Agile Proactive Cybercrime Evidence Analysis Model for Digital Forensics

Mohammad Al-Mousa
Department of Cybersecurity
Zarqa University, Jordan
mmousa@zu.edu.jo

Waleed Amer
Department of Cybersecurity
Zarqa University, Jordan
wamer.std@zu.edu.jo

Mosleh Abualhaj
Department of Networks and Cybersecurity
Al-Ahliyya Amman University, Jordan
m.abualhaj@ammanu.edu.jo

Sultan Albilasi
Department of Cybersecurity
Zarqa University, Jordan
Ss.m.b.2010@hotmail.com

Ola Nasir
Department of Computer Science
Zarqa University, Jordan
onasir@zu.edu.jo

Ghassan Samara
Department of Computer Science
Zarqa University, Jordan
gsamara@zu.edu.jo

Abstract: *Digital forensics is a critically important area of study dealing with the identification and combating of cyber threats in contemporary networked environments. In this paper, we investigate the possibility of utilizing Large Language Models (LLMs) to examine network traffic categorized as risky according to the University of New South Wales-Network-Based 2015 (UNSW-NB15) dataset. The study employs a multi-phase methodology that combines forensic analysis, evidence extraction, security recommendations, contextual evaluation, and detailed reporting. The results demonstrate high accuracy and qualitative performance across tasks. Automated metrics illustrate the forensic analysis with 95% accuracy, and evidence extraction with 94% precision and 95% coverage. Subjective self-assessment, followed by reviewing 100 examples processed through ChatGPT, shows that outputs have a very high level of clarity (5 out of 5) and relevance (4.5 out of 5). These results highlight the revolutionary role of LLMs in digital forensics with respect to precision, scope, and readability*.

Keywords: *Cybercrime evidence, digital forensics, proactive analysis, agile strategy, LLMs*.

Received January 16, 2025; accepted April 21, 2025
https://doi.org/10.34028/iajit/22/3/15

1. Introduction

Today, there is nothing more dangerous in the world of cybersecurity than cybercrime, which poses a significant threat to individuals, enterprises, and authorities [1]. Over the years, the technologies in use, as well as the intricacies of cybercriminal activity, have advanced. Therefore, identifying and interpreting electronic data is a task for the discipline of digital forensics, which has gained greater importance [18]. Previous techniques in digital forensics have been largely reactive, relying on systems and processes that often result in lengthy detection and analysis of cybercrimes [20]. This is a critical downside; as timely action could help prevent or limit damage or loss [11].

Agile is a proactive and adaptable strategy for addressing cybercrime, employing Agile principles to formulate dynamic solutions that can swiftly adjust to evolving cyber threats. This paradigm is grounded in ongoing surveillance, iterative assessment, and prompt reaction to prospective assaults prior to their manifestation or during their incipient phases. Given the emergence of sophisticated cyberattacks, there is an urgent need for nimble, proactive instruments and methodologies within the realm of digital forensics [16]. The incorporation of artificial intelligence, particularly Large Language Models (LLMs), into digital forensic practices have the potential to fundamentally transform the discipline [6]. LLMs, such as those developed by OpenAI, are engineered to comprehend and generate text akin to human language [4], rendering them exceptionally equipped for the analysis of intricate data patterns and the rapid generation of actionable insights [15]. This paper discusses the possibility of incorporating LLMs into an Agile proactive cybercrime evidence analysis model. This proposed framework seeks to positively transform the speed and accuracy of investigative efforts while enhancing the efficiency of the speedy detection, examination, and documentation of computer evidence in cybercrime cases. It will be shown that through the implementation of these advanced technologies, the overall digital forensic methodologies can shift from being a reactive discipline to one that proactively understands cyber threats before they are fully realized.

Despite progress in digital forensic technologies [2], a notable challenge persists: the time duration it takes for a cybercrime to be committed, identified, and investigated [5]. Traditional practices often require significant manpower to execute certain procedures, which may slow down the process of acquiring and analyzing evidence [10]. This delay not only hinders investigations but also reduces the likelihood of successful prosecution. Moreover, currently established forensic instruments face challenges with the complex and vast data processed and shared on digital platforms

[19]. These tools often lack the capacity to rapidly integrate this data within the context of a cybercrime, resulting in analysts having to filter through extensive amounts of potentially irrelevant information [12].

It is imperative to implement a new, dynamic, and responsive model that mobilizes the use of advanced technologies to increase not only the speed and efficacy of actions toward identifying cybercriminal activities [8] but also the scientific credibility of the procedures used [21]. The application of LLMs might help to fill this gap and offer new approaches that correspond to the continuously shifting and developing nature of electronic criminal investigations [23].

In consideration of the challenges that have been delineated, the principal paper inquiry engaged by this investigation is:

- How can LLMs enhance the flexibility and initiative of cybercrime evidence investigation in the context of digital forensics?

Based on this primary paper question, the analysis of specific goals oriented toward addressing the paper paradox is determined.

1. To determine the capability of LLM when working with big data and in digital cybercrime investigations.
2. To determine the extent to which models can be applied to improve data interpretation and evidence collection, ultimately contributing to the efficiency and accuracy of cybercrime investigations.

However, this paper will also seek to evaluate the efficiency and reliability of LLMs in providing accurate and timely intelligence information from computer evidence. Forensic Tool Integration based (FTI) and conventional approach will be used to analyze the effectiveness of the proposed model between the two. Last but not least, the goal is to create a repository of recommendations and policies for Language Forensic Model (LFM) practitioners concerning the use of LLMs in cybercrime investigation processes. This involves recommendations concerning training, practice application, and evaluation for ensuring the long-term applicability of LLM technologies to standard forensics.

The rest of the paper is organized as follows: Section 2 provides a literature review on Agile forensics and LLMs in cybersecurity. Section 3 discusses the methods used in this paper, including evidence extraction, recommendation, and report generation. Section 4 examines the results of the study, comparing automated metrics and manual evaluations across various activities. Section 5 summarizes the main findings, outlines the benefits of the proposed model in digital investigation, and discusses potential future developments.

2. Literature Review

Considering the prior literature, it is possible to identify both the drastic improvement in the sphere of digital forensics and crucial gaps in the capacity of the existing approaches in dealing with sophisticated cyber threats.

Flexibility in the context of digital scenarios is the capacity to achieve significant changes in response to new threats encountered in the sphere of cybersecurity. Puzis *et al*. [17], stressed on the need to promote agile methods by which the forensic processes can match up to rapidly evolving cyber threats. That is how Agile methodologies help forensic teams become more adaptive and proactive at the same time and develop their tactics in parallel to threats. This versatility is important because traditional approaches are not very effective because of the constantly evolving threats of cyber-criminal activities. Finally, agility brings about the principles of flexibility and collaboration, flexibly important when handling complications in the cybercrime cases. Applying the concept of agility allows teams to operate in cycles making changes as more information is received. Such methodology does not only enable a quicker approach to conducting an investigation but also guarantees wider reach with regards to identifying the possible risk factors that can be exploited. These methodologies can clearly mitigate these delays since they nurture a culture that embraces change and short, succinct iterations. However, agility in digital forensics has challenges; these are the enhancements of these methodologies within the current structures of forensic models. It means that the utilization of agile techniques requires the change in organizational culture and in some cases, the change of teams to include more of agile workflow approach. This then can sometimes be met with some form of resistance especially if the transformation is targeting an organization or department that operates in a structured format.

AI holds disruptive innovative value for digital forensics while still being confined by existing paradigms. The authors Jacob *et al*. [14] explain that mundane functions can be performed by AI technologies so that human analysts do not have to, large data sets can be managed and analyzed, and the accuracy of digital evidence will increase. However, adoption of these technologies in integration with conventional forensic approaches has been a bit slow than expected, largely due to technical and organizational factors despite the capability of improving detection and attribution in digital forensics through AI-based automation of evidence capture, processing, and analysis, the integration of these technologies into existing organizational approaches have featured significant technical and organizational challenges. They have made investigations time much shorter, and the chances of identifying the culprits to be much higher than it used to be before. Besides, in the case of using AI algorithms in forensic practice, more patterns and anomalies not seen by an investigator can be found and contribute to a higher level of precision

and results of forensic work. Nevertheless, the forensic use of innovative AI has a few challenges such as data privacy concerns, vagueness of algorithms, and difficulty due to integration of AI with current lengthy systems. It is also important to have professionals who can take charge of the systems with Artificial Intelligence. These are issues that need constant investigation and innovation and integration of efforts from both technologists and forensic analysts.

An equally relevant study by Usman *et al*. [22] hinges on the promotion of machine learning in digital evidence assessment and affirms the increase in investigation speed and accuracy of investigations, still, the existence of problems, for example replication data bias and human supervision. Soon, objects will require connection with each other, meaning that such devices can also collect personal data and a variety of security threats, increasing cybercriminal activity. New strategies of cyber security that can identify hostile IP addresses before connection needs to be established to prevent cybercrimes. The best method of profiling the behavior of security threats to the cyber-physical system include the IP reputation system. Current reputation systems are not very effective for the following reasons, high administrative coast, higher false positive rate, more time consuming, and consider a limited number of source data for IP address reputation claims. In this regard, we have introduced a new, technically proved hybrid solution which integrates data forensics, Machine Learning (ML), threat intelligence, and dynamic malware analysis to resolve these problems. The Decision Tree (DT) technique is incorporated in behavioral analysis of Internet Protocol commenced from big data forensics to classify closely related zero-day attack and to predict the reputation of an IP address at the pre-acceptance stage. The method followed in the present work suggests the identification of big data forensic problems and at the same time calculates the confidence and lifespan of the data and the severity and risk score associated with the data. Two methods are used to evaluate the proposed system: Firstly, we compare the respective ML techniques in order to achieve the highest F-measure or precision and recall ratings; secondly, we compare the whole concept of the reputation system with the already existing reputation systems. Moreover, it can also be avoided that the notorious security weaknesses that today's rudimentary reputation engines failed to detect.

A similar work by Al-Mousa [3] on proactive digital forensic models reveals that such approaches are grossly underdeveloped, especially with regard to predictive perspectives. It is therefore important to note that given the increased use of IoT applications across the globe a large amount of data is produced and processed. Therefore, the amount of electronic information that needs to be processed in the case of a cybercrime is huge. Thus, the measures devoted to security issues take more time and efforts to be implemented. The analysis stage is thus a critical and complex one in solely digital forensic investigations. In this study, an initial comprehensive proactive strategy for the Internet of Things cybercrime assessment is presented. Moreover, the methodology is aimed at the preliminary classification of the evidence relevant to prior offences, importance of the evidence in reference to the cybersecurity threat, and the gravity of the evidence in relation to the likelihood of a cybercrime occurrence. The automated forensic investigation procedure is expected to time and effort with this methodology.

As shown in Table 1, prior work focused on AI-driven forensics (e.g., [14, 22]) but lacked integration with LLMs for real-time analysis. While existing studies highlight AI's role in forensics [3, 14, 22], none explore LLMs' potential for agile, proactive evidence extraction. Our work bridges this gap by proposing a dynamic LLM-based framework tailored for cybercrime investigations."

Table 1. Summarize key studies.

Study	Methodology	Key findings	Limitations
Jacob *et al*. [14]	AI for data analysis	Improved evidence accuracy	No real-time application
Usman *et al*. [22]	ML for IP reputation	Faster threat detection	Bias in training data
Al-Mousa [3]	Proactive IoT forensics	Early threat classification	Limited to IoT environments

Husak *et al*. [13], present a description of the current reactive models as was already discussed, describe various aspects of the predictive methods in cyber defense and present three modern approaches to exemplify how they fail in anticipative handling of the cyber threats that align with the need for better correspondence of the method. The first method predicts continuing cybertrends in similar ongoing cyberattacks through the use of data mining to analyze typical attack patterns. The second method involves forecasting the risky malevolent actors using a dynamic network entity reputation score. The third method predicts network attack rates using time series analysis. In addition to the comparison of the approaches and to illustrate how predictive analysis may be used in current and future paper and cyber security practice this paper presents a novel evaluation of each of the three types of approaches in the context of an intrusion detection alert sharing platform. Thus, based on the results of our experiments, all three methods showed high levels of potential accuracy and significantly satisfactory levels of technological readiness for a combination of experimental use in an operating environment. However, these two methods are useful for the predictive blacklisting, especially that the projection technique gives more detailed result, although it is less expandable than the prediction technique. Being accurate and light in weight, the network security situation forecast does not reveal details of situations expected to happen.

Still, an investigation by Dragonas *et al*. [7], reveals

that there is a significant paper gap given that there exists limited evidence of enacted elaborate models with and or incorporating LLMs to mesh with forensic protocol. For instance, there is an AI ChatGPT from Open AI that gained millions of users within months of its release to the public. However, thus far, there is no application that became as popular as OpenAI's ChatGPT became. OpenAI introduced its ChatGPT application mobile last year. Sadly, until today, even open source and commercial tools cannot analyze this application, which is used in number of applications, including, maybe, malicious ones. However, the information this program saves like JSON files, which the customer uses to describe their conversations with ChatGPT, can be very important to apportion responsibility for user actions or behaviors, establish the offender's knowledge, and purpose, as well as draw conclusions in real-life scenarios. This paper considers only the actual content that exists in the OpenAI's ChatGPT mobile application for IOS and Android only, and the focus is mainly on any useful information that can be gleaned from it. Furthermore, considered are the other app-level cloud-native measurements which can be derived from requested user export data. The main purpose of this paper study is to identify actual case scenarios where investigators could effectively employ the particular mobile app and the artifacts discovered during the investigation. To assist these specialists, the writers have also participated in FOSS.

Last of all, Dunsin *et al*. [9] stress that there is high demand for digital forensic models that should reflect both virtual advancements and real-life operational concerns LLM incorporation being one of the beneficial aims, the purpose is to present in detail what types of AI and ML are used in incident response and digital forensics. The paper is an investigation of current high impact paper projects that cover areas ranging from data collection and data recovery, timeline reconstruction, big data analysis, pattern analysis, preservation of the chain of custody, and staging reactive strategies against hacker attacks. This undertaking goes further to explore and discover how these approaches are addressing these essential fields of digital forensic practice. AI remains a promising tool for developing digital forensics, but the emerging challenges in terms of increasing database sizes and evolving criminal tactics require a continuous exchange and collaboration with work in this subject area. To better understand the role of AI and ML in this subject area and in addressing the paper questions, this paper investigates both the contributions and limitations, as well as the areas left unexplored in the current literature. To make future advances toward the augmented use of AI in the context of digital forensics and incident response, we identify and discuss many paper areas by further highlighting the need for strategic planning and paper in this area. Specifically, this paper focuses on the guidelines for AI and ML implementation in digital forensics and the information on the benefits, limitations, and implications arising from it for countering modern threats in the field of cybersecurity.

Recent studies (e.g., [7, 9, 13]) emphasize predictive analytics in cyber defense but overlook LLMs' natural language capabilities for forensic reporting. Our model addresses this by combining proactive analysis with LLM-driven automation.

3. Methodology

The following subsection provides a step-by-step account of the process of constructing a generic Forensic Detection and Analysis (FDA) model based on the University of New South Wales-Network-Based 2015 (UNSW-NB15) dataset. To ensure rigorous and systematic evidence extraction, as well as coherent insights into clear security recommendations, a methodology based on LLMs was applied. Figure 1 illustrates the working methodology used in this paper.

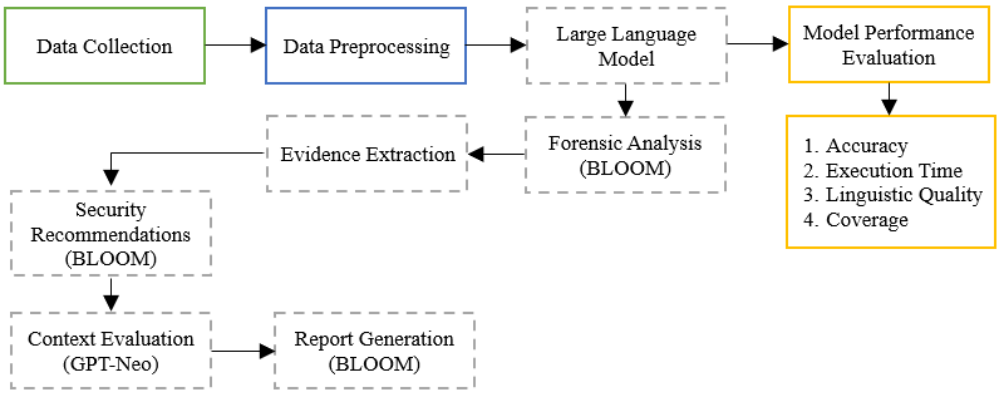

Figure 1. The methodology.

3.1. Dataset Description

For this paper, the dataset used was the UNSW-NB15 obtained from Kaggle which is widely used in network traffic analysis. This dataset has extensive information that relates to the natural environment of the networks, and has incorporated both the natural traffic as well as the attack traffic. This dataset is complex due to the nature of attack types including but not limited to the

following: Denial of Service (DoS), reconnaissance, exploits, among others. It features diverse attributes regarding networks that are protocol type, connection time, and the number of packets transmitted and received in addition, it is a great tool in data forensics and forensic evidence search. After an exhaustive search of the internet, you can directly visit by entering the following (https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15). Table 2 shows the characteristics of the dataset.

Table 2. Characteristics of the dataset.

Dataset name	UNSW-NB15
Number of records	Training data: 82,332 records. Test data: 17,000 records.
Features	Includes 45

3.2. Data Preprocessing

In the following part of the section, the practical measures for the preparation and analysis of the data are detailed. Due to the fact that the data in this case is fairly large and dispersed as Figure 2 below shows, it was required to preliminary check the data to guarantee the most correct and flawless analysis.

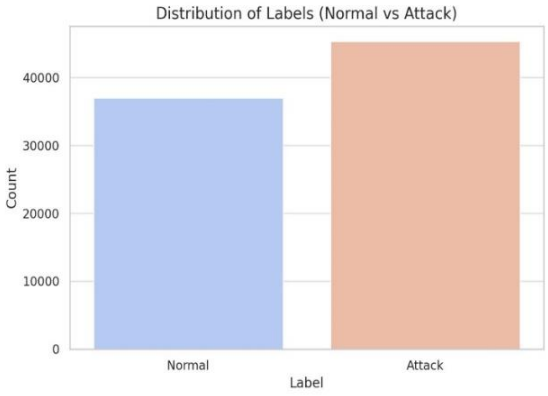

Figure 2. Distribution analysis of the data.

Records showing the label value as 1 which indicates the presence of offence related traffic was filtered for analysis. This step dealt specifically with records that are of criminal interest in an effort to enhance the quality of the analysis as well as the simplicity of the data. This led to the creation of a specific set of records whose accuracy boosted the efficiency of the used models and decreased the time which was needed to process them, all of which contributed to raising the accuracy of the results.

In Big-science Large Open-science Open-access Multilingual Language Model (BLOOM LLM), the technical characteristics of every record in the data are converted into clear, human-understandable descriptive text, enabling the language models to more easily analyze forensic evidence. For example, the following text represents a description of a particular record: Traffic detected for the following parameters: Protocol: Transmission Control Protocol (TCP), State: FIN (Finish), duration of the connection: 2.3 seconds, sent packets: 15, Rcvd packets: 20. Attack type: DoS." This optimization not only enhanced data readability but also facilitated subsequent phases of analysis for recommending features under suspicion and other vital forensic data most helpful for investigative purposes enhancing the efficiency of the analytical stages.

The derived descriptive texts meant that there were several sections of analysis to include in order to address the forensic and security requirements. The first method deployed in the work started with the process of identifying the evil patterns within the texts via forensic analysis and collection through Text-To-Text Transfer Transformer Large Language Model (T5LLM) then proceeded through the second step of identifying the evidence behind each record. Based on the result, specific security recommendations were given together with the identified vulnerabilities and risks using the BLOOM LLM approach. For this purpose, the texts obtained from the analysis process were checked for credibility and quality by implementing GPT-Neo. Lastly, reports were prepared by integrating text and image data, analyzed findings, extracted arguments, and security recommendations using BLOOM LLM which guaranteed the combination of all case aspects.

To reorganize the descriptive texts and the relative texts produced by the models into a structure, an approach was adopted to chop down the texts into basic elements and categories them into certain fields of the data frame. For instance, protocol type, state, time duration, number of packets transferred in and out and the kind of malicious activity detected, if any was extracted and tabularized. This format enables applying various criteria that define performance measurements including accuracy of performance, clarity of performance, relevant of performance and ability to inspire or provoke action. This also enabled the comparison between results in different models within the organization, as well as performance assessment in certain parameters, which made the evaluation more credible and accurate.

3.3. Performance Evaluation

Nevertheless, to check for the extensiveness and accuracy of the performed analysis, the LLM used was assessed against some basic performance benchmarks. These metrics included:

1. Accuracy: indicates the level of accuracy of the results that are given by the model.
2. Coverage: gives an analysis of how adequately each of the models addresses essential facets of the texts.
3. Actionability: consecutively confirms the precision and adequacy of the presented recommendations.
4. Execution time: records the span of time required to complete each model in order to analyze the texts.
5. Linguistic quality: helps to make certain the texts are comprehensible and stylistically correct.
6. The results were arrived at for each model after

applying the descriptive texts of the suspicious records and weightage based on these metrics was compared in quantitative terms.

4. Results

In this section the current state of development of the used LLMs is demonstrated and their performance analyzed in all stages of the forensic process. The described methodology was chosen to be rather strict and compare each model according to diverse parameters such as accuracy, relevance, language, and creativity. The results are designed to give a general idea of how efficient these models are in terms of identifying patterns of suspicion, identifying evidence, offering security suggestions, and creating detailed reports. The identification of areas that can be improved as a basis for better performance of models is the primary concern of the analysis carried out in this study.

The findings reported in this study were gathered in a systematic and systematic manner by employing key performance indicators of automated analysis complemented with qualitative analysis. The methodology followed these key steps:

4.1. Automated Metrics

Objective measures were employed to obtain the performance predications of models automatically. These metrics included:

1. Accuracy: computed by comparing their estimated values with true values on some tasks, like forensic analysis or evidence extraction, in the dataset.
2. Precision and coverage: used in the evidence extraction step to determine the accuracy of information retrieved and the coverage of information.
3. Execution time: captured during the pipeline run for each of the steps to determine the time taken to compute.
4. Output lengths: regulated to make sure that the generated outputs were as concise as required to suit their intended use and at the same time had conforming formatting.

These metrics offered a measure upon which to assess its effectiveness and productivity of the various stages.

4.2. Manual Evaluations via ChatGPT

For tasks that involved qualitative assessment, a manual process of evaluation was undertaken with support from prompt answers generated by ChatGPT OpenAI. The steps included:

- **Dataset Preparation**: to overcome this problem, 100 examples of the suspicious records were chosen for the study because they were varied in respect to different types of scenarios and patterns.
- **Evaluation Prompts**: as for each of these examples, there were specific test stimuli given to ChatGPT that corresponded to the task as follows: Clarity, Relevance, Action, and Innovation.
- **Example**: self-assessment: analyze the clarity and the relevance of the following forensic analysis output the score of simple vivid outcomes of each user appeal has ranged from 1 (poor) to 5 (excellent).
- **Systematic Scoring**: afterwards, ChatGPT also provided ratings with regard to quality as well as qualitative feedback for each example given and such feedback was used to derive performance ratings.

Such an approach made it possible to have a strong assessment of factors that could be hard to quantify using other automated systems. ChatGPT was chosen for the manual evaluations mainly because of the ability to bring different manuscripts for a completely impartial assessment. To eliminate subjectivity and possible variability within the questions, a single evaluation mechanism was used for all the examples provided above.

Additionally, ChatGPT provided an opportunity to assess qualitative factors in a large quantity. This choice provided sufficiently wide coverage for the dataset and considered various types of attacks and their patterns involving a quantitative number of 100 samples. This method includes quantitative results while also offering personal approaches with subjectivity on overall pipeline performance evaluation. Every quantitative or qualitative measure identified was obtained systematically, making the collected data credible, valid and replicable.

4.2.1. Forensic Analysis

This was evidenced from the high performance displayed by the forensic analysis module of detecting the aforementioned patterns in textual data. The data used in the evaluation incorporates presented in Table 3.

Table 3. Forensic analysis results.

Metric	Result
Accuracy	95%
Average relevance	4.5/5
Average clarity	5/5

The forensic accuracy was even closer to perfect the relevance of the descriptions was estimated as very high, 4.50/5 while their clarity was estimated as perfect, 5.00/5. This shows the model's capacity for converting complex technical information into actionable and easily comprehensible forensic information.

Figure 3 illustrates the distribution of data rates across various attack categories. The horizontal axis represents the target groups, which in this context are different types of attacks, including "normal", "backdoor", "DoS", and others. The vertical axis indicates the data rate, a numerical feature that quantifies the frequency or intensity of each attack type. This visualization helps in understanding the prevalence

and impact of different attack categories within the dataset.

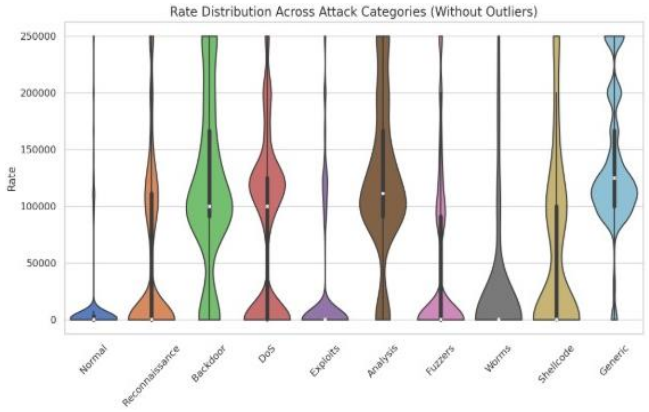

Figure 3. Attack categories.

Figure 4 shows the Horizontal axis: which represents the target groups (in this case, protocol types such as "TCP", "User Datagram Protocol (UDP)", etc.,). Vertical axis: means the data rate, which is a numerical feature.

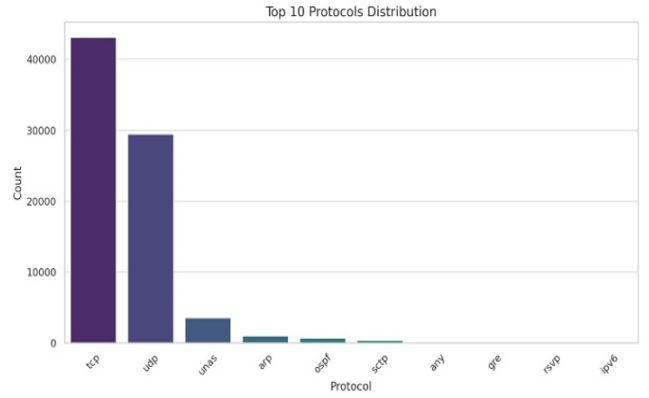

Figure 4. Protocols distribution.

4.2.2. Evidence Extraction

The evidence extraction module emphasized the identification of key information from the input data. Table 4 presents the performance of evidence extraction.

Table 4. Evidence extraction performance.

Metric	Result
Precision	94%
Coverage	95%
Average conciseness	48.5 characters

From the precision and coverage results, it proves that the extracted evidence is precise and relevant, lacking no important information. Moreover, the average length of extracted evidence proves the effectiveness of the model to be concise.

4.2.3. Recommendation Generation

In this step, implemented valuable and unique security strategies that were informed by the evaluated information as shown in Table 5.

Table 5. Recommendation generation evaluation.

Metric	Result
Average actionability	4.5/5
Average innovation	3.5/5
Average linguistic quality	5/5

These considerations were actionable (all from the 4th category, they are valuable and highly recommended (4.5)) and well-written (5). However, there is the hope for the improvement of the novelty of the suggestions it has received a rate of 3.5 out of 5.

4.2.4. Context Evaluation

In the context evaluation stage, the generated textual outputs were evaluated based on their relevance and coherency as shown in Table 6.

Table 6. Evaluation of coherence and comprehensiveness.

Metric	Result
Average consistency	4/5
Average comprehensiveness	4.5/5
Average Coherence	5/5

These evaluations appeared to be very precise and encompassing while making clear that there is potential for greater cohesiveness across different types of outputs.

4.2.5. Report Generation

Table 7 presents the results of the general reports were produced in order to present all the analysis phases as an entire process.

Table 7. Report generation results.

Metric	Result
Average comprehensiveness	4.5/5
Average linguistic quality	5/5
Report accuracy	96%

The reports were effective in providing both substantial analyses and high language standard to reach high overall accuracy and coverage.

4.2.6. Execution Metrics

In addition to the qualitative and quantitative results, the execution time and output lengths for each module were evaluated as presented in Table 8.

Table 8. Each module's execution time and output length.

Module	Avg. execution time (s)	Avg. output length (words)
Forensic analysis	57.10	881.0
Evidence extraction	0.67	9.4
Recommendation generation	57.42	881.0
Context evaluation	18.99	867.8
Report generation	56.65	881.0

The evaluation results showed that Evidence Extraction System (EES) had the shortest executing time, whereas forensic analysis and recommendation generation had longer time due to the nature of the processes. But as with all-time-consuming work, it is

well worthwhile because the quality of output is superior Figure 5 shows them.

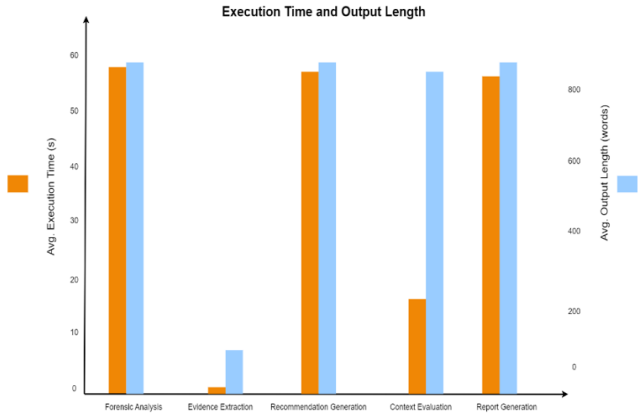

Figure 5. Execution time and output length chart.

5. Conclusions

The level of complexity regarding the variety of cyber threats has risen significantly, thus increasing the requirement for more complex instruments needed for accurate forensic investigation and sound decisions within the digital forensic domains. In this paper work, the strengths of LLMs have been unleashed to solve problems such as evidentiary analysis and forensic assessment and outline the necessary security measures within the context of a quantitative analytical framework. Some of the approaches used included converting the complex technical information into plain English, human-like texts, breaking tasks into analytical segments and the use of LLM such as T5, BLOOM and GPT-Neo for automated insight. Based on this, performance metrics could be obtained through automated comparisons together with the evaluation from the ChatGPT procedure. Through a comprehensive assessment of the work on both forensic analysis and evidence extraction, the study obtained promising results with a 95% accuracy achieved with forensic analysis, while maintaining a coverage rate of 94% in terms of the evidence extraction. Concepts like clarity and actionability were rated using a human-like check, thereby verifying the attributes of the listed qualitative characteristics. Further work will involve improving the size of the dataset, the method of evaluating the performance and possibly incorporating features that involve real-time analysis for improving the use of LLMs in digital forensics. In this paper, there is a significant potential of utilizing sophisticated AI tools in enhancing the procedures involved in recognising and solving offences, and highly beneficial in the fight against cybercrime as well as enhancing better cybersecurity strategies.

References

[1] Al-Khateeb M., Al-Mousa M.., Al-Sherideh A., Almajali D., Asassfeha M., and Khafajeh H., "Awareness Model for Minimizing the Effects of Social Engineering Attacks in Web Applications," *International Journal of Data and Network Science*, vol. 7, no. 2, pp. 791-800, 2023. DOI: 10.5267/j.ijdns.2023.1.010

[2] Al-Milli N., Jobair Z., Al-Mousa M., Alshaikh A., Asassfeh M., Alazaidah R., and Al-Daoud E., "Data Integrity Concerns, Requirements, and Proofing in Cloud Computing," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 12, pp. 5033-5043, 2024. https://jatit.org/volumes/Vol102No12/12Vol102No12.pdf

[3] Al-Mousa M., "Generic Proactive IoT Cybercrime Evidence Analysis Model for Digital Forensics," *in Proceedings of the International Conference on Information Technology*, Amman, pp. 654-659, 2021. https://ieeexplore.ieee.org/document/9491718

[4] Al-Sherideh A., Ismail R., Al-Mousa M., Al-Qawasmi K., Al-Shaikh A., Awwad H., Maabreh K., and Alauthman M., "Development of a Secure Model for Mobile Government Applications in Jordan," *Journal of Statistics Applications and Probability*, vol. 13, no. 1, pp. 145-155, 2024. https://digitalcommons.aaru.edu.jo/jsap/vol13/iss1/10/

[5] Casino F., Dasaklis T., Spathoulas G., Anagnostopoulos M., Ghosal A., Borocz I., and Patsakis C., "Paper Trends, Challenges, and Emerging Topics in Digital Forensics: A Review of Reviews," *IEEE Access*, vol. 10, pp. 25464-25493, 2022. DOI: 10.1109/ACCESS.2022.3154059

[6] Costantini S., De Gasperis G., and Olivieri R., "Digital Forensics and Investigations Meet Artificial Intelligence," *Annals of Mathematics and Artificial Intelligence*, vol. 86, no. 1, pp. 193-229, 2019. https://link.springer.com/article/10.1007/s10472-019-09632-y

[7] Dragonas E., Lambrinoudakis C., and Nakoutis P., "Forensic Analysis of OpenAI's ChatGPT Mobile Application," *Forensic Science International: Digital Investigation*, vol. 50, pp. 301801, 2024. https://doi.org/10.1016/j.fsidi.2024.301801

[8] Dubey H., Bhatt S., and Negi L., "Digital Forensics Techniques and Trends: A Review," *The International Arab Journal of Information Technology*, vol. 20, no. 4, pp. 644-654, 2023. DOI: 10.34028/iajit/20/4/11

[9] Dunsin D., Ghanem M., Ouazzane K., and Vassilev V., "A Comprehensive Analysis of the Role of Artificial Intelligence and Machine Learning in Modern Digital Forensics and Incident Response," *Forensic Science International: Digital Investigation*, vol. 48, pp.

301675, 2024. https://doi.org/10.1016/j.fsidi.2023.301675

[10] George J., "Advancing Enterprise Architecture for Post-Merger Financial Systems Integration in Capital Markets laying the Foundation for Machine Learning Application," *Australian Journal of Machine Learning Paper and Applications*, vol. 3, no. 2, pp. 429-475, 2023. https://sydneyacademics.com/index.php/ajmlra/article/view/155

[11] Harpring R., Maghsoudi A., Fikar C., Piotrowicz W., and Heaslip G., "An Analysis of Compounding Factors of Epidemics in Complex Emergencies: A System Dynamics Approach," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 11, no. 2, pp. 198-226, 2021. https://www.emerald.com/insight/content/doi/10.1108/jhlscm-07-2020-0063/full/html

[12] Hughes J., Chua Y., and Hutchings A., *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches*, Palgrave Macmillan, Cham, 2021. https://doi.org/10.1007/978-3-030-74837-1_10

[13] Husak M., Bartos V., Sokol P., and Gajdos A., "Predictive Methods in Cyber Defense: Current Experience and Paper Challenges," *Future Generation Computer Systems*, vol. 115, pp. 517-530, 2021. https://doi.org/10.1016/j.future.2020.10.006

[14] Jacob L., Thomas K., and Savithri M., *Artificial Intelligence for Cyber Defense and Smart Policing*, Chapman and Hall/CRC, 2024. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003251781-4/ai-forensics-lija-jacob-thomas-savithri

[15] Lauriola I., Lavelli A., and Aiolli F., "An Introduction to Deep Learning in Natural Language Processing: Models, Techniques, and Tools," *Neurocomputing*, vol. 470, pp. 443-456, 2022. https://doi.org/10.1016/j.neucom.2021.05.103

[16] Malik A., Bhatti D., Park T., Ishtiaq H., Ryou J., and Kim K., "Cloud Digital Forensics: Beyond Tools, Techniques, and Challenges," *Sensors*, vol. 24, no. 2, pp. 1-30, 2024. https://doi.org/10.3390/s24020433

[17] Puzis R., Zilberman P., and Elovici Y., "ATHAFI: Agile Threat hunting and Forensic Investigation," *arXiv Preprint*, vol. arXiv:2003.03663v1, pp. 1-12, 2020. https://arxiv.org/pdf/2003.03663

[18] Sharma S., "Digital Forensics: Legal Standards and Practices in Cybercrime Investigation," *in Proceedings of the 4th International Conference on Innovative Practices in Technology and Management*, Noida, pp. 1-6, 2024. https://ieeexplore.ieee.org/document/10563327

[19] Stoyanova M., Nikoloudakis Y., Panagiotakis S., Pallis E., and Markakis E., "A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues," *IEEE Surveys and Tutorials*, vol. 22, no. 2, pp. 1191-1221, 2024. https://ieeexplore.ieee.org/abstract/document/8950109

[20] Tanner A., Dancer F., Hall J., Parker N., Bishop R., and McBride T., "The Need for Proactive Digital Forensics in Addressing Critical Infrastructure Cyber Attacks," *in Proceedings of the International Conference on Computational Science and Computational Intelligence*, Las Vegas, pp. 976-982, 2022. https://ieeexplore.ieee.org/document/10216685

[21] Tok Y. and Chattopadhyay S., "Identifying Threats, Cybercrime and Digital Forensic Opportunities in Smart City Infrastructure Via Threat Modeling," *Forensic Science International: Digital Investigation*, vol. 45, pp. 301540, 2023. https://doi.org/10.1016/j.fsidi.2023.301540

[22] Usman N., Usman S., Khan F., Jan M., Sajid A., Alazab M., and Watters P., "Intelligent Dynamic Malware Detection Using Machine Learning in IP Reputation for Forensics Data Analytics," *Future Generation Computer Systems*, vol. 118, pp. 124-141, 2021. https://doi.org/10.1016/j.future.2021.01.004

[23] Wickramasekara A., Breitinger F., and Scanlon M., "Exploring the Potential of Large Language Models for Improving Digital Forensic Investigation Efficiency," *arXiv Preprint*, vol. arXiv:2402.19366v3, pp. 1-20, 2025. https://arxiv.org/pdf/2402.19366

Mohammad Al-Mousa is an Associate Professor in the Faculty of IT, Cybersecurity Department, Zarqa University. He received his PhD in Network Security from Universiti Sains Malaysia in 2013. His research interests include Digital Forensics, Cybersecurity, E-Government Strategy, Cloud Computing, and Software Engineering.

Waleed Amer is a postgraduate student in the Faculty of IT, Cybersecurity Department, Zarqa University. He received his Bachelor's in Cybersecurity in 2022. His research interests include Cybersecurity and Artificial Intelligence.

Mosleh Abualhaj is a Senior Lecturer in Al-Ahliyya Amman University. He received his first degree in Computer Science from Philadelphia University, Jordan, in 2004, master degree in Computer Information System from the Arab Academy for Banking and Financial Sciences, Jordan in 2007, and Ph.D. in Multimedia Networks Protocols from Universiti Sains Malaysia in 2011. His research area of interest includes VoIP, Congestion Control, and Cybersecurity Data Mining and Optimization.

Sultan Albilasi is a postgraduate student in the Faculty of Information Technology, Cybersecurity Department, at Zarqa University. He received his Bachelor's degree in Computer and Information Sciences in 2018. His Master's thesis focuses on the development of an automated DDoS attack detection model using AI techniques.

Ola Nasir is an Assistant Teacher in the Faculty of IT, Computer Science Department, Zarqa University. She received his Master's in computer science in 2024. Her research interests include Machine Learning, Deep Learning, and Artificial Intelligence.

Ghassan Samara is currently an Associate Professor and vice dean of the Faculty of Information Technology at Zarqa University, Jordan. He Holds a BSc. and MSc. in Computer Science and a Ph.D. in Computer Networks. He obtained his Ph.D. from Universiti Sains Malaysia (USM) in 2012. His research interests include the Internet of Things, Cryptography, Vehicular Ad-Hoc Networks, Wireless Sensor Networks, Internet Access Protocols, Accident Prevention, Alarm Systems, Cloud Computing, Computational Complexity, Computer Crime, Computer Network Security, Data Integrity, Data Privacy, Delays, Digital Forensics, Directed Graphs, Energy Conservation, Energy Management Systems, Fog Computing, Graph Theory, Home Automation, Intelligent Robots, and Intelligent Transportation Systems. He has authored more than 150 scientific research papers published in international journals and conference proceedings.