PriMera Scientific Engineering
Volume 8 Issue 1 January 2026
ISSN: 2834-2550

Using Artificial Intelligent Techniques to Standardize and Automate the Generation of Digital Forensic Reports

Type: Research Study
Received: December 03, 2025
Published: January 05, 2026

Citation:
Idani Mulaudzi., et al. "Using Artificial Intelligent Techniques to Standardize and Automate the Generation of Digital Forensic Reports". PriMera Scientific Engineering 8.1 (2026): 45-52.

Idani Mulaudzi* and Hein Venter

University of Pretoria, Engineering, Built Environment & IT, South Africa

***Corresponding Author:** Idani Mulaudzi, University of Pretoria, Engineering, Built Environment & IT, South Africa.

Abstract

The present study investigates if the digital forensics report can be generated automatically by using some of the artificial intelligence techniques, specifically the natural language processing. A model has been developed to assess if it is feasible to automate the generation of a digital forensic report using artificial intelligent techniques. One of the main purposes for this study is coming from a point where human errors, structure of the digital forensic reports, critical evidence that should take part of the digital forensic report are omitted during the generation of digital forensic report as well as the interpretation of the evidence drafted by an investigator during investigation. In addition, the standardization of this report happens to be imminent especially when it is being presented in a court of law. Given the rise of cybercrime, more research is needed to better improve the process of automating the generating digital forensic report using some intelligent techniques.

Keywords: digital forensics; artificial intelligence (AI); natural Language processing (NLP); digital forensic report (DF report); standardization

Introduction

The past few decades are commonly known to be decades of information technology, whereby interaction between individuals and multinational companies were simplified by the massive role out of communication technology. It is in this era wherein technology and its innovation has taken the world by a storm.

The immense rise of artificial intelligence (AI), machine learning, robotics processing and automation (RPA), blockchain, and internet of things (IoT) to name a few top technology trends, have dominated the way technology is needed and used. While the world is enabled by the luxury of having technology as a need, there has been an enormous rise in risks and threats associated with adoption and usage of these technologies.

As the rise of cybercrime and other malicious activities takes place in cyberspace, there has been immense use and reliability placed on a digital forensic report (DF report).

Courts of law and other forums has place reliance on the DF report due to the type of evidence they hold and present. These reports are guided by different standards such as ISO/IEC27037, 27042, 27043, and 27050. Furthermore, these standards allude to the reporting as a process. However, they do not provide the details that encompasses automation using AI techniques [6].

Generating the DF report refers to the process in which the digital evidence is gathered or obtained from the target system, and compiled in a structural format, be it done by an automated program or a human being [8]. The importance of the forensic report has been crucial in the legal fraternity. However, in many institutions that require the forensic report, it has also been noticed that the DF reports are generated and presented in different formats with different procedures and specifications, simply without standardized generation of the report. Consequently, there is no standardized process for generating a DF report.

Different tools and strategies are also in place to assist in extracting the digital evidence to build the forensic report. With the rise and reliability of technology, it has come to sense that there are different tools that make it easy to obtain digital evidence.

The remainder of this paper is constructed as follows; background on information security and cyber security, digital forensic (DF), and AI, critical evaluation and the conclusion. Following the introductory section, the next section provides the background starting with information security and cyber security.

Background

The background section is made up of the following sub-sections, information security and cybersecurity, digital forensic and AI. The next sub-section is information security and cybersecurity.

Information Security and Cyber Security

The field of information security (Infosec) has exponentially gained its relevance in Information Communication Technology (ICT) due to the nature in which data needs to be secured. Infosec and cyber security has become the matter of global interest due to the importance and the nature in which critical information and other types of data should be protected. Infosec and cyber security are often used interchangeably. However, these are two terms that substantially overlap, and they are not totally analogous [10].

Cybercrime has grown immensely over the years, becoming one of the most prominent crime concerns in the current era. With the continuous rise of cybercrime and its impact in different institutions, it has also become a subject to many organizations to combat the rise of cybercrime, wherein most of organizations have come up with different strategies to respond, detect, examine and investigate cybercriminal activities.

As a result of some of these strategies to reduce and to fight cybercrime, digital evidence is commonly used to prove and to identify the root cause, trace and analysis of the incident that occurred during the committed crime.

Therefore, this section gave a short review of the current literature on Infosec and cybercrime. The next section provides background on digital forensics

Digital forensics

This section provides the definition of digital forensics, some of the generic digital forensic process models. The digital forensic standards are briefly explained in this next section.

According to Awan (2022), digital forensics is a part of forensic science focusing on digital information produced, stored and transmitted by computer systems as a source of evidence in legal proceedings and investigations. Roussev (2009) stated that “digital forensics is the use of scientifically driven and proven strategies and methods towards collection, identification, preservation, analysis, documentation, presentation and interpretation of digital evidence emanates from digital sources for the purpose of investigation and legal proceedings”.

This also means that the interpretation of all collected evidence should be in a form that can be understood by anyone who happens to be involved in an investigation. This form is what is known as a DF report. There are numerous digital forensics process models. One of the first digital forensics process models is the Digital Forensics Research Workshops (DFRWS) investigative mode.

This process model was proposed and presented in 2001 at the first DFRWS conference. The said process model has different processes when comparing it with other digital forensics process models. The processes in this model are identification, preservation, collection, examination, analysis and presentation. It is important to note that the DFRWS is harmonized and eventually standardized in an international standard called ISO/IEC 27043 - Incident Investigation Principles and Processes.

The ISO/IEC 27043 international standard is imminent for this paper as it provides the baseline for the authors to determine where the AI techniques for automating the generation of the digital forensic report is to be applied. The ISO/IEC 27043 comprises of 4 processes (Readiness, initialization, acquisitive and Investigative process) and all these processes can run concurrently. Refer to figure 1 which illustrates the harmonized high-level representation of the ISO/IEC 27043 international standard.

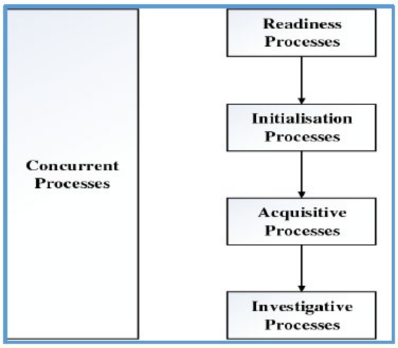

Figure 1: Harmonized high level ISO/IEC 27043 [23].

The AI technique to be used in the process of automatically generating digital forensic evidence is the Natural Language Processing (NLP). Due to NLP being able performs large scale analysis of data within a short space of time, provides accurate analysis and is more objective, and its ability to streamline processes and reducing costs to name a few of advantages, it makes NLP to be the most suitable AI to incorporate in the process of generating the digital forensic report automatically.

The next section expands on the short background which deals with AI, Natural Language Processing (NLP) specifically.

Natural Language Processing

AI-based applications are on the rise. AI is appearing to be dominating because of its ability of making it possible for technology to interact more logically and prudently with humans [20].

AI is one of the biggest game changers in technology and other sectors. AI can simply be defined as some kind of intelligence manifested or displayed by machines [22]. In the field of computing, AI is when machines or computer devices simulate human intelligence when solving problems, performing activities that are usually performed by human [22].

In AI, computers are being 'taught', i.e. similar as humans would be 'learning'. This process is known as 'machine learning'. For example, jet airliners rely more on the algorithms of the Flight Management Systems which contains the motion sensors, computer algorithms and global position systems to learn and track the position of the flight to make its own decisions on successfully flying the aircraft without human interaction. Therefore, the human element in flying a plane is limited and mostly the flight is spending more time flying itself (automated transportation/auto pilot).

Supervised learning is described by the usage of labelled datasets in predicting the outcome completely and accurately. Furthermore, supervised learning assists most organizations in solving various real-world problems such as classifying and identifying spam in different folders from email. Moreover, supervised learning involves a human being to be able to provide inputs on the learning process by manually manipulating some of the data to be processed by the algorithm.

In the previous sections, the background around information security and cybersecurity was discussed to gain an understanding of where the digital forensics investigations emanate from, digital forensics and the artificial intelligence technique namely the NLP process were also discussed to narrate down to the application of the AI in digital forensics process. The next section provides a model (high level and detailed) for the generation of digital forensic reports where the NLP is used to automatically generate DF report.

Model

The model in this section is derived from and extends the ISO/IEC 27043 international standard. However, this standard does not provide the technical guidelines to standardize the generation of a DF report, which led to the deployment of AI techniques. Furthermore, the process of generating the DF report is not yet standardized and usually the report is used in the court of law and for this paper, it is important to propose a solution that will guide and provide the procedures of generating such report for the report to be understood by anyone who happens to be using the report in the court of law.

Therefore, this section discusses and presents a high-level model as well as the detailed model for automating generation of DF report using NLP. The model consists of four different processes of the digital forensic process, namely:

1. ***Readiness process:*** taken exactly as from ISO/IEC 27043.
2. ***Initialization process:*** taken exactly as from ISO/IEC 27043.
3. ***Acquisitive process:*** taken exactly as from ISO/IEC 27043.
4. ***Investigative process:*** taken mainly from ISO/IEC 27043 and ISO/IEC 27037.

The naming and number of processes remain the same as in the said standards, however, there are changes within these processes, which call for the utilization/deployment of NLP in generating the DF report [20]:

As stated, intelligent techniques are used to automatically generate the DF report to save investigator's time and simplify the process of documenting the digital evidence, figure 2 provides a high-level model of how NLP is added and applied while expanding and adjusting some of the processes and sub-processes of ISO/IEC 27043 [23] and ISO/IEC 27037 [24] respectively. Please refer to figure 2 for the high-level model of how the NLP is incorporated.

1. ***Digital evidence acquisition:*** once the potential evidence has been collected from the seized device, the potential evidence is stored in a secured hard drive or an external drive where authorized members of the investigation can access the evidence.
2. ***Digital evidence analysis:*** the collected evidence is analyzed in this phase before being interpreted. In some instances, the forensic machine may be required whereby all stored evidence is analyzed using the digital forensics machine.
   For this paper, this is where AI is used for the relevant information or data from the evidence to be interpreted. Furthermore, NLP techniques will be used to filter, extract, and automatically generate a DF report.
3. ***Evidence interpretation:*** After the application of the NLP techniques, there is a need to interpret digital evidence before compiling the evidence into a report. The interpretation of digital forensics evidence is performed to ensure that the evidence can be understood by anyone who happens to be in the possession of the DF report.
4. ***Reporting:*** The reporting phase is whereby all the collected, analyzed and interpreted evidence is presented. In addition, this is the last phase where the DF report is being generated automatically.

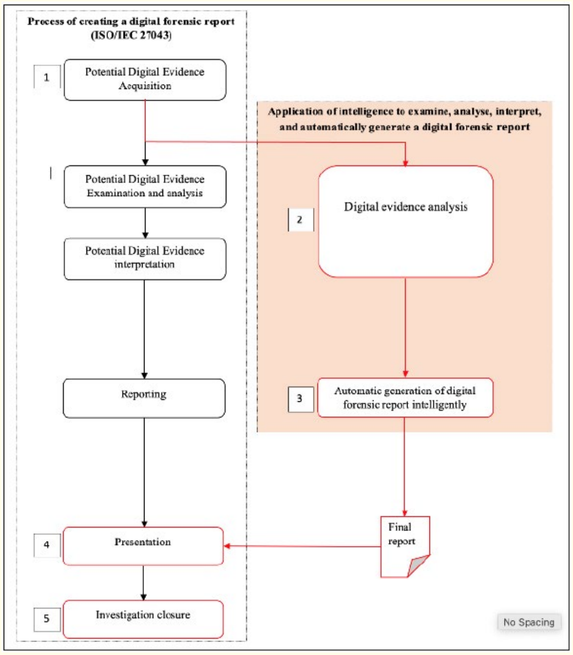
Process of creating a digital forensic report
(ISO/IEC 27043)
1
Potential Digital Evidence Acquisition
Potential Digital Evidence Examination and analysis
Potential Digital Evidence interpretation
Reporting
4
Presentation
5
Investigation closure
Application of intelligence to examine, analyse, interpret, and automatically generate a digital forensic report
2
Digital evidence analysis
3
Automatic generation of digital forensic report intelligently
Final report
No Spacing

Figure 2: High-level AI technique in the DF report process.

As this paper intends to propose an intelligent and automated solution to generate DF report it is important to understand that the generated report should be structured in a standardized reporting structure. Figure 2 provided a high-level overview of how the DF report is generated automatically using NLP.

The newly two introduced processes (intelligent digital evidence analysis and automatically and intelligently generating the DF report), each contain different subprocesses. Please note that for these sub-processes, the AI techniques are applied to reduce the manual intervention for the investigator during the generation of DF report. Thus, this is where the intelligent tool, using these intelligent techniques, performs most activities. The intelligent digital evidence analysis process is made up of the following three sub-processes:

- ***Intelligent process to evaluate digital evidence****:* The intelligent process to evaluate digital evidence is a sub-process on which the digital evidence is assessed whether the evidence is suitable enough to be analyzed intelligently.
- ***Intelligent process to examine digital evidence:*** In this sub-process, the acquired digital evidence is analyzed to ensure that all unwanted evidence that may be part of the collected digital evidence is excluded from the more useful evidence.
- ***Intelligent process to interpret digital evidence:*** within the digital evidence analysis where the collected evidence is interpreted in a more intelligent way.

The Automatically and intelligently generate the DF report process replaces the original reporting process from the (ISO/IEC 27043) international standard, and it is made up of the following four sub-processes which are dependent on each other, i.e., the output of one sub-process is the input of the next sub-process):

- ***Automated and intelligent process to arrange evidence by file types:*** Once the evidence has been interpreted intelligently, the evidence is arranged automatically according to their file types, for example, all text files are arranged according to their order, and all images are also arranged accordingly.

- ***Automated and intelligent process to arrange file types by last modified dates:*** This sub-process arranges the digital evidence according to the last modified date. In addition, if the file is recently updated or modified, it will be on top of the list of evidence. This is to ensure that files can be searched by date.
- ***Automated and intelligent process to separate required evidence (files) from the unwanted types:*** The intelligent tool on this sub-process separates the useful or most appropriate evidence from the unwanted or irrelevant digital evidence. For example, if the NLP tool is searching for files with specific values or names, it will exclude the files with irrelevant details.
- ***Final investigator's input on the automatically and intelligently generated report:*** In this sub-process, the investigator needs to have input by making sure that the report extracted all the needed evidence. This is the only part of the intelligent and automated processes that still requires the manual input of the investigator. Therefore, the investigator needs to review the automated generated report and ensure that the automated processes included all the details that need to be in the report. If some changes or modifications should be made to the automated report, then the investigator can do so at this stage.

The implementation and use of the AI tool to automatically generate a DF report has been displayed in figure 2.

The newly proposed model has its limitations and benefits which are stated in the next section. The next section highlights the critical evaluation for the proposed model.

Discussion or critical evaluation

As a proposed digital forensic model to automatically generate a DF report intelligently, it also comes with its own benefits and limitations. This section evaluates the research conducted for the purposes of this paper and provides a critical evaluation of the contribution made by this research.

The proposed model to automate the generation of DF report intelligently is to assist a digital forensic investigator to be able to generate a DF report quicker and have the minimal manual intervention as far as possible.

It may be difficult for the model to deal with password protected files. Cracking encryptions or recovering passwords can be time-consuming or even impossible to certain extent where the encryption is too long. This problem, however, is a general problem with digital forensic investigations in general, and not unique to this study.

In case of files that are corrupted, if a seized device has a corrupt file system or damage sectors, it can be challenging to retrieve data, and some information may be lost or unrecoverable. Furthermore, data fragmentation is a challenge as large files or file system with significant fragmentation can make extracting data more time consuming and complex. This may require reassembling fragmented data.

Perpetrators may use anti-forensic techniques to hide or destroy digital evidence, and this could include tools or methods to delete files securely, alter timestamps or obfuscate data which can make it difficult for the proposed model to examine content. Like the previous issue mentioned above, this problem is also a general problem with digital forensic investigations, and not unique to this study.

One of the benefits of using this model is to save the investigator`s time during DF report. Deploying AI techniques also assist in extracting what human/an investigator may omit during the investigation, and this will not be intentional but will be due to the amount of data being evaluated. This model is also beneficial because it eliminates a lot of mistakes that may be human errors when generating a DF report. In addition, to human error, there are few things such as exhaustion and fatigue while working on large amount of data, but with this model such will be eliminated and will eventually speed the process of generating DF report effectively and efficiently with minimal or sometimes no errors.

Conclusion

The proposed model seeks to improve the process of generating digital forensic reports using AI techniques. This model also standardizes the way in which DF reports will be generated using AI techniques (NLP). In addition, the model being proposed in this study is derived and guided by the ISO/IEC 27043 International Standard which focuses on providing “guidelines based on idealized models for common incidents investigation processes across various incident investigation scenarios involving digital evidence” [20].

As the process of generating the DF report has been in the center of how the digital findings should be presented in legal fraternity and other institutions that make use of the DF report, it has been generated in different structures, formats with different specifications while guided by different standards and procedures, which raised the question of not having a standardized procedures and specifications to generate such report.

Taking into consideration different standards such as ISO/IEC 2737, 27042, 27043 and 27050 which promotes good practice and processes for digital forensic and investigation capture of digital evidence. The author noted a need to propose a solution to automate the generation of DF report intelligently with the benefit of standardizing this report. This was further motivated by the fact that none of these internation standards goes into detail on how to draft a proper DF report.

Considering the relevance and importance of the DF report, there is a need to standardize the way in which the DF report is generated as this will enable different report interpreters to be able to understand and make use of reports where necessary.

With the proposed model to solve the stated problem, there is still more improvements that needs to be taken into consideration in standardizing the generation of DF report using other different types and techniques of AI that can potentially replace forensic investigator as the proposed model still require some level of manual intervention by the digital forensics’ investigator.

References

1. Adeyanju IA., et al. “Machine learning methods for sign language recognition: A critical review and analysis”. Intelligent Systems with Applications 12 (2021): 200056.
2. Awan SA., et al. “Digital Forensics and Cyber Forensics Investigation: security challenges, limitations, pen issues, and future direction”. International Journal of Electronic Security and Digital Forensics 1.1 (2022): 1.
3. Buhalis D and Law R. “Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research”. Tourism Management 29.4 (2008): 609-623.
4. Casey E and Stellatos GJ. “The impact of full disk encryption on digital forensic”. Operating Systems Review 42.3 (2008): 93-98.
5. Chan AW., et al. “SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials”. BMJ 346 (2013): e7586.
6. Collobert Ronan., et al. “Natural language processing (Almost) from scratch” Journal of Machine Learning Research (2011).
7. Costantini S, De Gasperis G and Olivieri R. “Digital forensics and investigations meet artificial intelligence”. Annals of Mathematics and Artificial Intelligence 86.1-3 (2019): 193-229.
8. De Alwis C., et al. “Survey on 6G frontiers: trends, applications, requirements, technologies and future research”. IEEE Open Journal of the Communications Society 2 (2021): 836-886.
9. Fahlevi M., et al. “Cybercrime Business digital in Indonesia”. E3S Web of Conferences 125 (2019): 21001.
10. Fuchs K. “Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse”. Frontiers in Education (2023): 8.
11. Goh OS., et al. “Top-down natural language query approach for embodied conversational agent”. International Multiconference of Engineers and Computer Scientists (2006): 470-475.
12. Hirschberg J and Manning CD. “Advances in natural language processing”. Science 349.6245 (2015): 261-266.
13. Roussev V. “Hashing and data fingerprinting in digital forensics” IEEE Security & Privacy 7.2 (2009): 49-55.
14. Shahbazi Z and Byun YC. “NLP-Based Digital Forensic Analysis for Online Social Network based on System Security”. International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health 19.12

(2022): 7027.

15. Shalaginov A, Iqbal A and Olegård J. "IoT Digital Forensics Readiness in the Edge: A Roadmap for Acquiring Digital Evidence from Intelligent Smart Applications". in Lecture notes in computer science (2020): 1-17.
16. Siponen MT. "A conceptual foundation for organizational information security awareness". Information Management & Computer Security 8.1 (2000): 31-41.
17. Van Niekerk JF Von Solms R. "Information security culture: A management perspective". Computers & Security 29.4 (2010): 476-486.
18. Von Solms R and Van Niekerk J. "From information security to cyber security". Computers & Security 38 (2013a): 97-102.
19. Artificial intelligence in online shopping using natural language processing (nlp)'. Journal of Critical Reviews 7.4 (2020).
20. Kothapalli S and Appavu R. "The application of artificial intelligence and machine learning to anesthesiology". Journal of Student Research 12.2 (2023).
21. SANS 27043, information technology - Security techniques - incident investigation principles and processes.
22. SANS 27037, Information technology — Security techniques — Guidelines for identification, collection, acquisition, and preservation of digital evidence.