МЕТОДОЛОГІЯ АВТОМАТИЗОВАНОЇ КЛАСИФІКАЦІЇ ТА АНАЛІЗУ ЦИФРОВИХ АРТЕФАКТІВ ЗА ДОПОМОГОЮ ЛОКАЛЬНИХ LLM

Самарець В.Є.

Кафедра інфокомунікаційної інженерії ім. В.В. Поповського
Харківський національний університет радіоелектроніки,
Україна.

E-mail: vladyslav.samarets@nure.ua

Abstract

This paper presents a comprehensive methodology for automated classification and analysis of digital forensic artifacts using local Large Language Models. The research explores the integration of LLM-driven analysis with traditional forensic workflows, focusing on artifact extraction, intelligent categorization, and automated metadata processing. We examine practical implementation approaches including Retrieval-Augmented Generation (RAG) frameworks, timeline analysis automation, and intelligent search capabilities for evidence management. The paper discusses techniques for optimizing LLM performance through quantization methods while maintaining forensic accuracy. Experimental results demonstrate significant improvements in investigation efficiency, with automated systems reducing manual analysis effort and enhancing evidence retrieval precision. Particular attention is given to maintaining chain of custody and ensuring reliability of AI-discovered evidence.

Зростаючий обсяг цифрових доказів у сучасних розслідуваннях створює критичні виклики для правоохоронних органів щодо своєчасного проведення як кібер-розслідувань, так і традиційних розслідувань. Автоматизація процесів збору, класифікації та аналізу цифрових артефактів стає необхідною умовою ефективної криміналістичної роботи.

Інтеграція великих мовних моделей у процеси цифрової криміналістики має потенціал суттєво підвищити ефективність розслідувань, однак потребує ретельного дослідження питань надійності, точності та юридичної прийнятності доказів, виявлених за допомогою штучного інтелекту. Локальне розгортання LLM забезпечує необхідний рівень контролю над процесами обробки чутливих даних та дозволяє створити відтворювані результати аналізу.

Ефективна методологія автоматизованого аналізу цифрових артефактів повинна включати кілька ключових компонентів. Перший етап передбачає автоматизовану екстракцію криміналістичних артефактів із різних джерел даних, включаючи файлові системи, мережевий трафік та оперативну пам'ять.

Структурована система автоматизує екстракцію криміналістичних артефактів, уточнює дані через аналіз на основі LLM та валідує результати. Використання архітектури Retrieval-Augmented Generation (RAG) дозволяє моделі отримувати доступ до спеціалізованої бази знань криміналістичних процедур та типових артефактів.

Фреймворк GenDFIR демонструє ефективність поєднання великих мовних моделей, зокрема Llama 3.1 8B у режимі zero-shot, із RAG-агентом для аналізу часових ліній кіберінцидентів. Система аналізує артефакти та події, особливо часові мітки та метадані, для виявлення аномалій у цифрових слідах.

Постійно зростаючий обсяг даних у цифрових криміналістичних розслідуваннях є одним із найбільш обговорюваних викликів у цій галузі. Зазвичай більшість файлових артефактів на вилучених пристроях не є релевантними для розслідування, що створює необхідність автоматичної пріоритизації підозрілих артефактів.

Методологія автоматичної пріоритизації передбачає роботу в режимі human-in-the-loop, тобто система прогнозує та рекомендує, що артефакт є ймовірно релевантним для розслідування. AI-керована система управління доказами автоматизує процеси через використання машинного навчання, обробки природної мови та автоматизації для покращення ефективності, точності та масштабованості розслідувань.

Розумне тегування за допомогою штучного інтелекту автоматизує категоризацію шляхом аналізу вмісту, ідентифікації релевантних сутностей та призначення відповідних міток файлам. Це забезпечує швидкий пошук та перехресне посилання на докази без ризику втрати інформації, особливо в масштабних або резонансних справах.

Знаходження специфічних фрагментів цифрових доказів у масивних наборах даних є значним викликом для слідчих. Можливості пошуку на основі штучного інтелекту дозволяють офіцерам швидко локалізувати релевантні файли, текст, усне мовлення та об'єкти через розпізнавання ключових слів, ідентифікацію облич, екстракцію метаданих та багатомовну підтримку.

Замість ручного перегляду годин відеозаписів або тисяч документів слідчі можуть використовувати AI-керовані системи управління цифровими доказами для проведення миттєвого пошуку, що радикально скорочує час обробки справи. Це забезпечує швидше отримання критичних доказів, покращуючи прийняття рішень у часово критичних розслідуваннях.

Система може відповідати на запити щодо ідентифікації та детального опису цифрових артефактів, виявлених у дослідженні, зосереджуючись на їх типах та криміналістичній релевантності. Також можливі запити щодо специфічних системних локацій (мережа, диск, пам'ять), де ці артефакти можуть бути знайдені.

Квантизація дозволяє використовувати потужність LLM навіть там, де апаратні ресурси далекі від ідеальних. Завдяки цій техніці модель стає компактнішою та швидшою, водночас зберігаючи точність, яка задовольняє більшість завдань криміналістичного аналізу. Зі зростанням розміру та складності моделей вимоги до пам'яті та обчислювальної потужності можуть здаватися непосильними для локальних середовищ. Квантизація переводить параметри моделі з 32-бітної плаваючої точки у 16-бітні чи навіть 8-бітні цілі числа, що значно зменшує вимоги до ресурсів. Квантовані версії різних моделей доступні на платформах HuggingFace, Ollama та LMStudio. Завдяки квантизації можливо використовувати навіть моделі з 32 мільярдами параметрів на звичайних ноутбуках, що робить потужні LLM доступними для локальних криміналістичних лабораторій.

Людські помилки в управлінні цифровими доказами можуть призвести до неправильної класифікації, пропуску файлів або неточного введення даних. Такі помилки можуть затримати справи, ввести в оману розслідування або навіть вплинути на судові процеси. Штучний інтелект усуває невідповідності шляхом автоматичного тегування, індексації та верифікації цифрових доказів. Просунуті алгоритми забезпечують правильну класифікацію, перехресне посилання та безпечне зберігання кожного фрагмента доказів. З автоматизацією на основі AI правоохоронні органи можуть покладатися на точну категоризацію та отримання доказів, зменшуючи ризик пропуску критичної інформації. Це забезпечує створення сильніших справ без помилок, які витримують перевірку в суді.

Критичним аспектом використання великих мовних моделей у цифровій криміналістиці є забезпечення надійності та юридичної прийнятності виявлених доказів. Запропоновані структуровані фреймворки включають автоматизацію екстракції криміналістичних артефактів, уточнення даних через LLM-керований аналіз та валідацію результатів. Дослідження надійності цифрових криміналістичних доказів, виявлених великими мовними моделями, має вирішальне значення для забезпечення їх юридичної обґрунтованості. Процедури валідації повинні включати перехресну перевірку результатів, документування процесу аналізу та підтримку ланцюга зберігання доказів. Питання прозорості, підзвітності та надійної стандартизації у криміналістичному застосуванні LLM потребують подальших досліджень. Також критично важливими є питання інтерпретованості результатів, можливої упередженості моделей та етичних міркувань щодо автоматизованого прийняття рішень у правовому контексті.

Практична цінність локальних LLM значно підвищується при їх інтеграції з існуючими криміналістичними інструментами. Дослідження оцінювали життєздатність використання LLM для створення нових плагінів для криміналістичних інструментів Autopsy та Volatility 3, використовуючи доступні онлайн моделі.

Таким чином, локальне розгортання LLM дає розробникам змогу експериментувати з ідеями швидше й економніше, зменшуючи витрати на створення прототипів. Інтеграція з локальними API, сумісними з OpenAI, забезпечує плавну роботу системи та зберігає гнучкість для майбутнього переходу на різні архітектури моделей.

Література

1. Fatma Yasmine Loumachi, Mohamed Chahine Ghanem, Mohamed Amine Ferrag. GenDFIR: Advancing Cyber Incident Timeline Analysis Through Retrieval Augmented Generation and Large Language Models // arXiv. 2024. URL: https://arxiv.org/abs/2409.02572.
2. Akila Wickramasekara, Frank Breitinger, Mark Scanlon. Exploring the Potential of Large Language Models for Improving Digital Forensic Investigation Efficiency // arXiv. 2025. URL: https://arxiv.org/abs/2402.19366.
3. Xiaoyu Du, Mark Scanlon. Methodology for the Automated Metadata-Based Classification of Incriminating Digital Forensic Artefacts. URL: https://dl.acm.org/doi/10.1145/3339252.3340517.
4. Sarim Suleman**.** 12 Ways to Boost Efficiency with AI in Evidence Management // Digital Evidence. 2024. URL: https://digitalevidence.ai/blog/ai-digital-evidence-management.
5. Jeel Piyushkumar Khatiwala, Daniel Kwaku Ntiamoah Addai, Weifeng Xu. Evaluating the Reliability of Digital Forensic Evidence Discovered by LLM // IEEE Xplore. 2025. URL: https://ieeexplore.ieee.org/document/11126744.
6. Clinton Joel Walker. Artifact Analysis and Using AI in the Forensics Domain // Louisiana State University Repository. 2025. URL: https://repository.lsu.edu/gradschool_dissertations/6884/.
7. Binaya Sharma, James Ghawaly, Kyle McCleary, AndrewM. Webb, Ibrahim Baggili. Forensic LLM: A Local Large Language Model for Digital Forensics // DFRWS. 2025. URL: https://www.sciencedirect.com/science/article/pii/S2666281725000113.
8. Козлов І. Розгортання LLM локально: усе, що потрібно знати // Dev.ua. 2025. URL: https://dev.ua/blogs/posts/rozhortannia-llm-lokalno.
9. Zhipeng Yin , Zichong Wang , Weifeng Xu , Jun Zhuang , Pallab Mozumder , Antoinette Smith , Wenbin Zhang. Digital Forensics in the Age of Large Language Models // arXiv. 2025. URL: https://arxiv.org/abs/2504.02963.