Використання мультимодальних великих мовних моделей для цифрової криміналістики з метою виявлення військовослужбовців на зображеннях із мобільних пристроїв

Application of Multimodal LLMs for Identifying Military Individuals in Mobile-Captured Images within Digital Forensics

Тарас Фединишин [A]
Corresponding author: аспірант, асистент кафедри захисту інформації, e-mail: taras.o.fedynyshyn@lpnu.ua, ORCID: 0009-0006-8233-8057

Сергій Висоцький [A]
студент кафедри захисту інформації, e-mail: serhii.vysotskyi.kb.2021@lpnu.ua, ORCID: 0009-0000-5685-7503

Марія Хомік [A]
студент кафедри захисту інформації, e-mail: mariia.khomik.kb.2021@lpnu.ua, ORCID: 0009-0004-6031-5618

Олександр Гимза [A]
студент кафедри захисту інформації, e-mail: oleksandr.hymza.kb.2021@lpnu.ua, ORCID: 0009-0009-7928-7545

Анастасія Василиця [A]
студент кафедри захисту інформації, e-mail: anastasiia.vasylytsia.kb.2021@lpnu.ua, ORCID: 0009-0000-9133-8338

Богдан Гарасимчук [A]
студент кафедри захисту інформації, e-mail: bohdan.harasymchuk.kb.2021@lpnu.ua, ORCID: 0009-0008-6075-4820

Taras Fedynyshyn [A]
Corresponding author: Postgraduate student, Assistant Lecturer, e-mail: taras.o.fedynyshyn@lpnu.ua, ORCID: 0009-0006-8233-8057

Serhii Vysotskyi [A]
student, e-mail: serhii.vysotskyi.kb.2021@lpnu.ua, ORCID: 0009-0000-5685-7503

Mariia Khomik [A]
student, e-mail: mariia.khomik.kb.2021@lpnu.ua, ORCID: 0009-0004-6031-5618

Oleksandr Hymza [A]
student, e-mail: oleksandr.hymza.kb.2021@lpnu.ua, ORCID: 0009-0009-7928-7545

Anastasia Vasylytsia [A]
student, e-mail: anastasiia.vasylytsia.kb.2021@lpnu.ua, ORCID: 0009-0000-9133-8338

Bohdan Harasymchuk [A]
student, e-mail: anastasiia.vasylytsia.kb.2021@lpnu.ua, ORCID: 0009-0008-6075-4820

[A] Національний університет "Львівська політехніка", Львів, Україна
[A] Lviv Polytechnic National University, Lviv, Ukraine

Received: March 28, 2025 | Revised: April 23, 2025 | Accepted: April 30, 2025

DOI: 10.33445/sds.2025.15.2.15

Мета роботи: Дослідження проведено для оцінки можливості використання мультимодальних великих мовних моделей для виявлення військовослужбовців на зображеннях з мобільних пристроїв. Метою було з'ясувати, чи можуть ці моделі ефективно розрізняти реальних військових та манекенів у складних реалістичних умовах.

Метод дослідження: Кількісні та експериментальні методи, зокрема застосування мультимодальних моделей штучного інтелекту (Google Gemini 1.5 Pro та LLAVA) для генерації описів та класифікації зображень, а також структурованого аналізу даних з використанням статистичних метрик (точність, повнота та точність класифікації) для оцінки ефективності виявлення військовослужбовців. Дослідження проведено на вибірці з 436 зображень, що включає фото військових, манекенів у військовій формі та цивільних осіб, здобутих із резервної копії iOS.

Результати дослідження: Обидві моделі показали високу точність у виявленні військових (точність = 1.0), повнота = 0.99 (Gemini) / 0.98 (LLAVA). Проте 88 з 99 манекенів були помилково класифіковані Gemini як військові, LLAVA — 86. Gemini значно перевищив LLAVA у виявленні країни (0.7875 проти 0.1218) та підрозділу (0.2544 проти 0.0051). Загальна точність: 0.768 (Gemini), 0.764 (LLAVA).

Теоретична цінність дослідження: Дослідження розширює застосування мультимодальних великих мовних моделей

Purpose: The purpose of this research was to evaluate the effectiveness of multimodal Large Language Models (LLMs) in detecting military personnel in mobile device images, particularly in challenging scenarios involving mannequins dressed in military uniforms. The study aimed to assess whether such AI models can support forensic analysts by automating parts of the visual identification process in large-scale digital investigations.

Method: This study applied quantitative and experimental methods, including the use of multimodal AI models (Google Gemini 1.5 Pro and the open-source LLAVA) for image analysis. A structured analysis was conducted using statistical performance metrics such as precision, recall, and accuracy. The sample consisted of 436 images divided into three categories: military personnel (198), military mannequins (99), and civilians (137), all extracted from an iOS backup to simulate real-world forensic conditions.

Findings: Both models demonstrated high precision (1.0) and strong recall (0.99 for Gemini, 0.98 for LLAVA) in detecting military presence. However, they struggled to differentiate between real individuals and mannequins—Gemini misclassified 88 out of 99 mannequin images, while LLAVA misclassified 86. Gemini significantly outperformed LLAVA in identifying contextual attributes such as country (0.7875 vs. 0.1218) and unit name (0.2544 vs. 0.0051).

Theoretical implications: This research expands the application of multimodal LLMs in digital forensics by demonstrating their

у сфері цифрової криміналістики. Визначено межі існуючих моделей у розпізнаванні реальних людей та манекенів, що вимагає перегляду підходів до семантичного аналізу зображень.

Практична цінність дослідження: Мультимодальні великі мовні моделі можуть бути ефективними інструментами допомоги судовим експертам при первинному аналізі великих масивів зображень. Забезпечують прискорене виявлення потенційно релевантних зображень у цифрових розслідуваннях.

Оригінальність: Перше дослідження, що цілеспрямовано аналізує ефективність мультимодальних великих мовних моделей для виявлення військовослужбовців на фото з мобільних пристроїв. Вперше проведено порівняльну оцінку Google Gemini 1.5 Pro та LLAVA у цьому контексті.

Обмеження дослідження: Моделі не можуть ефективно відрізнити манекенів від справжніх людей. Проблеми при обробці зображень з низькою якістю, слабким освітленням чи незвичайними позами. Майбутні дослідження мають зосередитися на покращенні контекстного розуміння та підвищенні стійкості моделей до помилкової класифікації.

Тип статті: Емпіричне дослідження.

Ключові слова: мультимодальні великі мовні моделі, штучний інтелект в криміналістиці, мобільна криміналістика, автоматичне розпізнавання зображень.

potential and limitations in semantically complex image recognition tasks. While the study did not challenge existing theories, it revealed the need for enhanced model architectures or training paradigms capable of better contextual interpretation in forensic scenarios.

Practical implications: The study highlights the practical potential of multimodal LLMs as auxiliary tools for forensic analysts, capable of rapidly identifying relevant images in large datasets. This can reduce manual workload, streamline investigative workflows, and improve initial triage in digital forensic investigations.

Value: This research is among the first to specifically explore the use of multimodal LLMs for detecting military personnel in images under real-world forensic conditions. The inclusion of mannequins as visual distractors adds unique value, uncovering significant model limitations and informing the development of more reliable AI tools for forensic applications.

Limitations: The findings are limited by the models' inability to differentiate between mannequins and real individuals, especially in cases of high visual realism. Future research should focus on improving robustness against such misclassifications, enhancing contextual reasoning, and extending evaluations to other object categories or threat identification use cases in digital forensics.

Paper type: Empirical study.

Key words: Multimodal Large Language Models (LLMs), Artificial Intelligence in Forensics, Mobile Forensics, Automated Image Recognition.

Вступ

Сфера цифрової криміналістики відіграє дедалі важливішу роль у сучасних розслідуваннях, що зумовлено експоненційним зростанням обсягів цифрових даних та зростаючою складністю кіберзлочинів [1, 2]. Мобільні пристрої стали справжнім джерелом доказової інформації, часто містячи великі обсяги зображень, що мають значення для кримінальних, контртерористичних та контррозвідувальних розслідувань. Величезний обсяг таких зображень становить серйозну проблему для криміналістів, яким доводиться вручну переглядати й аналізувати кожне з них з метою виявлення осіб, об’єктів та дій, що становлять інтерес [3, 4]. Цей ручний процес є не лише трудомістким і затратним за ресурсами, але й схильним до людських помилок, що може призвести до втрати важливих зачіпок або до хибних висновків.

Традиційні методи аналізу зображень у цифровій криміналістиці, такі як вилучення метаданих, аналіз відбитків пальців і розпізнавання обличь, використовуються для автоматизації окремих аспектів обробки зображень [5]. Однак ці методи часто не справляються зі складністю реальних зображень, зокрема через варіації освітлення, поз, часткові перекриття об’єктів, а також дедалі більш витончені методи маніпуляції зображеннями [6]. Крім того, завдання ідентифікації окремих категорій осіб, наприклад військовослужбовців, які можуть носити різноманітну форму або перебувати в цивільному середовищі, становить унікальну проблему, що потребує глибшого розуміння візуального контексту та семантичних взаємозв’язків. Додаткову складність для автоматизованої ідентифікації створює також наявність у публічному просторі експозицій і манекенів у військовій формі.

Останні досягнення в галузі штучного інтелекту (ШІ), зокрема у сфері великих мовних моделей (LLM) та мультимодальних LLM [7], відкривають перспективні нові підходи до подолання згаданих викликів [8, 9, 10]. LLM продемонстрували вражаючі можливості у розумінні та генеруванні людської мови, тоді як мультимодальні LLM розширюють ці можливості, інтегруючи візуальну інформацію. Це дозволяє їм виконувати завдання, пов’язані з описом зображень, відповідями на візуальні запитання та розумінням сцен.

Основні дослідницькі питання, розглянуті в цій роботі:

- Чи здатні мультимодальні LLM ефективно виявляти та розпізнавати військовослужбовців на зображеннях з мобільних пристроїв, навіть у випадках із реалістичними манекенами та різноманітним фоном?
- Якими є характеристики продуктивності (точність, повнота та загальна точність) моделей Gemini 1.5 Pro [11] та LLAVA [12] у межах цього конкретного криміналістичного завдання, зокрема щодо розрізнення справжніх військовослужбовців і манекенів, а також правильного визначення супутніх атрибутів, таких як країна походження, приналежність і підрозділ?

Які обмеження та виклики виникають під час використання цих LLM, і в якому напрямі варто проводити подальші дослідження?

Теоретичні основи дослідження

В цьому розділі розглянуто останні дослідження, що стосуються застосування мультимодальних великих мовних моделей (LLM) у сфері цифрової криміналістики, зокрема для ідентифікації військового персоналу на зображеннях. Ми аналізуємо попередні дослідження в традиційній цифровій криміналістиці зображень, використання штучного інтелекту для судово-криміналістичного аналізу та можливості мультимодальних LLM для розуміння зображень.

Традиційна цифрова криміналістика зображень охоплює широкий спектр методів, спрямованих на перевірку автентичності та цілісності цифрових зображень. Ці методи часто базуються на аналізі метаданих зображень, таких як дані EXIF, для виявлення невідповідностей або аномалій, що можуть свідчити про їхнє редагування [8]. Такі техніки, як ідентифікація моделі камери [13] і визначення джерела зображення на основі сенсорного шуму [14], розроблені для встановлення походження зображень і виявлення можливих підробок. Проте ці методи мають обмеження, оскільки покладаються на специфічні характеристики зображень і не здатні аналізувати їхній семантичний зміст. Крім того, як зазначено у [1], зростаюча складність інструментів для маніпуляції зображеннями ускладнює застосування традиційних методів для протидії новим загрозам.

Останнім часом зростає інтерес до використання штучного інтелекту (ШІ) у цифровій криміналістиці. Дослідники вивчають потенціал методів машинного навчання та глибокого навчання для автоматизації криміналістичних завдань і підвищення точності аналізу. Solanke та ін. [11] наголошують на необхідності оцінки та стандартизації методів криміналістичного аналізу цифрових доказів, керованих ШІ, з акцентом на створення надійних та ефективних підходів. Karthik і Shankar [2] розглядають, як можна інтегрувати ШІ та технологію блокчейн для покращення цифрових криміналістичних розслідувань, застосовуючи передові інструменти та алгоритми. Водночас ці методи на основі ШІ часто потребують великих обсягів навчальних даних і можуть стикатися з труднощами при обробці складних реальних криміналістичних сценаріїв.

Останні досягнення в галузі великих мовних моделей і мультимодальних LLM відкрили нові можливості для аналізу зображень. Ключовим досягненням стало створення моделей, здатних обробляти як візуальну, так і текстову інформацію, що дозволяє їм виконувати складні завдання, які раніше були недосяжні. Такі мультимодальні LLM проходять попереднє навчання на величезних наборах даних, що містять як зображення, так і текст, для формування глибоких семантичних представлень.

Одним із найбільш відомих прикладів є модель CLIP (Contrastive Language-Image Pre-training) [15], яка навчається за принципом передбачення відповідності зображень і текстових фрагментів. CLIP продемонструвала чудову здатність до перенесення знань у різні завдання комп’ютерного зору без додаткового навчання. Ще однією важливою моделлю є LLaVA (Large

Language and Vision Assistant) [12], яка поєднує візуальний енкодер із великою мовною моделлю та налаштовується на виконання мультимодальних інструкцій. LLaVA продемонструвала високі результати у мультимодальних діалогах та задачах візуального запитання-відповіді.

Оглядові дослідження, такі як роботи Yi et al. (2025) [16] і Chen et al. (2024) [17], детально аналізують швидко розвиваючу сферу мультимодальних LLM, зокрема їхні архітектури, методи навчання та застосування в різних галузях. Вони також обговорюють виклики та перспективи розвитку мультимодальних LLM. Хоча LLM уже продемонстрували значний потенціал у різних сферах, їхнє застосування у цифровій криміналістиці залишається малодослідженим. У [9] розглядається використання генеративного ШІ в криміналістичному аналізі даних, що показує його потенціал для підвищення точності та ефективності криміналістичних досліджень у хмарному середовищі. Проте застосування мультимодальних LLM для ідентифікації військового персоналу на криміналістичних зображеннях, особливо у випадках, коли присутні манекени, залишається малодослідженою темою.

Це дослідження спрямоване на подолання існуючої прогалини в літературі шляхом аналізу можливостей та ефективності використання комерційних і відкритих мультимодальних LLM для ідентифікації військового персоналу на криміналістичних зображеннях. На відміну від попередніх робіт, які зосереджувалися на традиційних методах аналізу зображень або загальному використанні ШІ у криміналістиці, це дослідження досліджує, як семантичне розуміння мультимодальних LLM може підвищити точність та ефективність виявлення військового персоналу. Крім того, наша робота включає аналіз складного набору даних, що містить зображення манекенів, які часто зустрічаються у реальних сценаріях, для оцінки стійкості LLM до таких факторів.

Постановка проблеми

Автоматизоване виявлення військовослужбовців на зображеннях з мобільних пристроїв є складним завданням у цифровій криміналістиці, особливо за наявності візуальних перешкод, наприклад, манекенів у військовій формі, що візуально імітують реальних осіб. Традиційні методи аналізу зображень не забезпечують достатньої точності в таких сценаріях. Це зумовлює необхідність дослідження можливостей сучасних мультимодальних великих мовних моделей для підвищення ефективності та достовірності автоматизованого аналізу.

Методологія дослідження

У цьому розділі описано методологію, використану для оцінки ефективності мультимодальних великих мовних моделей (LLM) у завданні ідентифікації військовослужбовців на зображеннях. Описано набір даних, деталі реалізації LLM, експериментальне середовище та метрики оцінки продуктивності.

Набір даних. Для проведення експериментів був зібраний набір даних з 436 зображень, що включав три різні категорії:

- Зображення військовослужбовців: Цей набір містив 198 зображень, на яких зображено військових у різних умовах, таких як навчальні вправи, паради та розгортання. Зображення включали широкий спектр форменого одягу, поз, умов освітлення та фону, щоб відобразити різноманітність, з якою стикаються в реальних судово-цифрових сценаріях.
- Зображення манекенів: Цей набір містив 99 зображень манекенів, одягнених у військову форму на військових виставках та в музеях. Ці зображення були включені для оцінки здатності LLM розрізняти реальних людей та неживі об’єкти, що є ключовою

задачею в цьому застосуванні. Манекени демонстрували реалістичні пози та деталі, що додатково ускладнювало завдання.

- Не військові зображення: Цей набір містив 137 зображень цивільних осіб в різних побутових умовах, що виконували роль негативної контрольної групи. Ці зображення дозволили переконатися, що LLM не ідентифікують цивільних осіб як військовослужбовців на основі загальних візуальних ознак.

Зображення були отримані з комбінації відкритих наборів даних та онлайн-репозиторіїв зображень. Для моделювання реалістичного цифрового судово-розслідувального сценарію всі зображення були надіслані через WhatsApp за допомогою iPhone. Потім зображення були витягнуті з резервної копії файлової системи iOS за допомогою програмного забезпечення iMazing [18]. Цей процес був обраний для імітації того, як зображення можуть бути відновлені під час реального мобільного розслідування. Детальніше інформація про процес отримання зображень із резервної копії iOS описана у нашого попередньому дослідженні [19].

Реалізація моделі. Було протестовано дві мультимодальні LLM:

- Gemini 1.5 Pro: Gemini 1.5 Pro — це передова мультимодальна LLM, розроблена Google AI. Вона здатна обробляти як зображення, так і текстові запити, що дозволяє виконувати складні завдання, такі як опис зображень, відповіді на візуальні запитання та розпізнавання об’єктів. Доступ до Gemini 1.5 Pro ми отримали через Google AI API, використовуючи стандартні налаштування для аналізу зображень.
- LLAVA: LLAVA — це мовна модель із відкритим кодом, що дозволяє користувачам запускати LLM локально. Для експерименту ми використали попередньо натреновану модель LLAVA, доступну з 01.03.2025 року [20]. Модель LLAVA працювала на локальній машині з Apple M2 Max та 64 ГБ оперативної пам'яті.

Для взаємодії з LLM ми використали Python та бібліотеку Langchain [21]. Langchain надав зручний інтерфейс для подачі зображень та текстових запитів до LLM та отримання їхніх відповідей. Запити були розроблені для отримання інформації про наявність військовослужбовців на зображеннях та відповідні атрибути.

Сценарій експерименту. Кожне зображення з набору даних оброблялося через такі етапи:

1. Подання зображення до LLM: Кожне зображення подавалося до моделей Gemini 1.5 Pro та LLAVA. Моделі отримували запит на детальний опис зображення, включаючи видимих осіб, об’єкти та сцени. Специфічні запити були сформульовані для отримання інформації про наявність військовослужбовців та відповідні атрибути, англійською: “Analyze the photo in detail and provide a comprehensive description. Include as much verifiable information as possible. Describe the scene, environment, and any notable objects. Determine if there are individuals who appear to be military personnel based on uniform, equipment, or insignia. If military chevrons, patches, or insignia are visible, identify and analyze them. If weapons, tactical gear, or other military equipment are present, describe them in detail and assess their possible origin and use. Affiliation and Context (if applicable). Identify any indications of country or organizational affiliation (flags, insignia, text, symbols, etc.). If unit or organization names are present, provide details on their role, historical background, and activities. Place findings in an operational or historical context if verifiable from the image. Important: Do not generate assumptions or fabricate information. Only analyze what is visible in the image and cross-check visual elements before drawing conclusions.”
2. Перетворення опису в формат JSON: Текстові описи, згенеровані моделями Gemini 1.5 Pro та LLAVA, були передані до моделі OpenAI o3-mini-2025-01-31, яка отримала

завдання перетворити вільні текстові описи в структурований формат JSON. Модель OpenAI була доступна через OpenAI API з використанням стандартних налаштувань.

3. Зберігання даних у таблиці CSV: Отримані JSON-об'єкти були витягнуті та збережені у таблиці CSV. Кожен рядок таблиці відповідав одному зображенню з набору даних, а стовпці містили ім'я файлу зображення, описи від Gemini 1.5 Pro та LLAVA, а також відповідні JSON-поля (включаючи військових, країну, ім'я тощо), що були отримані за допомогою моделі OpenAI. Ця структурована таблиця CSV стала основою для подальшого аналізу та оцінки продуктивності.

Результати були зареєстровані та проаналізовані для генерування метрик продуктивності, які описані в наступному розділі.

Результати

У цьому розділі наведено результати наших експериментів з оцінки ефективності моделей Gemini 1.5 Pro від Google та відкритої моделі LLAVA у завданні розпізнавання військовослужбовців на зображеннях. Ми подаємо як кількісні результати, засновані на метриках оцінювання, описаних у Розділі 3, так і якісні спостереження, отримані під час аналізу вихідних даних моделей. Далі ми обговорюємо значення отриманих результатів у контексті цифрової судової експертизи та можливі напрями для майбутніх досліджень. Кількісні результати. У таблиці 1 узагальнено загальну ефективність двох моделей на всьому наборі даних:

Таблиця 1 – Ефективність моделей

Метрика	Gemini 1.5 Pro	LLAVA
Точність	1.0	1.0
Повнота	0.99	0.98
Точність класифікації	0.768	0.764
Правильне визначення країни	0.7875	0.1218
Правильне визначення підрозділу	0.2544	0.0051

Як показано в таблиці 1, моделі Gemini 1.5 Pro та LLAVA досягли ідеальних показників точності, що дорівнюють 1.0. Це свідчить про те, що в усіх випадках, коли модель ідентифікувала зображення як таке, що містить військовослужбовця, ця класифікація була правильною. Показники повноти також були високими: 0.99 для Gemini 1.5 Pro та 0.98 для LLAVA, що вказує на здатність обох моделей виявляти майже всі зображення, які дійсно містили військовослужбовців. Водночас загальна точність класифікації була нижчою — 0.768 для Gemini 1.5 Pro та 0.764 для LLAVA. Ця невідповідність зумовлена насамперед помилковою класифікацією зображень із манекенами, що буде детальніше розглянуто нижче. Приклади зображень, на яких штучний інтелект правильно ідентифікував військову особу, наведені на рис. 1.

Моделі суттєво відрізнялися за здатністю правильно визначати супровідні атрибути військовослужбовців. Gemini 1.5 Pro продемонструвала точність 0.7875 за метрикою правильне визначення країни та 0.2544 за метрикою правильне визначення підрозділу.

Рисунок 1 – Приклади зображення, на якому моделі правильно ідентифікували військову особу

Показники моделі LLAVA за цими метриками були значно нижчими — 0.1218 та 0.0051 відповідно. Це свідчить про те, що Gemini 1.5 Pro краще розуміє військовий контекст і здатна ефективніше витягувати релевантну інформацію із зображень. Приклади зображень, на яких ШІ правильно ідентифікував військову символіку, нашивки або прапори, наведено на рис. 2.

Рисунок 2 – Приклади зображення, на якому моделі правильно ідентифікували емблему військового підрозділу, нашивки або прапори

Кількість випадків помилкової класифікації додатково підкреслює відмінності між моделями: LLAVA помилково класифікувала 193 зображення, тоді як Gemini 1.5 Pro — 128.

Помилкова класифікація манекенів. Суттєвим викликом, з яким ми зіткнулися під час експериментів, стала помилкова класифікація зображень із манекенами як таких, що містять військовослужбовців. Модель Gemini 1.5 Pro неправильно класифікувала 88 із 99 зображень манекенів, тоді як LLAVA — 86. Це свідчить про те, що обидві моделі мають труднощі з розрізненням реальних людей і манекенів, особливо коли манекени одягнені в реалістичну військову форму. Приклади хибно-позитивних результатів — манекенів, помилково класифікованих як військовослужбовці, наведено на рис. 3.

Рисунок 3 – Приклади хибно-позитивного результату — манекени, помилково класифіковані як військовослужбовці

Якісний аналіз результатів, отриманих від моделей, показав, що великі мовні моделі часто зосереджувалися на уніформі та спорядженні, які носили манекени, не враховуючи інші ознаки — такі як риси обличчя, текстура шкіри чи мова тіла — які могли б вказувати на те, що перед ними не справжня людина.

Якісний аналіз. Окрім кількісних результатів, ми також провели якісний аналіз вихідних даних моделей з метою кращого розуміння їхніх сильних і слабких сторін. Було виявлено, що Gemini 1.5 Pro загалом точніше визначала країну походження та приналежність військовослужбовців, тоді як LLAVA часто надавала загальні або некоректні відповіді. Наприклад, при аналізі зображення військовослужбовця армії США, Gemini 1.5 Pro правильно ідентифікувала країну як Сполучені Штати та приналежність як Армію США, тоді як LLAVA лише узагальнено визначила особу як “військовослужбовця”.

Однак обом моделям було складно обробляти зображення з частковими перекриттями, поганим освітленням або нетиповими позами. У таких випадках моделі часто

не змогли ідентифікувати військовослужбовця або надавали неточні описи. Крім того, обидві моделі виявляли схильність до надмірного акцентування уваги на наявності зброї чи військового спорядження, навіть якщо ці елементи не були основним об'єктом на зображенні.

Обговорення

Результати наших експериментів демонструють потенціал мультимодальних великих мовних моделей (LLM) у завданні ідентифікації військовослужбовців на зображеннях, але водночас висвітлюють низку ключових викликів. Високі показники точності, досягнуті як Gemini 1.5 Pro, так і LLAVA, свідчать про те, що ці моделі можуть стати корисними інструментами для попереднього перегляду великих наборів зображень у рамках судових розслідувань, дозволяючи оперативно виявляти зображення, які ймовірно містять релевантних осіб. Водночас нижчі показники загальної точності та висока частота помилкової класифікації манекенів свідчать про те, що ці моделі наразі не готові до повністю автоматизованого використання.

Кращі результати Gemini 1.5 Pro у визначенні супровідних атрибутів свідчать про її глибше розуміння військового контексту порівняно з LLAVA. Це може бути зумовлено відмінностями у навчальних даних або архітектурі моделі. Втім, обидві моделі продемонстрували обмеження у здатності працювати з зображеннями в складних умовах та розрізняти реальних людей і манекенів.

Отримані результати мають низку важливих наслідків для цифрової судової експертизи. По-перше, вони свідчать про те, що мультимодальні великі мовні моделі можуть бути корисним інструментом для підтримки судових аналітиків у завданні ідентифікації військовослужбовців на зображеннях, однак не повинні розглядатися як єдине джерело інформації. По-друге, результати підкреслюють важливість ретельної оцінки ефективності LLM у конкретних судово-експертних контекстах, оскільки їх продуктивність може суттєво варіюватися залежно від набору даних і характеру завдання. По-третє, виявлено кілька ключових напрямів для майбутніх досліджень, зокрема покращення здатності LLM працювати в складних умовах обробки зображень та розрізняти реальних людей і неживі об'єкти.

Висновки

У цьому дослідженні було проаналізовано ефективність застосування мультимодальних великих мовних моделей Gemini 1.5 Pro та LLAVA для виявлення військовослужбовців на зображеннях з мобільних пристроїв. Отримані результати засвідчили високі показники точності та повноти для обох моделей у базовому розпізнаванні військових, але також виявили серйозні труднощі з розрізненням реальних осіб і манекенів. Модель Gemini продемонструвала суттєво кращу здатність до визначення контекстних атрибутів, таких як країна або підрозділ, порівняно з LLAVA. Водночас, обидві моделі потребують подальшого вдосконалення для зменшення кількості хибнопозитивних результатів та покращення роботи в складних візуальних умовах. Результати дослідження свідчать про перспективність використання мультимодальних LLMs як допоміжного інструменту в цифровій криміналістиці. Подальші дослідження мають бути спрямовані на підвищення стійкості моделей до візуальних імітацій та розширення їхніх контекстуальних можливостей.

Фінансування

Це дослідження не отримало конкретної фінансової підтримки.

Конкуруючі інтереси

Автори заявляють, що у них немає конкуруючих інтересів.

Список використаних джерел

1. Zangana, Hewa Majeed, and Marwan Omar. "Introduction to Digital Forensics and Artificial Intelligence." Digital Forensics in the Age of AI, edited by Marwan Omar and Hewa Majeed Zangana, IGI Global, 2025, pp. 1-30. https://doi.org/10.4018/979-8-3373-0857-9.ch001
2. Karthikeyan, P., Pande, H.M., & Sarveshwaran, V. (Eds.). (2023). Artificial Intelligence and Blockchain in Digital Forensics (1st ed.). River Publishers. https://doi.org/10.1201/9781003374671
3. "AI IN DIGITAL FORENSICS", IJSRMST, vol. 3, no. 5, pp. 01–06, May 2024, https://doi.org/10.59828/ijsrmst.v3i5.208.
4. Moses Ashawa, Ali Mansour, Jackie Riley, Jude Osamor, Nsikak Pius Owoh. Digital Forensics Challenges in Cyberspace: Overcoming Legitimacy and Privacy Issues Through Modularisation. Cloud Computing and Data Science [Internet]. 2023 Dec. 25];5(1):140-56. https://doi.org/10.37256/ccds.5120233845.
5. Mishra, Pallavi. (2020). Big Data Digital Forensic and Cybersecurity. https://doi.org/10.1201/9781003024743-9.
6. Javed, Abdul Rehman & Jalil, Zunera & Zehra, Wisha & Gadekallu, Thippa & Suh, Doug & Jalil Piran, Md. (2021). A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions. Engineering Applications of Artificial Intelligence. https://doi.org/10.1016/j.engappai.2021.104456.
7. Hao Tan and Mohit Bansal. 2019. LXMERT: Learning Cross-Modality Encoder Representations from Transformers. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 5100–5111, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1514.
8. Moustafa, N. (2022). Digital Forensics in the Era of Artificial Intelligence (1st ed.). CRC Press. https://doi.org/10.1201/9781003278962.
9. Emehin, Oluwatobi & Emeteveke, Isaac & Adeyeye, Oladele & Akanbi, Ibrahim. (2024). Generative AI in Forensic Data Analysis: Opportunities and Ethical Implications for Cloud-Based Investigations. *International Journal of Research Publication and Reviews*. 6. 2941-2957. https://doi.org/10.55248/gengpi.5.1024.2904.
10. Solanke, Abiodun & Biasiotti, Maria. (2022). Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques. KI - Künstliche Intelligenz. https://doi.org/10.1007/s13218-022-00763-9.
11. Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati et al, Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, 2024, https://doi.org/10.48550/arXiv.2403.05530.
12. Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee, Visual Instruction Tuning, 2023, https://doi.org/10.48550/arXiv.2304.08485.
13. Kirchner, Matthias & Gloe, Thomas. (2015). Forensic Camera Model Identification. https://doi.org/10.1002/9781118705773.ch9.
14. Filler, Tomás & Fridrich, Jessica & Goljan, Miroslav. (2008). Using sensor pattern noise for camera model identification. Proceedings - International Conference on Image Processing, ICIP. 1296-1299. https://doi.org/10.1109/ICIP.2008.4712000.
15. Radford, Alec & Kim, Jong & Hallacy, Chris & Ramesh, Aditya & Goh, Gabriel & Agarwal, Sandhini & Sastry, Girish & Askell, Amanda & Mishkin, Pamela & Clark, Jack & Krueger, Gretchen & Sutskever, Ilya. (2021). Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020.

16. Yi, Z.; Xiao, T.; Albert, M.V. A Survey on Multimodal Large Language Models in Radiology for Report Generation and Visual Question Answering. Information 2025, 16, 136. https://doi.org/10.3390/info16020136.
17. He, Yingqing & Liu, Zhaoyang & Chen, Jingye & Zeyue, Tian & Liu, Hongyu & Chi, Xiaowei & Liu, Runtao & Yuan, Ruibin & Xing, Yazhou & Wang, Wenhai & Dai, Jifeng & Zhang, Yong & Xue, Wei & Liu, Qifeng & Guo, Yike & Chen, Qifeng. (2024). LLMs Meet Multimodal Generation and Editing: A Survey. https://doi.org/10.48550/arXiv.2405.19334.
18. iMazing, 2025. [Online]. Retrieved from : https://imazing.com/.
19. Mykhaylova, O. et al., Person-of-Interest Detection onMobile Forensics Data—AI-Driven Roadmap, in:Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654 (2024) 239–251.
20. LLaVA: Large Language and Vision Assistant, 2025. [Online]. Retrieved from : https://llava-vl.github.io/.
21. Langchain, 2025. [Online]. Retrieved from : https://www.langchain.com/.

References

1. Zangana, Hewa Majeed, and Marwan Omar. "Introduction to Digital Forensics and Artificial Intelligence." Digital Forensics in the Age of AI, edited by Marwan Omar and Hewa Majeed Zangana, IGI Global, 2025, pp. 1-30. https://doi.org/10.4018/979-8-3373-0857-9.ch001
2. Karthikeyan, P., Pande, H.M., & Sarveshwaran, V. (Eds.). (2023). Artificial Intelligence and Blockchain in Digital Forensics (1st ed.). River Publishers. https://doi.org/10.1201/9781003374671
3. "AI IN DIGITAL FORENSICS", IJSRMST, vol. 3, no. 5, pp. 01–06, May 2024, https://doi.org/10.59828/ijsrmst.v3i5.208.
4. Moses Ashawa, Ali Mansour, Jackie Riley, Jude Osamor, Nsikak Pius Owoh. Digital Forensics Challenges in Cyberspace: Overcoming Legitimacy and Privacy Issues Through Modularisation. Cloud Computing and Data Science [Internet]. 2023 Dec. 25];5(1):140-56. https://doi.org/10.37256/ccds.5120233845.
5. Mishra, Pallavi. (2020). Big Data Digital Forensic and Cybersecurity. https://doi.org/10.1201/9781003024743-9.
6. Javed, Abdul Rehman & Jalil, Zunera & Zehra, Wisha & Gadekallu, Thippa & Suh, Doug & Jalil Piran, Md. (2021). A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions. Engineering Applications of Artificial Intelligence. https://doi.org/10.1016/j.engappai.2021.104456.
7. Hao Tan and Mohit Bansal. 2019. LXMERT: Learning Cross-Modality Encoder Representations from Transformers. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 5100–5111, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1514.
8. Moustafa, N. (2022). Digital Forensics in the Era of Artificial Intelligence (1st ed.). CRC Press. https://doi.org/10.1201/9781003278962.
9. Emehin, Oluwatobi & Emeteveke, Isaac & Adeyeye, Oladele & Akanbi, Ibrahim. (2024). Generative AI in Forensic Data Analysis: Opportunities and Ethical Implications for Cloud-Based Investigations. *International Journal of Research Publication and Reviews*. 6. 2941-2957. https://doi.org/10.55248/gengpi.5.1024.2904.
10. Solanke, Abiodun & Biasiotti, Maria. (2022). Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques. KI – Künstliche Intelligenz. https://doi.org/10.1007/s13218-022-00763-9.

11. Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati et al, Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, 2024, https://doi.org/10.48550/arXiv.2403.05530.
12. Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee, Visual Instruction Tuning, 2023, https://doi.org/10.48550/arXiv.2304.08485.
13. Kirchner, Matthias & Gloe, Thomas. (2015). Forensic Camera Model Identification. https://doi.org/10.1002/9781118705773.ch9.
14. Filler, Tomás & Fridrich, Jessica & Goljan, Miroslav. (2008). Using sensor pattern noise for camera model identification. Proceedings - International Conference on Image Processing, ICIP. 1296-1299. https://doi.org/10.1109/ICIP.2008.4712000.
15. Radford, Alec & Kim, Jong & Hallacy, Chris & Ramesh, Aditya & Goh, Gabriel & Agarwal, Sandhini & Sastry, Girish & Askell, Amanda & Mishkin, Pamela & Clark, Jack & Krueger, Gretchen & Sutskever, Ilya. (2021). Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020.
16. Yi, Z.; Xiao, T.; Albert, M.V. A Survey on Multimodal Large Language Models in Radiology for Report Generation and Visual Question Answering. Information 2025, 16, 136. https://doi.org/10.3390/info16020136.
17. He, Yingqing & Liu, Zhaoyang & Chen, Jingye & Zeyue, Tian & Liu, Hongyu & Chi, Xiaowei & Liu, Runtao & Yuan, Ruibin & Xing, Yazhou & Wang, Wenhai & Dai, Jifeng & Zhang, Yong & Xue, Wei & Liu, Qifeng & Guo, Yike & Chen, Qifeng. (2024). LLMs Meet Multimodal Generation and Editing: A Survey. https://doi.org/10.48550/arXiv.2405.19334.
18. iMazing, 2025. [Online]. Retrieved from : https://imazing.com/.
19. Mykhaylova, O. et al., Person-of-Interest Detection onMobile Forensics Data—AI-Driven Roadmap, in:Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654 (2024) 239–251.
20. LLaVA: Large Language and Vision Assistant, 2025. [Online]. Retrieved from : https://llava-vl.github.io/.
21. Langchain, 2025. [Online]. Retrieved from : https://www.langchain.com/.