The Use of Artificial Intelligence in Digital Forensics and Incident Response (DFIR) in a Constrained Environment

Dipo Dunsin
School of Computing and Digital Media,
London Metropolitan University, London, UK
tsdduns1@londonmet.ac.uk

Mohamed C. Ghanem
School of Computing and Digital Media,
London Metropolitan University, London, UK
ghanem@londonmet.ac.uk

Karim Ouazzane
School of Computing and Digital Media,
London Metropolitan University, London, UK
k.quazzane@londonmet.ac.uk

***Abstract*—** Digital investigators often have a hard time spotting evidence in digital information. It has become hard to determine which source of proof relates to a specific investigation. A growing concern is that the various processes, technology, and specific procedures used in the digital investigation are not keeping up with criminal developments. Therefore, criminals are taking advantage of these weaknesses to commit further crimes. In digital forensics investigations, artificial intelligence (AI) is invaluable in identifying crime. It has been observed that an algorithm based on AI is highly effective in detecting risks, preventing criminal activity, and forecasting illegal activity. Providing objective data and conducting an assessment is the goal of digital forensics and digital investigation, which will assist in developing a plausible theory that can be presented as evidence in court. Researchers and other authorities have used the available data as evidence in court to convict a person. This research paper aims at developing a multiagent framework for digital investigations using specific intelligent software agents (ISA). The agents communicate to address particular tasks jointly and keep the same objectives in mind during each task. The rules and knowledge contained within each agent are dependent on the investigation type. A criminal investigation is classified quickly and efficiently using the case-based reasoning (CBR) technique. The proposed framework development is implemented using the Java Agent Development Framework, Eclipse, Postgres repository, and a rule engine for agent reasoning. The proposed framework was tested using the Lone Wolf image files and datasets. Experiments were conducted using various sets of ISAs and VMs. There was a significant reduction in the time taken for the Hash Set Agent to execute. As a result of loading the agents, 5% of the time was lost, as the File Path Agent prescribed deleting 1,510, while the Timeline Agent found multiple executable files. In comparison, the integrity check carried out on the Lone Wolf image file using a digital forensic tool kit took approximately 48 minutes (2,880 ms), whereas the MADIK framework accomplished this in 16 minutes (960 ms). The framework is integrated with Python, allowing for further integration of other digital forensic tools, such as AccessData Forensic Toolkit (FTK), Wireshark, Volatility, and Scapy.

***Keywords*—**Artificial Intelligence, Computer Science, Criminal Investigation, and Digital Forensics.

I. Introduction

CUTTING-EDGE intelligence tools are not limited to a single field; instead, it has nearly limitless applications in almost every area. One has to consider the fact that the use of AI has helped researchers to analyse complex and large data within a minimum time. Moreover, the insights generated by AI are far more advanced than any other available technology. It has aided researchers and various authorities in looking into available data, analysing it, and using relevant data to convict criminals in the court of justice [8].

The concept of analysing available digital data gave birth to the field of digital forensics. AI has two applications in digital forensics: academic and practitioner. Both the have faced multiple challenges in digital forensics. The major challenges can be divided into three major categories: 1) the growth in the use of storage devices, 2) the sophistication of attacks on the available systems, and 3) the incompatibility of existing tools to keep pace with the opponent's ability to conduct the attack and remove the digital evidence [9]. To solve these challenges, a better, easy, and effective technique needs to be adapted, which is the reason why AI emerged as a significant branch of computer science, which can help both researchers and practitioners to solve the respective challenges. There is an increasing concern that various processes, technologies, and certain procedures which are used for digital investigation purposes have failed in keeping abreast, and that is the reason why criminals are using these weaknesses to commit further criminal activities. It is to be noticed that overabundance is to some extent because of the expanding sizes of datasets. A contributor to the issue exists in light of the expanding number of examinations that are combined with a limited specialist to research them.

Algorithms which are based on AI are very helpful in the detection of risks along with the prevention and prediction of criminal activities. Different researchers and practitioners use algorithms for the analysing of data that is available on the cloud and other devices. In forensics, the use of AI helps to quickly analyse the different aspects that are used by standard algorithms so that the relevant data can be swiftly flagged and the insights can be classified in an appropriate category [6]. It is widely known that AI is dependent on the use of advanced algorithms. These algorithms can help one identify, observe, analyse, and predict different possible outcomes based on time, location, and history. AI has various sub-fields which play an essential role collectively in the detection of risk factors and their solutions. AI is classified into three categories: Narrow, general and super, artificial intelligence. They are used to solve problems that cannot be solved by human intelligence, and hence they surpass human intelligence. Some major AI techniques include supervised and unsupervised learning, machine learning, optimisation, deep learning, hybrid learning, and reinforcement learning. Digital forensics also uses Case-Based Reasoning (CBR), which is an effective technique that effectively handles all the problems which are related to different conventional systems like the symbolic rule-based systems. CBR can handle a situation that is not faced by the system earlier.

II. LITERATURE REVIEW

A. Digital Forensics

Digital forensics is the preservation, removal, analysis, and recording of a computer for lawful reasons [7]. It is the science of finding certain shreds of evidence on the digital media. It provides various forensic tools to handle and solve digital-related cases, and which are complicated. It is also said to be a process that involves not a single but a group of different processes and tasks that are involved in an investigation. In various digital forensic investigations, there are over a hundred different tasks and procedures to be conducted. Some of these tasks and practices are focused on the technologies in data acquisition, whereas some of them focus on the data analysis area of the research. A large number of these procedures were produced for handling distinctive innovation utilised in the reviewed gadget; when the underlying technology of the target gadget changes, new techniques must be created. Many procedures are used in these cases, which are quoted, and they are said to be standard procedures in digital forensic methodologies. However, sometimes certain discrepancies are also found between them [7].

B. Artificial Intelligence

Artificial intelligence is the development of exceptional computers or systems that can perform tasks well beyond the capacities of the majority of humans. AI can also be described as the procedure of replicating or simulating human intelligence in machines that allows an individual to respond to Turing's questions affirmatively. This term is sometimes used to describe the development of processes that have regular human intellectual cycles, such as the ability to reason, find meaning, summarise, or learn from experiences [2]. Since the 1940s, when AI was first created, it has been demonstrated that computers can be programmed to perform complex tasks such as finding confirmations for a numerical theory or playing chess with exceptional skill

C. Narrow Artificial Intelligence

The most promising type of AI to date is Narrow Artificial Intelligence (NAI). This intelligence is generally goal-oriented, and they are destined to perform a single task like facial recognition, voice assistants, etc. These machines seem to be very vigilant and intelligent, but they have certain limitations and constraints; this is the reason why this type of intelligence is called narrow artificial intelligence [5]. NAI does not totally stimulate human intelligence; it only replicates human behaviours depending on a narrow range of contexts and parameters, for example, Siri in iPhone.

D. Artificial Super Intelligence (ASI)

Artificial super intelligence is a type of knowledge that does not only stimulate or replicate human behaviours and intelligence; instead, it becomes self-aware and surpasses the capacity of human intelligence. ASI is also referred to as a problem-solving system with high intellectual capabilities. Super intelligence is a technique that utilises robots that overthrow, overrun, and surpass humanity in all aspects. ASI seems to evolve all the human emotions, and it does not only understand them but also comprehends these beliefs, emotions, and desires on its own [12].

E. Limited Memory

This type of AI contains machines that can analyse previous histories like autonomous vehicles, quickly observing other cars, speed, and other directions. But none of this can be accomplished in a single moment; it necessitates the identification of individual objects as well as ongoing surveillance. These findings are applied to the pre-programmed autonomous vehicles, such as detecting the traffic light signs, motion, and other activated sensor's responses around the elements such as road maps and curves [11]. They are included when the car decides to avoid the drivers in other vehicles or when the driver decides to change the lane so that it can avoid hitting a nearby car. But this straightforward information about the past is temporary and is not saved in the library of cars.

F. Optimisation

Today, AI is supported by various optimisation techniques and powerful data sciences. The optimisation is a procedure in which a decision is set for the variable values in a way that the question is optimised in AI. The optimal solution is a collection of variable decisions that can maximise or minimise the neural functions and satisfy the constraints. Generally, the optimisation technique in AI is obtained when the corresponding variables of the related variables give the best yield of the objective function that satisfies all the constraint models. Some new and modern optimisation methods are also introduced apart from the gradient-based optimisation techniques that help to solve some very complex issues.

G. Machine Learning

Artificial machine learning is the technique of solving solutions and transforming risk management. Machine learning techniques benefit greatly in risk management, which is proved from the current practices and evidence [1]. There are various applications of machine learning in AI which are utilised for risk management. One application of machine learning is in credit management. It is the economic loss from the fulfilment of the contractual obligations to the failures of counterparties or during the increased transaction time risk [3]. Traditionally, to deal with such problems, regression models are used. Since some conventional risk management approaches are ineffective, institutions are increasingly interested in using artificial intelligence and machine learning to improve risk management practices.

H. Supervised Learning

Supervised learning in AI is a program in which the system is provided with labelled input data with an output that is expected. It is an approach in which the computer algorithm is trained for input data to get a particular output out of it. This model is trained until it recognises underlying patterns, relationships, and similarities in the input and output data that enable accurate labelling. Supervised learning is excellent in regression and classification risk management, such as

determination of the category of an article or prediction of the volume of sales in the future [4]. Supervised learning is improved in the neural network of algorithms by measuring the outputs of a model constantly to get closer to the target accuracy of a system. The accuracy obtained is determined by two factors: the availability of labelled data and the algorithm employed.

I. Deep Learning

A subfield of machine learning is deep learning which is concerned with the algorithms that are inspired by functions and structures of brains called the neural networks [10]. It is, however, a feature of AI that is designed to be used in object detection, language translation, speech recognition, and critical decision-making by simulating how human brains process data. Deep learning can learn without the supervision of humans, which are drawn from data that is both unlabelled and unstructured. It is a kind of machine learning that is used to detect money laundering and other types of fraud. Deep learning uses networks that can learn unsupervised, unstructured data and are often referred to as a neural or deep neural network.

III. RESEARCH METHODOLOGY

The Multi-Agent Digital Investigation Tool Kit (MADIK) is an application of AI in computer forensics which generally refers to the computational systems that are composed of more than one agent. AI also uses an intelligent software agent (ISA) to accomplish the set goals. ISA is actually an interaction with the environment that perceives and acts autonomously to achieve the goals. The multiagent system is developed in an environment where different tools communicate with one another to address specific tasks and retain the same objectives jointly. In MADIK, each intelligent software agent contains rules and knowledge which are based on the experiences of the investigation type. MADIK employs CBR to quickly identify the agents involved in a criminal investigation, allowing both agents to reason the facts in a more appropriate manner on a particular case. Here, the proposed MADIK has seven highly trained agents that are intelligently deployed:

- The Hard Disk Geometry Agent examines the bytes per sector and the number of sectors for any compromises.
- The File Path Agent keeps a list of frequently used files.
- The Keyword Agent is associated with the regular expressions used to find keywords.
- The File Signature Agent checks the file's first 8 bytes.
- The Hash Set Agent calculates and compares the MD5 hash.
- The Windows Registry Agent examines Windows registry-related files.
- The Timeline Agent investigates the development, access, and alteration dates.

The recommended ISA are a simplified collection that concedes to the system's configuration and the problems forensic investigators face daily in analysing evidence files, allowing the majority of the investigations to be carried out seamlessly. Embedding tool features such as Volatility, Scapy, Wireshark, and AccessData Forensic Toolkit (FTK) can also be used to make new agents. Furthermore, the agent's output can be improved using a case-based system.

IV. THE PROPOSED DFIR FRAMEWORK

We propose a framework for digital forensics and incident response (DFIR), as shown in Fig. 1. The DFIR framework describes the digital forensics lifecycle, the Java development framework, Case-Based Reasoning, Eclipse, Postgres Repository, and the embedded intelligent software agents needed for additional forensics investigations.

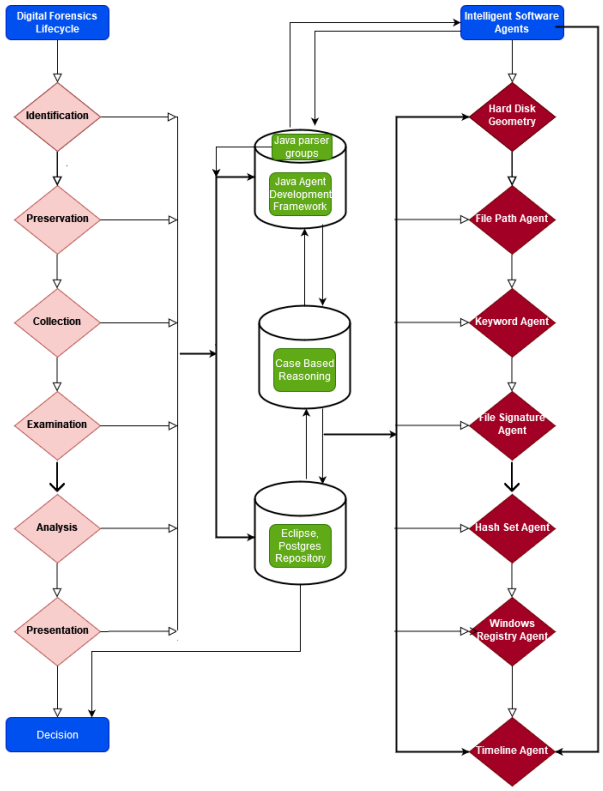

Fig. 1: Digital Forensics and Incident Response Framework

V. IMPLEMENTATION AND TESTING

The Java Agent Development Framework is used to implement the MADIK proof of concept. Eclipse, an interactive development platform for computer programming, was used to implement JADE. The Java Run Time Environment version 5 is the only software needed to run JADE, and also, the JADE source distribution includes Java parser groups. The agent's reasoning is also implemented using a Postgres database and a rule engine. JADE is chosen because it makes it easier to incorporate multi-agent frameworks on a distributed platform. Also, it has the capabilities to call Python, which enhances the integration of other tools we have discussed in this paper, such as Scapy, Wireshark, AccessData FTK, and Volatility. Jython is a Java-based Python implementation built around a Python interpreter that translates Python code into Java bytecode for execution, further supporting other plugins integrations.

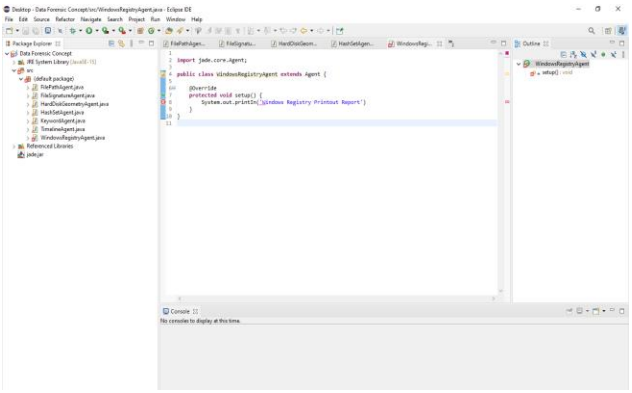

Fig. 2: Imported JADE-bin-4.5.0

By dragging and dropping the extracted JADE-bin-4.5.0 core Agent into the Data Forensic Concept, the JADE-bin-4.5.0 core Agent is imported into each Agent's code. A message appears on how the files should be imported; the copy files is selected, and then the okay button. The imported JADE core Agent is compatible with the classes when extended agent syntax is used on each agent's classes. However, the method setup is used on each Agent's codes to override JADE.

A. FTK Integrity Checks

The Lone Wolf E01 files were obtained from the Digital Corpora website and uploaded to an FTK virtual machine. An integrity check was carried out, taking approximately 48 minutes to upload and verify. The output result showed that the integrity check was 100% accurate. The hard disc geometry was 512 bytes per sector, for a total of 1,000,215,216 sectors. The MD5 Computed Hash Stored matches the MD5 Calculated recorded, indicating that the file is not abnormal. The SHA1 Computed value stored corresponds to the SHA1 Calculated value recorded.

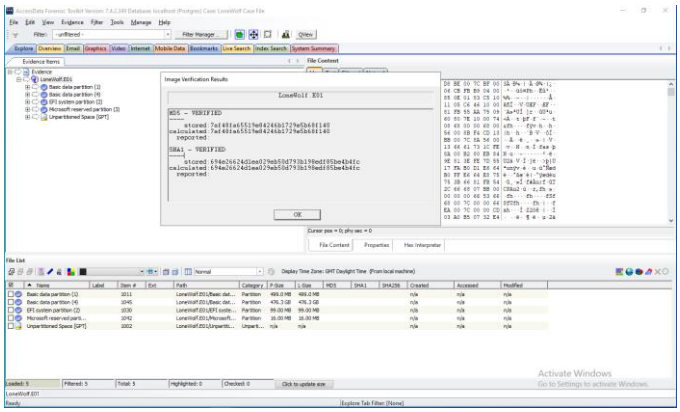

Fig. 3: Image Verification Results

VI. Results and Discussion

Experiments were conducted on MADIK to evaluate the effectiveness and integrity check of Lone Wolf image files. When carrying out an integrity check on the Lone Wolf Image file, there was a significant decrease in the execution times of the Hash Set Agent compared to when we executed the Hash Set Agent on the AccessData FTK. Furthermore, experiments were conducted to understand how the distribution impacted the intelligent software agent's interconnectivity. We experimented with 1–4 VMs running 1–8 Hash Set Agents. The output results indicated a 5%-time loss due to the burden imposed on the agents. The File Path Agent recommended deleting 1,510 tiny GIF files and ignoring 1,064 cookies. The Timeline Agent found different points of programme installation and executable files.

MADIK is one of the AI applications in computer forensics that provides the most up-to-date data. It assists investigators and practitioners during the investigation process to complete the examinations timely. The use of MADIK as an application of AI shows that it is an exciting approach for the utilisation of all available computational resources and is very helpful in reducing the time to conduct an examination. MADIK is not simply a substitute for forensic tools like the Scapy, Volatility, AccessData Forensic Toolkit and Wireshark but a proposed integrated framework with other agents to reduce the time frame to carry out forensic investigations. The integrity check carried on an image file using AccessData Forensic Toolkit took approximately 48 minutes (2,880 ms), whereas it took the MADIK system 16 minutes (960 ms).

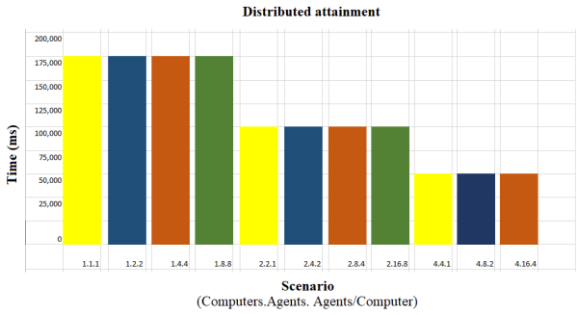

Fig. 4: Result Evaluation

VII. Conclusion

One of the areas of jeopardy regarding technological advancements is cyber-criminal activities. As the use of the internet and smart devices has increased, the risk of cyber-attacks has also increased. To address these issues and solve problems regarding cyber criminology, we introduced the term "artificial intelligence". In this research, AI is discussed along with its various types and classifications, like optimisation, machine learning, hybrid learning, deep learning, and reinforcement learning. The primary purpose of AI is to solve problems that humans cannot solve and provide evidence that can easily be overlooked when the data are examined separately. The correlated features show possibilities that reveal the reduction in time taken to check the integrity of evidence files, which could help reduce the presently overloaded caseloads facing law enforcement agencies around the world.

This paper focuses on the practical use of AI in digital forensic investigations. One of the applications of AI is MADIK systems, which generally refer to computational systems that are composed of more than one agent. It involves various techniques that are utilised in solving an investigation. The results identify variations in evidence that might be oversight due to the vast amount of data examined by an investigator. One suggestion for a future obligation is the combination and integration of MADIK with tools such as Access-Data, Scapy, Wireshark, and Volatility. It should focus on expanding the scope of the investigation and task automation and the system's architecture, making new agents a little easier.

REFERENCES

[1] Adams, R., 2013. The Emergence of Cloud Storage and the Need for a New Digital Forensic Process Model. *Cybercrime and Cloud Forensics*, pp.79-104.

[2] Agarwal, R. and Karahanna, E., 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, 24(4), p.665

[3] Biggs, S. and Vidalis, S., 2009. Cloud Computing: The impact on digital forensic investigations. *2009 International Conference for Internet Technology and Secured Transactions, (ICITST)*,.

[4] Damshenas, M., Dehghantanha, A., Mahmoud, R. and bin Shamsuddin, S., 2012. Forensics investigation challenges in cloud computing environments. *Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*,.

[5] Grosz, B. and Stone, P., 2018. A century-long commitment to assessing artificial intelligence and its impact on society. *Communications of the ACM*, 61(12), pp.68-73.

[6] Hoelz, B., Ralha, C. and Geeverghese, R., 2009. Artificial intelligence applied to computer forensics. *Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09*,.

[7] Ieong, R., 2006. FORZA – Digital forensics investigation framework that incorporate legal issues. *Digital Investigation*, 3, pp.29-36.

[8] Iqbal, S. and Abed Alharbi, S., 2020. Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics. *Digital Forensic Science*,.

[9] Kelly, L., Sachan, S., Ni, L., Almaghrabi, F., Allmendinger, R. and Chen, Y., 2020. Explainable Artificial Intelligence for Digital Forensics: Opportunities, Challenges and a Drug Testing Case Study. *Digital Forensic Science*,.

[10] Kim, T., Hooker, J. and Donaldson, T., 2021. Taking Principles Seriously: A Hybrid Approach to Value Alignment in Artificial Intelligence. *Journal of Artificial Intelligence Research*, 70, pp.871-890.

[11] Mittal, U. and Sharma, D., 2021. Artificial Intelligence and its Application in Different Areas of Indian Economy. *International Journal of Advanced Research in Science, Communication and Technology*, pp.160-163.

[12] Yan, D., Zhou, Q., Wang, J. and Zhang, N., 2016. Bayesian regularisation neural network based on artificial intelligence optimisation. *International Journal of Production Research*, 55(8), pp.2266-2287.