Designing a Trustworthy and Private LLM Architecture for Cross Platform Forensic Investigations

*Gauri Mathur
**Avni Jain
***Dr. Radha Krishna Rambola

Abstract

In order to undertake research on cybercrimes, the digital forensics will be necessary. It entails the study of data of various gadgets such as computers, cell phones, and cloud-based servers. Nevertheless, traditional techniques tend to be labor-intensive and/or time consuming and are not able to match the rising frequency or variety of digital evidence. Alternatives can be promising and are through ChatGPT and other large language models (LLMs). Detecting anomalies and text analysis are some of the tasks that a person can automate. However, many researchers identify a variety of roadblocks to the incorporation of such tools in forensic practice, such as the lack of standardized data sources, the lack of explainability, problems with cross-device management, and privacy threats. The given research proposes a secure and explainable LLM-based forensic model. The objectives of the model include aggregating evidence based on different sources to safeguard the privacy of the data, provide clearly understandable results, and establish specific standards of forensic analysis. Its first goal is to enhance the performance, reliability, and admissibility of digital inquiries conducted with the help of AI.

1. INTRODUCTION

1. OVERVIEW

Digital forensics is a precious area of cybersecurity and police due to the increased complexity and the rise in the frequency of computer crimes. It involves locating, extracting, preserving, and processing data on the computer systems to support criminal proceedings and lawsuit. Since it is shifting towards cloud storage and the domain of Internet of Things (IoT), we will need more scalable and quicker solutions.

Meanwhile, Large Language Models (LLMs), such as GPT-4 by OpenAI, have transformed the field of natural language processing (NLP). These enable highly fluent machines to comprehend, produce and process human language. They are also experimented in a number of fields such as discovery forensics in the context to be applied in techniques of automation, summarization, and anomaly detection [5].

However, this prompts even more problems when LLMs are implemented in the forensic data pathway. The majority of the prevailing systems are underdeveloped at processing information driven by various devices, lack transparency and follow privacy principles. Moreover, faith-based testing and use are limited by the absence of standardised data and metrics of performance prepared specifically to apply to forensics.

This study offers a Secure and Explainable LLM-Based Framework on Digital Forensics as a way of addressing these issues. The model attempts to provide a local choice to the privacy, merge data between different devices, offer a transparent infrastructure in explaining the outcome, and new standards in gauging accuracy and credibility [1]. It is meant to quicken the investigation process without compromising technical and legal soundness.

2. BACKGROUND AND EVOLUTION

2.1 Overview of Digital Forensics

From the recovery and analysis of digital evidence from electronic sources, digital forensics has grown into an interdisciplinary process that supports legal investigations. It includes subjects such as network forensics, cloud forensics, mobile forensics, and computer forensics. Traditional methods often involve rule-based investigation and manual tools, which are time-consuming and susceptible to human error, especially when dealing with large or unstructured data sets.

Digital forensics has to get faster and more automated to keep up with increasing instances of ransomware attacks, data breaches, and cyber fraud. This has, therefore, created increased interest in using artificial intelligence (AI) to support forensic work.

2.2 Development of LLM's in Digital Forensics

Digital forensics AI first focused on dedicated tasks such as file categorization or machine learning-driven anomaly detection. These models required extensive domain-specific fine-tuning and tended to have limited scope. It was enabled by the emergence of transformer-based models, specifically LLMs such as BERT and GPT, to introduce more language understanding abilities, background knowledge extraction, and summarization of large-scale datasets [3].

For automating functions such as timeline analysis, chat log summary, detecting unusual communication behavior, and aiding forensic reporting, LLMs are now being researched for application in digital forensics [4]. They are not employed currently though due to issues such as generation of fake data, intractability, and privacy and admissibility in court.

3. ADDRESSING THE GAPS IN CURRENT RESEARCH

While applying Large Language Models (LLMs) to digital forensics holds tremendous promise, the available research and real-world experiments to date indicate that there are several issues which remain to be resolved. The practical forensic application suffers from these shortcomings in scalability, reliability, as well as legal acceptability.

3.1 Shortage of Standardized Forensic Datasets

The key obstacle is represented by the absence of publicly disclosed forensics specific information on LLM testing. The majority of the standards, such as BLEU and ROUGE, pay much attention to the similarity in language, but not to the forensic relevance or the accuracy of facts. In absence of standardized data and assessment protocols, validating the correctness, and performance of forensic outputs generated by LLM is hard [2].

3.2 Privacy and security issues

Chain-of-custody and data privacy are threatening with most LLMs based on cloud-based APIs [6]. The transfer of forensic sensitive data to third-party servers is done most of the time in violation of moral and legal protocols. LLMs that are deployable locally and whose data maintains integrity and privacy regulation.

3.3 Shortcomings of multi-device data integration

Digital evidence is frequently provided by a number of devices and platforms, including smartphones, PCs, cloud, and IoT devices. The majority of the currently deployed LLM applications is a single-input model that is weak when it comes to correlating events that may have occurred in varied sources. This restrains their capabilities to recreate timelines or spot concerted efforts in complex inquiry.

4. PROPOSED SOLUTION: SECURE AND EXPLAINABLE FORENSIC FRAMEWORK

To fill the gap of the existing methods, the proposed research develops the Secure and Explainable LLM-Based Forensic Framework. Facilitating trust and transparency by conducting cross-device evidence analysis, the goal is to present a scalable, privacy-aware, and admissible solution to capitalize on the potential of Large Language Models (LLMs).

4.1 Overview of the Framework

The suggested framework is a locally implementable, modular solution that will help forensic examiners to analyze huge volumes of digital evidence. The framework incorporates both forensic-specific enhancements and beyond-state-of-the-art LLM capabilities to process input data coming in through multiple sources and generate explainable output. Data privacy, correlation among the events, transparency in the use of the model as well as its performance to consider.

4.2 Key Characteristics of the Framework

Multi-Device Evidence Correlation: The system is also capable of being connected to information (i.e. computer, mobile, cloud logs, and external devices sources). Investigators can find event sequences and patterns in data irrespective of platforms using LLM-driven timeline reconstruction module.

Local Deployment and Privacy Control: The system operates diligently off line following the forensic processes of handling the data. This supports the chain of custody, which is a very important requirement in legal matters in case any criminal investigations are exerted and secures the privacy of the data by avoiding them being consulted by other outside servers.

Explainable AI Module: The system will have a layer of explainability built in that can give clear outputs in terms that a person can comprehend. These are the levels of confidence, reasons behind conclusions and citations to the particular pieces of evidence. This kind of transparency will increase the confidence and strengthen the applicability of the results generated by AI to the legal landscape.

Standard Dataset and Evaluation Protocol: An entirely new standard dataset that is comprised of forensic scenarios artificially generated is proposed. Using this dataset, the framework compares the performance of LLMs in terms of forensic metrics, including fact accuracy, evidence traceability and anomaly detection precision in conjunction with more general NLP metrics like BLEU and ROUGE.

5. SYSTEM COMPONENTS AND ARCHITECTURE

Secure and Explainable LLM-Based Forensic Framework is a modular framework with interconnected subcomponents, which accommodates the peculiarities of digital forensic research, in particular, concerning the concept of multi-device integration, explainability, scalability, privacy. The crucial structures and their purposes are provided in this section.

5.1 Explainability Module

Among the largest obstacles to the application of the LLMs in digital forensics, there are issues of transparency on LLMs. To fix this issue, the Explainability Module produces explanations that human readers can understand each output of this system. Upon identifying a suspicious communication, e.g. the model will give the exact wording or pattern which brought on the detection, and correlate the same with the initial source of the evidence. This module contains links to evidence traceability, explanation of the reasoning behind it in plain language, confidence scoring. Such transparency increases the level of legal admissibility and results in the investigator confidence in discoveries made with the aid of AI.

5.2 Privacy and Compliance Layer

To ensure no cloud environment or external services are exposed to, the system can be used completely offline since the information of digital evidence is delicate. The layer maintains chain of custody throughout the period of investigation. One of them is the characteristic of local implementation of LLM and on-premise resources management.

- Encryption of data at Rest and Data encryption in Transit.
- Compliance checkers of GDPR, HIPAA and jurisdiction-dependent forensic standards.

This architecture ensures the security of sensitive data, besides, it is possible to use the tool in high-security environments like defense and law enforcement.

5.3 Multi-Device Evidence Aggregator

Digital crimes involve many gadgets and platforms. Computation of structured and unstructured data of laptops and mobile phones, cloud logs, and other digital sources are on the Multi-Device Evidence Aggregator.

- Standardization of the artifacts for these operating systems and the file type.
- Establishment of an accurate capture of the chronological sequence of events by matching the events.

This factor allows cross-platform forensic reconstruction and searching trends and correlations between the activities of various devices.

5.4 Evaluation Engine and Dataset

Forensic-specific standard database which includes created simulated investigation events, logs and anomalies is also proposed to provide an objective measure. Evaluation Engine makes the use of this database to evaluate performance of the framework using the following classical measures of NLP: BLEU, ROUGE. Factors that are specific to forensics:

- Accuracy of factual statements.
- Evidence coverage.
- Timeline coherence.
- A deviance-spotting.

This aspect will remain that the system can be equated with the changing industry standards and constantly upgraded.

6. CASE STUDY- EXPERT SURVEY ON LLM ADOPTION IN DIGITAL FORENSICS

A survey comprising of 22 government and state forensic science laboratory officers spread across India was conducted to investigate the feasibility and applicability of proposed secure and explainable based on LLM forensic system. The respondents were more than 85 percent above ten years of experience in forensic science comprising of Assistant Directors, Senior Scientific Officers, Laboratory Scientists, and Crime Scene Investigators.

6.1 Key Points

Knowledge regarding LLMs: As shown in Figure 1, more than 90 percent of the respondents were familiar with such LLMs as ChatGPT, BERT, and GPT-4. Though 13.6 of the respondents asserted that it has a fully-manual process, over 80 percent assert partial automation of processes demonstrating their readiness to intelligent automation tools LLMs. Nevertheless, the majority of the respondents (77 percent) noted that they did not use AI or NLP tools when handling forensic cases, which suggests more integration is possible.

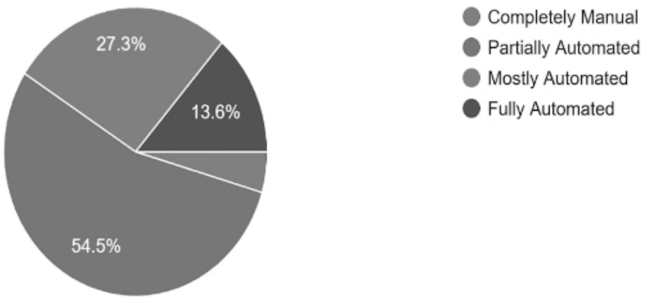

Figure 1: Expert Familiarity with LLMs and AI Tools

Issues with Investigations: Cross-device correlation, large amounts of data (31.8 percent) and limitations of tools (45.5 percent) were found to have been the most frequent concerns. The other highlighted major issues were legal compliance and report production. Figure 2 represents the concerns around this.

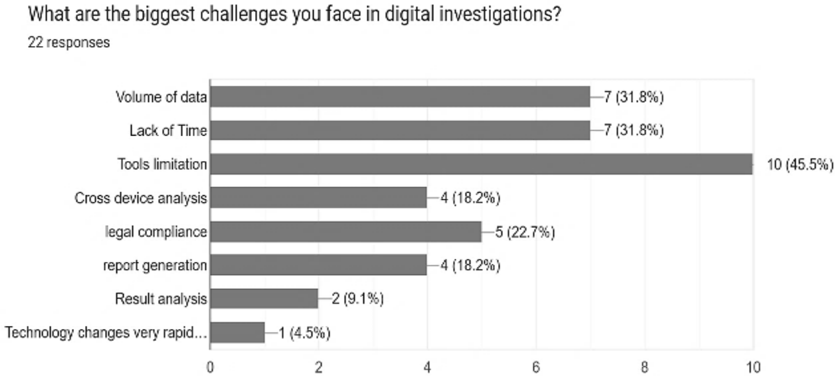

Figure 2: Concerns Around the Use of LLMs in Forensics

Visualization Insights: According to Figure 3,4 and 5, over three-quarters of experts identified explainability of AI forensic tools as being of importance, of great importance, or critical and 59.1 percent experts chose offline LLM system due to privacy and integrity concerns. In addition, 91 percent of the professionals felt that LLMs could certainly or possibly make digital investigations faster. It means that there would be high demand in well-timed, open, secure, and run locally AI tools complying with forensic validation standards.

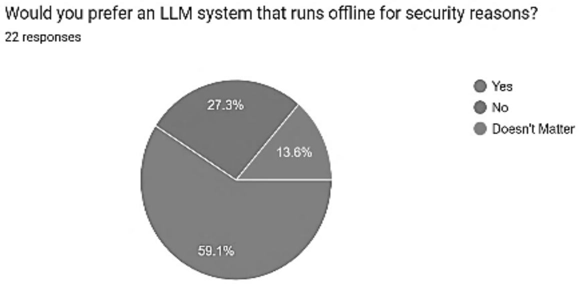

Figure 3: Preferred Deployment Mode for LLMs in Forensics

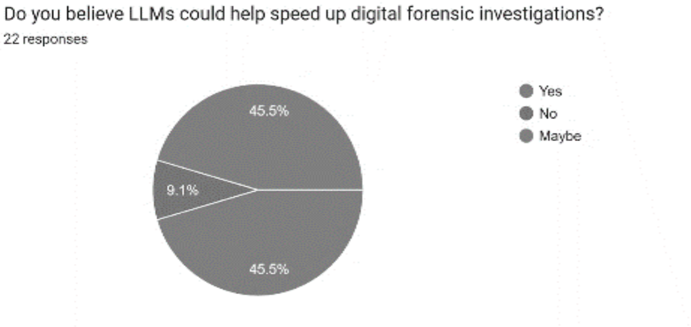

Figure 4: Confidence Level in AI-Generated Forensic Outputs

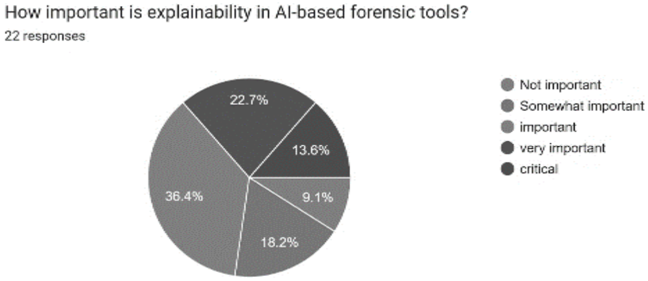

Figure 5: Importance of Explainability in AI-Based Forensic Tools

The most relevant issues related to applying LLMs include: As Figure 6 shows, the most pressing issues concerned the legality of using LLMs and the privacy of the data (63.6 and 59.1 percent, respectively), whereas the second stage of concerns related to the reliability of AI output (40.9 percent). The lesser but still highly prevalent issues were lack of technical skills, implementation costs, and the handling of multimedia evidence that is complex. The above results highlight the urgency of the issue of the availability of secure, interpretable, and legally acceptable AI solutions in forensics.

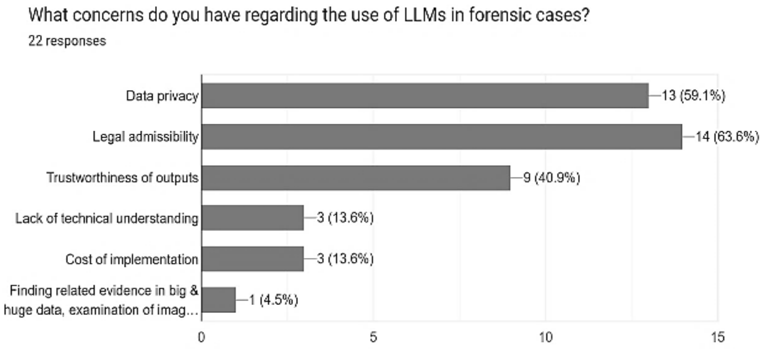

Figure 6: Concerns Around the Use of LLMs in Forensics

7. LIMITATIONS AND CHALLENGES

Large Language Models (LLMs) have enormous potential in digital forensics, but this category of tools is not devoid of limitations and challenges that will have to be met prior to a functional, and legally successful implementation:

7.1 Legal Admissibility of AI-Generated Evidence

The first issue that forensic professionals are concerned about (63.6%) is, in turn, whether the results found by AI can be proven in the court. Unless standards and certification in forensics are put in place, the LLM outputs may be rejected by the courts and thus their use will be limited to formal investigations.

7.2 Offline Requirements and Data Privacy

Many of the respondents (59.1%) stated the need of privacy of the data and supported the procedure of offline deployment to maintain the chain of custody and avoid patronizing data leakage. This is a technical challenge to the LLM developers, who normally rely on cloud processing.

7.3 Explainability and Trust in AI Results

The need to have Explainable AI also emerges due to the high share of respondents who did not trust how trustworthy the results provided by LLM are. Transparency and credibility requires forensic experts to be able to trace and verify conclusions arrived at by AI.

7.4 With the costs curbed, infrastructure remained as a concern.

Some of the participants cited limited resources in computing and implementation costs particularly in government laboratories. Edge-friendly options or light-weighted versions might be the necessities when considering LLMs to be applied in the local settings, the settings with limited resources.

8. CONCLUSION AND FUTURE WORK

8.1 Final Thoughts

The paper explored using an explainable, secure and locally deployed AI framework to adopt the use of Large Language Models (LLMs) on digital forensics analysis. This was triggered by the complexities and growing volume of evidence in a digital form and inadequacy of existing forensic tools.

The participation in a professional poll by 22 government laboratory forensic specialists showed that the experts were educated fairly well on LLMs and cautiously optimistic about the potential usefulness of them. One of the key concerns, including admissibility in the court, the protection of the privacy of the data, and an accountable AI output were required, and they were all addressed by the proposed framework explicitly. Offline deployment and results that are human-readable were also mentioned by the professionals to assure their source and promote faith in investigation and the court.

The successful integration of LLMs is subject to tackling ethical, technical, and legal issues despite providing significant opportunities in terms of automation of the evidence analysis, reconstitution of the timeline, and cross-device correlation. The remarks determine that forensic-conscious, subject-specific LLM system is needed to expand human knowledge without abating legal terms or the chain of custody.

8.2 Future Work

The construction of a locally deployable LLM- that has data isolation and explainability-based forensic assistant, and it is known as a prototype development.

Standardized Forensic Dataset Generation A collection of differing and diverse datasets of anonymized forensic cases to examine the accuracy, degree of bias, and reliability of LLMs on forensic applications.

Some explainability extensions include having traceable chains of logic, chain of evidence, and generation of rationale to accommodate court requirements to explain and justify.

Legal and Ethical Systems: Using legislators and lawyers, particular rules were being created concerning the verification and admissibility of the forensic products produced using AI.

***B.Tech Computer Science**
****B.Tech Computer Science**
*****Computer Science**
NMIMS, Shirpur, India

9. REFERENCES

[1] Mittal, B., Pahwa, N., & Sharma, T. (2023). A Analysis of the Role of Artificial Intelligence and Machine Learning in Digital Forensics. International Journal of Computer Applications.

[2] Manogaran, G., & Thota, C. (2024). Towards a methodology and dataset for testing LLM-based digital forensic tools. Forensic Science International: Digital Investigation, Elsevier.

[3] Mittal, B., & Walia, E. (2023). Challenges and Applications of AI in Digital Forensics: A Review. In Proceedings of the International Conference on Computing, Communication and Networking Technologies (ICCCNT).

[4] Adedayo, W., & Akintade, O. (2022). Artificial Intelligence in Digital Forensics: Challenges and Opportunities. Journal of Cybersecurity and Digital Forensics, 5(2), 15–27.

[5] Li, Z., & Li, T. (2023). Explainable AI in Criminal Justice and Digital Evidence Processing. ACM Digital Threats: Research and Practice, 4(1), 1–19.

[6] Alsmadi, I., & Alharkan, I. (2021). the Role of NLP and LLMs in Cybercrime Detection and Forensic Reporting. Computers & Security, 105, 102247.