Схема дослідження нейромережевих моделей

Вимоги до моделей:

1. Сконвертована у формат GGUF;
2. Здатність роботи з текстом у якості «питання->відповідь»;
3. Вміння надавати правильну (іноді розгорнуту і обґрунтовану) відповідь;
4. Здатність запуску і отримання відповіді від моделі засобами командного рядка Windows, без користувацького інтерфейсу;
5. Здатність видавати відповідь за прийнятний час (чим швидше тим краще). В ідеалі — менше ніж 1 хв на відповідь, проте, зважаючи на різницю між корпоративним та домашнім обладнанням, будуть прийнятними всі моделі, що надали відповідь менше ніж за 10 хв;
6. Розмір моделі менший за 100 ГБ.

ISSN 2663 - 4023

З розміщених на HuggingFace [20] обрано 135 моделей формату GGUF, 57 з яких навчено під задачі кібербезпеки (різного розміру, рівня квантування та від різних користувачів). Для тестування підготовлено промпти для моделей, які б містили таку інформацію про цифрові сліди, на основі яких остаточно можна було б сказати, що конкретний ПК заражений, навіть не маючи додаткової інформації про файли, див. [10], [21]. При цьому під терміном «промпт» розуміються спеціальні інструкції на природній мові, які надаються ШІ для виконання певної роботи і надання відповіді у вигляді тексту, картинок, аудіо, відео тощо. Це можуть бути запити, питання або завдання яке ШІ має виконати «відповідаючи» на інструкцію і чим точніше буде інструкція — тим більш релевантною буде відповідь ШІ.

Протестовано різні методи запуску конкретної моделі засобами командного рядка у Windows (оскільки не всі працюють однаково швидко і лише llama-cli може приймати файли у якості промптів): llama-cli, llama-run, koboldcpp. А також розглядалася реакція моделі на запит надання короткої і розширеної відповіді для виявлення коректних обґрунтувань та зміни швидкості відповіді.

Промт для короткої відповіді: "Hello, on my PC I have next digital artifacts in Registry: HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Run RealtekHD C:\ProgramData\WindowsTasks\RealtekHD.exe and in Windows Defender Exclusions: C:\ProgramData\WindowsTasks\RealtekHD.exe. Is it really a virus? Answer in only one word Yes or No. No other words needed.".

Промт для довгої відповіді: "Hello, on my PC I have next digital artifacts in Registry: HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Run RealtekHD C:\ProgramData\WindowsTasks\RealtekHD.exe and in Windows Defender Exclusions: C:\ProgramData\WindowsTasks\RealtekHD.exe. Is it really a virus? Give detailed report.".

Файл промпту містить кілька розділів та додаткову надлишкову для визначення коректної відповіді інформацію, тому що, саме в такому вигляді отримуються цифрові сліди з досліджуваного ПК [10], [21], [22].

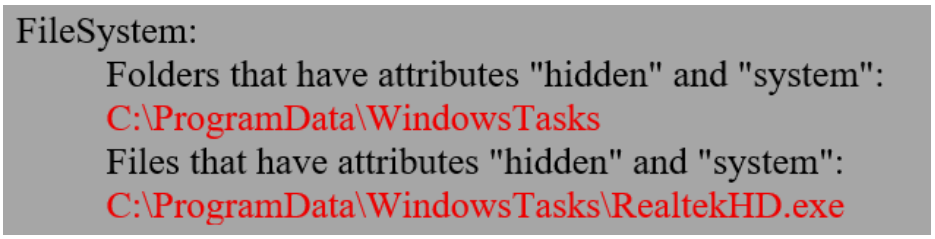
Приклад структури такого файлу:

Hello, I have next digital artifacts on my PC:

Processes:
Node,ExecutablePath
<list of average processes on PC>
<PC name>,C:\ProgramData\WindowsTasks\RealtekHD.exe
<list of average processes on PC>

Registry:
Key,Value Name,Value Type,Value data
HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Run,RealtekHD,REG_SZ,C:\ProgramData\WindowsTasks\RealtekHD.exe
<additional legitimate software>

Windows Defender exclusions (files or folders):
C:\
C:\ProgramData\WindowsTasks\RealtekHD.exe



Is my PC infected by computer virus?
Коротка відповідь: *Answer with only one word Yes or No.*
Розширена відповідь: *Say "Yes" or "No" and give detailed report why do you think so.*

Можливість моделі змінювати свою відповідь при повторних запитах перевірялась на моделі Gemma. Результати тестування зведено у табл. 1 і 2.

Варто зазначити, що незважаючи на явну вказівку надавати коротку відповідь, модель іноді видає більше слів ніж звичайне «так» або «ні», тому в таблиці введено наступні позначення:

No — модель явно надала відповідь НІ;

Yes — модель явно надала відповідь ТАК;

Suspicious — модель не надала явної відповіді, проте вважає файл підозрілим і таким, що потребує подальших перевірок;

Irrelevant — модель не надала нормальної відповіді, проте відповіла беззвязний текст або відповідь була надана зовсім на інше питання (були ситуації, коли модель сама собі поставила питання яке ніяк не зв'язано з промтом та почала відповідати на нього);

Error або пропуск — модель не видала нічого, або зависла при відповіді, або це був непечатний текст.

Оскільки різниці у відповідях не виявлено, наступні моделі тестувалися по одному разу.

Таблиця 1

Результати тестування 5 моделей Gemma з короткою відповіддю

Model name		Short Answer					
	Size of model (GB)	llama-cli with promptfile (answer)	llama-cli with promptfile (seconds)	llama-run with prompt (answer)	llama-run with prompt (seconds)	koboldcpp with prompt (answer)	koboldcpp with prompt (seconds)
gemma-3-270m-it-Q4_K_M	0,24	irrelevant	4.5	Yes	1	irrelevant	11.1
gemma-3-270m-it-Q4_K_M	0,24	irrelevant	4.7	Yes	1	irrelevant	11.3
gemma-3-1b-it-Q4_K_M	0,75	No	10.2	No	1.5	No	12
gemma-3-1b-it-Q4_K_M	0,75	No	9.1	No	1.4	No	12.4
gemma-3-4b-it-Q4_K_M	2,32	Yes	30.3	No	2.6	No	12

gemma-3-4b-it-Q4_K_M	2,32	Yes	29.9	No	2.5	No	52.3
gemma-3-12b-it-Q4_K_M	6,8	No	48.7	No	5.3	No	19.7
gemma-3-12b-it-Q4_K_M	6,8	No	48.4	No	4.9	No	16.8
gemma-3-27b-it-Q4_K_M	15,41	No	101	Yes	10.8	No	117
gemma-3-27b-it-Q4_K_M	15,41	No	102	Yes	12.4	No	118

Таблиця 2

Результати тестування 5 моделей Gemma з довгою відповіддю

Model name		**Long Answer**					
	Size of model (GB)	**llama-cli with long promptfile (answer)**	**llama-cli with long promptfile (seconds)**	**llama-run with long prompt (answer)**	**llama-run with long prompt (seconds)**	**koboldcpp with long prompt (answer)**	**koboldcpp with long prompt (seconds)**
gemma-3-270m-it-Q4_K_M	0,24	irrelevant	2.3	irrelevant	1.3	irrelevant	10.9
gemma-3-270m-it-Q4_K_M	0,24	irrelevant	2.1	suspicious	6.5	irrelevant	11.4
gemma-3-1b-it-Q4_K_M	0,75	irrelevant	5.2	suspicious	42.9	irrelevant	11.6
gemma-3-1b-it-Q4_K_M	0,75	irrelevant	5.3	suspicious	42.1	irrelevant	12.6
gemma-3-4b-it-Q4_K_M	2,32	irrelevant	3.6	suspicious	108	irrelevant	16
gemma-3-4b-it-Q4_K_M	2,32	No	12.7	suspicious	98.6	irrelevant	18.1
gemma-3-12b-it-Q4_K_M	6,8	No	39.3	suspicious	431	suspicious	38.2
gemma-3-12b-it-Q4_K_M	6,8	No	34	suspicious	317	suspicious	36.5
gemma-3-27b-it-Q4_K_M	15,41	No	140	-	-	-	-
gemma-3-27b-it-Q4_K_M	15,41	irrelevant	77.8	suspicious	912	irrelevant	168

Результати тестування моделей Gemma показали, що жодна з них не дає стабільно правильної відповіді залежно від методу запуску, а більшість з них надає неправильну відповідь, отже доцільно розділити досліджувані моделі за методом запуску і не включати у вибірку моделі, які неправильно відповідають на питання. Результати тестування моделей Cybersecurity представлено у табл. 3–8.

Таблиця 3

Результати тестування моделей Cybersecurity з короткою відповіддю

Model name	Size of model (GB)	llama-cli with promptfile (answer)	llama-cli with promptfile (seconds)
Lily-Cybersecurity-7B-v0.2.Q8_0_Quantization-made-by-Richard-Erkhov	7,17	yes	91.5
mistral-v0.3-7b-cybersecurity_unsloth.Q4_K_M	4,07	yes	34.6
senecallm_x_qwen2.5-7b-cybersecurity-q5_k_m_Nekuromento	5,07	yes	34.1
forensicmistra-unsloth.Q8_0	7,17	yes	46.7
forensicmistral_v0.3-unsloth.Q4_K_M	4,07	yes	29.6
CyberSecurity-CHMP-AS-DP8B-V3-GigoFof-unsloth.Q4_K_M	4,58	suspicious	52

Таблиця 4

Результати тестування моделей Cybersecurity з короткою відповіддю

Model name	Size of model (GB)	llama-run with prompt (answer)	llama-run with prompt (seconds)
Cybersecurity-alarm-llama3.2-1B-GGUF-unsloth.Q8_0	1,76	yes	25.8
senecallm-q4_k_m	4,58	yes	3.3
senecallm-x-qwq-32b-q4_k_m	18,49	yes	230
llama-3-8b-instruct-cybersecurity.Q4_K_M	4,58	yes	3.7
forensicmistra-unsloth.Q4_K_M	4,07	suspicious	5.5

Таблиця 5

Результати тестування моделей Cybersecurity з короткою відповіддю

Model name	Size of model (GB)	koboldcpp with prompt (answer)	koboldcpp with prompt (seconds)
forensicmistral_v0.3-unsloth.Q4_K_M	4,07	yes	13.8
SenecaLLM_x_Qwen2.5-7B-CyberSecurity.i1-Q4_K_M	4,36	yes	13.3
SenecaLLM_x_Qwen2.5-7B-CyberSecurity.Q4_K_M	4,36	yes	14.1
Mistral-7B-Instruct-v0.3-Forensics-v1.Q8_0	7,17	yes	14.6
nomic-ai-gpt4all-falcon-Q4_K_M	4,63	yes	12.2
lily-cybersecurity-7b-v0.2-q8_0_Nekuromento	7,17	suspicious	42.6
Lily-Cybersecurity-7B-v0.2.Q8_0_quantized_version	7,17	suspicious	41

Таблиця 6

Результати тестування моделей Cybersecurity з довгою відповіддю

Model name	Size of model (GB)	llama-cli with long promptfile (answer)	llama-cli with long promptfile (seconds)
forensicmistral_v0.3-unsloth.Q4_K_M	4,07	yes	26.6
senecallm_x_qwen2.5-7b-cybersecurity-q5_k_m_Nekuromento	5,07	yes	23.3
qwq-32b-preview-senecallmv1.2-q4_k_m	18,49	yes	194

Таблиця 7

Результати тестування моделей Cybersecurity з довгою відповіддю

Model name	Size of model (GB)	llama-run with long prompt (answer)	llama-run with long prompt (seconds)
~~senecallm-x-qwq-32b-q4_k_m~~	~~18,49~~	~~yes~~	~~942~~
senecallm_x_qwen2.5-7b-cybersecurity-q5_k_m_Nekuromento	5,07	suspicious	34.6
senecallm-q4_k_m	4,58	suspicious	87.3
qwen-cybersecurity-2.5-7b-Armandotrsg-unsloth.F16	14,19	suspicious	79
lily-cybersecurity-7b-v0.2-q5_k_m_Nekuromento	4,78	suspicious	61.2
~~seneca-x-deepseek-r1-distill-qwen-32b-v1.3-q4_k_m~~	~~18,49~~	~~suspicious~~	~~906~~
BaronLLM_Offensive_Security_abliterated_(by_huihui-ai)_llama3.1-v1-q6_k	6,14	suspicious	388
baronllm-llama3.1-v1-q6_k	6,14	suspicious	89

Таблиця 8

Результати тестування моделей Cybersecurity з довгою відповіддю

Model name	Size of model (GB)	koboldcpp with long prompt (answer)	koboldcpp with long prompt (seconds)
BaronLLM_Offensive_Security_abliterated_(by_huihui-ai)_llama3.1-v1-q6_k	6,14	suspicious	38.4
mistral-v0.3-7b-cybersecurity_unsloth.Q4_K_M	4,07	suspicious	29
DeepSeek-Qwen7B-CyberSecurity-unsloth.Q4_K_M	4,36	suspicious	31.1

Додатково було протестовано моделі загального призначення (General), результати представлені у табл. 913.

Таблиця 9

Результати тестування моделей General з короткою відповіддю

Model name	Size of model (GB)	llama-cli with promptfile (answer)	llama-cli with promptfile (seconds)
gemma-2-9b-it-abliterated(by_bartowski)-Q6_K_L	7,27	yes	53
gemma-3-12b-it-abliterated(by_mlabonne)-v2.q8_0	11,65	yes	81.3
gemma-3-27b-it-q4_0_(by_google)	16,05	yes	189

meta-llama-3.1-8b-claude-q8_0	7,95	yes	74.8
mistral-claude-merged.Q5_K_M	4,78	yes	46.9
deepseek-ai_DeepSeek-R1-0528-Qwen3-8B-Q4_K_L	5,11	suspicious	86.2
deepseek-ai_DeepSeek-R1-0528-Qwen3-8B-Q8_0	8,11	suspicious	148
DeepSeek-R1-0528-Qwen3-8B-UD-Q8_K_XL	10,08	suspicious	183
mistral-7b-claude-chat.Q8_0	7,17	suspicious	102

Таблиця 10

Результати тестування моделей General з короткою відповіддю

Model name	Size of model (GB)	llama-run with prompt (answer)	llama-run with prompt (seconds)
~~gemma-3-27b-pt-q4_0_(by_google)~~	~~16,05~~	~~yes~~	~~1250~~
gemma-3-27b-it-q4_0_(by_google)	16,05	yes	11.7
gemma-3-27b-it-Q3_K_M	12,51	yes	12
gemma-3-27b-it-Q4_K_M	15,41	yes	13.8
gemma-3-27b-it-UD-Q4_K_XL	15,67	yes	12.7
gemma-3-27b-it-abliterated(by_mlabonne)-v2.q4_k_m	15,41	yes	13.6
~~gemma-3-27b-it-abliterated(by_mlabonne)-v2.q8_0~~	~~26,74~~	~~yes~~	~~1490~~
~~Kimiko-Claude-FP16.f16~~	~~13,49~~	~~yes~~	~~666~~
mistral-7b-claude-chat.Q4_K_M	4,07	yes	311
Phi4-Reasoning-Merged-15B-Q8_0	14,51	yes	124
tinyllama-claude_16bit_GGUF-unsloth.F16	2,05	yes	4.1

Таблиця 11

Результати тестування моделей General з короткою відповіддю

Model name	Size of model (GB)	koboldcpp with prompt (answer)	koboldcpp with prompt (seconds)
gemma-3-27b-it-abliterated(by_mlabonne)-v2.q8_0	26,74	yes	386
gemma-3-12b-it-abliterated(by_mlabonne)-v2.q8_0	11,65	yes	97.2
mistral-7b-claude-chat.Q8_0	7,17	yes	40.5
gemma-3-12b-it-abliterated(by_mlabonne)-v2.q4_k_m	6,8	yes	16.9
gemma-3-1b-it-Q4_K_M	0,75	yes	12.1
DeepSeek-R1-Distill-Llama-8B-Q4_K_M	4,58	yes	13.5
meta-llama-3.1-8b-instruct-abliterated(by_mlabone).Q4_K_M	4,58	yes	22.7
Kimiko-Claude-FP16.Q4_K_M	4,07	yes	19.5
deepseek-ai_DeepSeek-R1-0528-Qwen3-8B-Q8_0	8,11	suspicious	49.3
DeepSeek-R1-0528-Qwen3-8B-UD-Q8_K_XL	10,08	suspicious	173

Таблиця 12

Результати тестування моделей General з довгою відповіддю

Model name	Size of model (GB)	llama-cli with long promptfile (answer)	llama-cli with long promptfile (seconds)
gemma-3-27b-it-Q4_K_M	15,41	yes	173
meta-llama-3.1-8b-claude-q8_0	7,95	yes	35.1
meta-llama-3.1-8b-claude-q4_k_m	4,58	yes	20.5
llama-7b.Q5_K_M	4,45	yes	25.5
claude-3.7-sonnet-reasoning-gemma3-12B.Q8_0	11,65	suspicious	108
oh-dcft-v3.1-claude-3-5-haiku-20241022-qwen-q4_k_m-imat	4,36	suspicious	19.5
Kimiko-Claude-FP16.f16	13,49	suspicious	87.1
DeepSeek-R1-0528-Qwen3-8B-Q4_K_M	4,68	suspicious	30.6
Kimiko-Claude-FP16.Q8_0	7,17	suspicious	32.1
Phi-4-mini-instruct-Q4_K_M	2,32	suspicious	4.5

Таблиця 13

Результати тестування моделей General з довгою відповіддю

Model name	Size of model (GB)	llama-run with long prompt (answer)	llama-run with long prompt (seconds)
~~gemma-3-27b-it-Q3_K_M~~	~~12,51~~	~~yes~~	~~851~~
gemma-3-27b-pt-q4_0_(by_google)	16,05	yes	251
oh-dcft-v3.1-claude-3-5-haiku-20241022-qwen-q4_k_m-imat	4,36	suspicious	32.6
DeepSeek-R1-0528-Qwen3-8B-Q4_K_M	4,68	suspicious	303
DeepSeek-R1-Distill-Llama-8B-Q4_K_M	4,58	suspicious	153
gemma-3-12b-it-Q5_K_M	7,87	suspicious	471
Qwen2.5-32B-Instruct-Q4_K_L	19,03	suspicious	364
meta-llama-3.1-8b-instruct-abliterated(by_mlabone).Q8_0	7,95	suspicious	167
mistral-7b-instruct-v0.2.Q4_K_M	4,07	suspicious	60
~~Qwen3-32B-Q4_K_M~~	~~18,4~~	~~suspicious~~	~~1171~~
deepseek-ai_DeepSeek-R1-0528-Qwen3-8B-Q8_0	8,11	suspicious	603
DeepSeek-R1-0528-Qwen3-8B-UD-Q8_K_XL	10,08	suspicious	518
~~deepseek-ai_DeepSeek-R1-0528-Qwen3-8B-bf16~~	~~15,26~~	~~suspicious~~	~~792~~
~~gemma-3-27b-it-UD-Q4_K_XL~~	~~15,67~~	~~suspicious~~	~~1019~~
mistral-7b-claude-chat.Q4_K_M	4,07	suspicious	324

Моделі General із довгою відповіддю під час запуску з koboldcpp не давали правильної відповіді, а сам koboldcpp іноді використовував всю доступну пам’ять на ПК, тому результат їх виконання у таблиці не відображений.

Для визначення переліку доцільних для використання моделей, отримані результати у зведених таблицях узагальнено за часом, що був витрачений на відповідь а також за способом запуску моделі.

Для моделей Gemma

Для використання з файлом промпту llama-cli, більше підходить gemma-3-4b-it-Q4_K_M, що важить всього 2,32 ГБ з 30 сек.

Для використання з промптом llama-run, більше підходить gemma-3-270m-it-Q4_K_M, що важить всього 0,24 ГБ з 1 сек та gemma-3-27b-it-Q4_K_M що важить 15,41ГБ з 11 сек.

Для моделей Cybersecurity з короткою відповіддю

Для використання з файлом промпту llama-cli, більше підходить:

- forensicmistral_v0.3-unsloth.Q4_K_M(29.6sec);
- senecallm_x_qwen2.5-7b-cybersecurity-q5_k_m_Nekuromento (34.1sec);
- mistral-v0.3-7b-cybersecurity_unsloth.Q4_K_M (34.6sec);
- forensicmistra-unsloth.Q8_0 (46.7sec);
- Lily-Cybersecurity-7B-v0.2.Q8_0_Quantization-made-by-Richard-Erkhov (91.5sec).

Для використання з промптом llama-run, більше підходить:

- senecallm-q4_k_m (3.3 sec);
- llama-3-8b-instruct-cybersecurity.Q4_K_M (3.7 sec);
- Cybersecurity-alarm-llama3.2-1B-GGUF-unsloth.Q8_0 (25.8 sec);
- senecallm-x-qwq-32b-q4_k_m (230 sec).

Для використання з промптом koboldcpp, більше підходить:

- nomic-ai-gpt4all-falcon-Q4_K_M (12.2 sec);
- SenecaLLM_x_Qwen2.5-7B-CyberSecurity.i1-Q4_K_M (13.3 sec);
- forensicmistral_v0.3-unsloth.Q4_K_M (13.8 sec);
- SenecaLLM_x_Qwen2.5-7B-CyberSecurity.Q4_K_M (14.1 sec);
- Mistral-7B-Instruct-v0.3-Forensics-v1.Q8_0 (14.6 sec).

Для моделей Cybersecurity з розширеною відповіддю

Для використання з файлом промпту llama-cli, більше підходить:

- senecallm_x_qwen2.5-7b-cybersecurity-q5_k_m_Nekuromento (23.3 sec);
- forensicmistral_v0.3-unsloth.Q4_K_M (26.6 sec);
- qwq-32b-preview-senecallmv1.2-q4_k_m (194 sec).

Для моделей General з короткою відповіддю

Для використання з файлом промпту llama-cli більше підходить:

- mistral-claude-merged.Q5_K_M (46.9 sec);
- gemma-2-9b-it-abliterated(by_bartowski)-Q6_K_L (53 sec);
- meta-llama-3.1-8b-claude-q8_0 (74.8 sec);
- gemma-3-12b-it-abliterated(by_mlabonne)-v2.q8_0 (81.3 sec);
- gemma-3-27b-it-q4_0_(by_google) (189 sec).

Для використання з промптом llama-run більше підходить:

- tinyllama-claude_16bit_GGUF-unsloth.F16 (4.1 sec);
- gemma-3-27b-it-q4_0_(by_google) (11.7 sec);
- gemma-3-27b-it-Q3_K_M (12 sec);

- gemma-3-27b-it-UD-Q4_K_XL (12.7 sec);
- gemma-3-27b-it-abliterated(by_mlabonne)-v2.q4_k_m (13.6 sec);
- gemma-3-27b-it-Q4_K_M (13.8 sec);
- Phi4-Reasoning-Merged-15B-Q8_0 (124 sec);
- mistral-7b-claude-chat.Q4_K_M (311 sec).

Для використання з промптом koboldcpp більше підходить:

- gemma-3-1b-it-Q4_K_M (12.1 sec);
- DeepSeek-R1-Distill-Llama-8B-Q4_K_M (13.5 sec);
- gemma-3-12b-it-abliterated(by_mlabonne)-v2.q4_k_m (16.9 sec);
- Kimiko-Claude-FP16.Q4_K_M (19.5 sec);
- meta-llama-3.1-8b-instruct-abliterated(by_mlabone).Q4_K_M (22.7 sec);
- mistral-7b-claude-chat.Q8_0 (40.5 sec);
- gemma-3-12b-it-abliterated(by_mlabonne)-v2.q8_0 (97.2 sec);
- gemma-3-27b-it-abliterated(by_mlabonne)-v2.q8_0 (386 sec).

Для моделей General з розширеною відповіддю

Для використання з файлом промпту llama-cli більше підходить:

- meta-llama-3.1-8b-claude-q4_k_m (20.5 sec);
- llama-7b.Q5_K_M 4,45 (25.5 sec);
- meta-llama-3.1-8b-claude-q8_0 (35.1 sec);
- gemma-3-27b-it-Q4_K_M (173 sec).

Для використання з промптом llama-run більше підходить gemma-3-27b-pt-q4_0_(by_google) (251 sec).

Така варіативність у відповідях моделей і у методах запуску, дозволяє зробити висновок, що використання моделей ШІ, а особливо тих, які були спеціально навчені під задачі кібербезпеки, вже зараз може давати прийнятні результати при визначенні зараженості ПК по цифрових слідах, що у свою чергу підвищить швидкість реагування на інциденти на підприємстві, проте остаточний вибір моделі буде залежати від актуальних задач та наявних ресурсів.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі розглянуто типи і технології штучного інтелекту, та їх вплив на кібербезпеку і в якості захисту від кібератак, і в якості одного із компонентів для атак на інформаційну інфраструктуру.

Щоб оцінити можливості застосування існуючих моделей ШІ для вирішення актуальних задач кіберзахисту, зокрема виявлення заражених ПК на базі цифрових слідів із застосуванням ШІ, визначено критерії для моделі ШІ які будуть прийнятними для використання у корпоративному середовищі та проведено тестування 135 моделей формату GGUF на предмет виявлення або невиявлення ними ознак вірусної активності та індикаторів компрометації у промпті, що надавався користувачем.

Оскільки виявлено, що при запуску однієї і тієї ж нейромережевої моделі з однаковими промптами але різними програмами, що можуть запускати локальні моделі на ПК, її відповідь кардинально змінюється, підготовлено низку зведених таблиць де є назва моделі і варіанти відповідей під кожну програму для запуску ШІ моделей, без

урахування тих, що надали неправильну відповідь, витратили надто багато часу на відповідь або завершилися з помилкою.

Визначено перелік доцільних для використання моделей ШІ у форматі GGUF для вирішення задач кібербезпеки, зокрема для виявлення заражених ПК на базі цифрових слідів. Проте, так як кожна модель краще проявляє себе у специфічних умовах із різними сценаріями запуску, вибір моделі буде залежати від актуальних задач та наявних ресурсів.

Подальші дослідження можуть бути зосереджені на вдосконаленні методики дослідження моделей для обробки цифрових слідів, перетворенні цифрових слідів з ПК у зрозумілий для ШІ промт та автоматизацію аналізу відповіді ШІ.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Microsoft. (2025, August 3). *What is AI for cybersecurity? | Microsoft Security Essentials*. Microsoft. https://www.microsoft.com/uk-ua/security/business/security-101/what-is-ai-for-cybersecurity
2. Kostiuk, Yu. V., Skladannyi, P. M., Bebeshko, B. T., Khorolska, K. V., Rzaieva, S. L., & Vorokhob, M. V. (2025). *Information and communication systems security* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.
3. Kostiuk, Yu. V., Skladannyi, P. M., Hulak, H. M., Bebeshko, B. T., Khorolska, K. V., & Rzaieva, S. L. (2025). *Information security systems* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.
4. Hulak, H. M., Zhyltsov, O. B., Kyrychok, R. V., Korshun, N. V., & Skladannyi, P. M. (2023). *Enterprise information and cyber security* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.
5. World Economic Forum. (2025, August 5). *Global cybersecurity outlook 2025*. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf
6. Oberig IT. (2025, August 3). *Artificial Intelligence (AI) and Privileged Access Management (PAM) Blog*. https://oberig-it.com/statti/shtuchnyj-intelekt-shi-ta-upravlinnya-pryvilejovanym-dostupom-pam/
7. Weigand, S. (2025, August 4). *2025 forecast: AI to supercharge attacks, quantum threats grow, SaaS security woes*. SC Media. https://www.scworld.com/feature/cybersecurity-threats-continue-to-evolve-in-2025-driven-by-ai
8. AV-ATLAS. (2025, September 22). *AV-ATLAS – & PUA*. https://portal.av-atlas.org/malware
9. Kalash, M., et al. (2018). Malware classification with deep convolutional neural networks. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. https://doi.org/10.1109/NTMS.2018.8328749
10. Chernihivskyi, I., & Kriuchkova, L. (2025). Systematic approach to solving the task of protecting information in the infocommunication network from the influence of computer viruses. *Cybersecurity: Education, Science, Technique*, (27), 572–590. https://doi.org/10.28925/2663-4023.2025.27.781
11. BlackBerry. (2025, August 5). *Predictive AI for cybersecurity. What actually works and how to understand it*. https://blackberry.bakotech.com/ua/predictive-ai-for-cybersecurity
12. Otal, H. T., & Canbaz, M. A. (2024). LLM Honeypot: Leveraging large language models as advanced interactive honeypot systems. *IEEE Conference on Communications and Network Security (CNS)*. https://doi.org/10.1109/CNS62487.2024.10735607
13. Gholami, Y. (2024). Large language models (LLMs) for cybersecurity: A systematic review. *World Journal of Advanced Engineering Technology and Sciences, 13*(1), 57–69. https://doi.org/10.30574/wjaets.2024.13.1.0395
14. Coppolino, L., et al. (2025). The good, the bad, and the algorithm: The impact of generative AI on cybersecurity. *Neurocomputing, 623*, 129406. https://doi.org/10.1016/j.neucom.2025.129406
15. Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security, 81*, 123–147. https://doi.org/10.1016/j.cose.2018.11.001
16. Unite.AI. (2025, September 8). *Top 10 AI cybersecurity tools (September 2025)*. Unite.AI – AI News. https://www.unite.ai/uk/ai-cybersecurity-tools/
17. Chernihivskyi, I., & Kriuchkova, L. (2025). Testing antivirus solutions for the corporate segment. *Information Security. Scientific Journals of the State University "Kyiv Aviation Institute"*. https://jrnl.nau.edu.ua/index.php/Infosecurity/article/view/20362

18. Hugging Face. (2025, August 8). *GGUF*. *Hugging Face – The AI community building the future*. https://huggingface.co/docs/hub/gguf
19. ggml-org. (2025, August 8). *GGUF ggml/docs/gguf.md at master*. GitHub. https://github.com/ggml-org/ggml/blob/master/docs/gguf.md
20. Hugging Face. (2025, August 8). *Hugging Face – The AI community building the future*. https://huggingface.co
21. Chernihivskyi, I., & Kriuchkova, L. (2025). Effective solutions for rapid detection of committed PCs in the infocommunication networks. *Telecommunication and Information Technologies*, *87*(2). https://doi.org/10.31673/2412-4338.2025.029875
22. Bohdanov, O., & Chernihivskyi, I. (2024). Types of digital forensic artifacts in Windows computers. *Cybersecurity: Education, Science, Technique, 4*(24), 221–228. https://doi.org/10.28925/2663-4023.2024.24.221228

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023

Ivan Chernihivskyi
Graduate Student
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0009-0003-4568-3212
i.chernihivskiy@gmail.com

Larysa Kriuchkova
Doctor of Technical Sciences, Professor
Professor of the Volodymyr Buriachok Department of Information and Cyber Security
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0002-8509-6659
l.kriuchkova@kubg.edu.ua

TESTING NEURAL NETWORK MODELS FOR SOLVING THE PROBLEM OF DETECTING INFECTED PCS BASED ON DIGITAL TRACES

Abstract. The development of artificial intelligence has made great progress and already today has a significant impact on a large number of industries and with the development of LLM will have an even greater impact in the future, especially on cybersecurity. AI can both help save data by early detection of cyberattacks, and harm cybersecurity by facilitating the writing of convincing phishing emails, reproducing fragments of malicious code, helping to identify weak points in the network, and finding vulnerabilities in the operating system, programs, etc. that are still unknown to software manufacturers (zero day vulnerability). Therefore, in order not to be lagging behind in this "arms race", it is necessary to already implement AI as one of the components of cyber protection in the enterprise. The relevance of the work lies in the need to find such artificial intelligence models that can already be involved in solving the problems of protecting infocommunication networks. The purpose of the article is to test neural network models of the GGUF format to assess the possibility of their application in solving the problem of detecting infected PCs based on digital traces. The paper considers the types and technologies of artificial intelligence, and their impact on cybersecurity both as protection against cyberattacks and as one of the components for attacks on information infrastructure. In order to assess the possibilities of using existing AI models to solve current cyberdefense problems, in particular, detecting infected PCs based on digital traces using AI, criteria were determined for an AI model that would be acceptable for use in a corporate environment and 135 GGUF format models were tested for their detection or non-detection of signs of viral activity and indicators of compromise in the prompt provided by the user. Since it was found that when running the same neural network model with the same prompts but different programs that can run local models on a PC, its response changes dramatically, a number of summary tables were prepared with the name of the model and answer options for each program for running AI models, excluding those that gave the wrong answer, took too long to answer, or ended with an error. A list of AI models in the GGUF format that are appropriate for use in solving cybersecurity problems, in particular for detecting infected PCs based on digital traces, was determined. However, since each model performs better in specific conditions with different launch scenarios, the choice of model will depend on the current tasks and available resources. Further research can be focused on improving the methodology for studying models for processing digital traces, converting digital traces from a PC into a prompt understandable for AI, and automatically analyzing the AI response.

Keywords: artificial intelligence; neural network models; LLM; computer viruses; digital artifacts; testing; prompt; cybersecurity.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Microsoft. (2025, August 3). *What is AI for cybersecurity? | Microsoft Security Essentials*. Microsoft. https://www.microsoft.com/uk-ua/security/business/security-101/what-is-ai-for-cybersecurity
2. Kostiuk, Yu. V., Skladannyi, P. M., Bebeshko, B. T., Khorolska, K. V., Rzaieva, S. L., & Vorokhob, M. V. (2025). *Information and communication systems security* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.

3. Kostiuk, Yu. V., Skladannyi, P. M., Hulak, H. M., Bebeshko, B. T., Khorolska, K. V., & Rzaieva, S. L. (2025). *Information security systems* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.
4. Hulak, H. M., Zhyltsov, O. B., Kyrychok, R. V., Korshun, N. V., & Skladannyi, P. M. (2023). *Enterprise information and cyber security* [Textbook]. Kyiv: Borys Grinchenko Kyiv Metropolitan University.
5. World Economic Forum. (2025, August 5). *Global cybersecurity outlook 2025*. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf
6. Oberig IT. (2025, August 3). *Artificial Intelligence (AI) and Privileged Access Management (PAM) Blog*. https://oberig-it.com/statti/shtuchnyj-intelekt-shi-ta-upravlinnya-pryvilejovanym-dostupom-pam/
7. Weigand, S. (2025, August 4). *2025 forecast: AI to supercharge attacks, quantum threats grow, SaaS security woes*. SC Media. https://www.scworld.com/feature/cybersecurity-threats-continue-to-evolve-in-2025-driven-by-ai
8. AV-ATLAS. (2025, September 22). *AV-ATLAS – & PUA*. https://portal.av-atlas.org/malware
9. Kalash, M., et al. (2018). Malware classification with deep convolutional neural networks. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. https://doi.org/10.1109/NTMS.2018.8328749
10. Chernihivskyi, I., & Kriuchkova, L. (2025). Systematic approach to solving the task of protecting information in the infocommunication network from the influence of computer viruses. *Cybersecurity: Education, Science, Technique*, (27), 572–590. https://doi.org/10.28925/2663-4023.2025.27.781
11. BlackBerry. (2025, August 5). *Predictive AI for cybersecurity. What actually works and how to understand it*. https://blackberry.bakotech.com/ua/predictive-ai-for-cybersecurity
12. Otal, H. T., & Canbaz, M. A. (2024). LLM Honeypot: Leveraging large language models as advanced interactive honeypot systems. *IEEE Conference on Communications and Network Security (CNS)*. https://doi.org/10.1109/CNS62487.2024.10735607
13. Gholami, Y. (2024). Large language models (LLMs) for cybersecurity: A systematic review. *World Journal of Advanced Engineering Technology and Sciences, 13*(1), 57–69. https://doi.org/10.30574/wjaets.2024.13.1.0395
14. Coppolino, L., et al. (2025). The good, the bad, and the algorithm: The impact of generative AI on cybersecurity. *Neurocomputing, 623*, 129406. https://doi.org/10.1016/j.neucom.2025.129406
15. Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security, 81*, 123–147. https://doi.org/10.1016/j.cose.2018.11.001
16. Unite.AI. (2025, September 8). *Top 10 AI cybersecurity tools (September 2025)*. Unite.AI – AI News. https://www.unite.ai/uk/ai-cybersecurity-tools/
17. Chernihivskyi, I., & Kriuchkova, L. (2025). Testing antivirus solutions for the corporate segment. *Information Security. Scientific Journals of the State University "Kyiv Aviation Institute"*. https://jrnl.nau.edu.ua/index.php/Infosecurity/article/view/20362
18. Hugging Face. (2025, August 8). *GGUF. Hugging Face – The AI community building the future*. https://huggingface.co/docs/hub/gguf
19. ggml-org. (2025, August 8). *GGUF ggml/docs/gguf.md at master*. GitHub. https://github.com/ggml-org/ggml/blob/master/docs/gguf.md
20. Hugging Face. (2025, August 8). *Hugging Face – The AI community building the future*. https://huggingface.co
21. Chernihivskyi, I., & Kriuchkova, L. (2025). Effective solutions for rapid detection of committed PCs in the infocommunication networks. *Telecommunication and Information Technologies, 87*(2). https://doi.org/10.31673/2412-4338.2025.029875
22. Bohdanov, O., & Chernihivskyi, I. (2024). Types of digital forensic artifacts in Windows computers. *Cybersecurity: Education, Science, Technique, 4*(24), 221–228. https://doi.org/10.28925/2663-4023.2024.24.221228