# Artificial Intelligence in Digital Forensics: Augmented Analysis and Emerging Evidence

Albi Marini, Danny Trainor

## Introduction

The field of digital forensics is ever-evolving, constantly seeing new sophisticated threats with increasing amounts of data to analyze. Manual analysis has become a thing of the past in this modern era, as digital forensic analysts are relying more and more on artificial intelligence and machine learning technologies for their investigations. Professionals in the field have touted how these technologies "amplify the efficiency and precision of digital forensic investigations". (Dunsin et al., 2024) Many popular tools now integrate AI & ML to help the identification, collection, and analysis steps of the forensic process (Hagan, D. K., 2021). At the same time. The rise of AI technology in consumer products has resulted in new types of digital "artifacts", like AI-generated or "genAI" images and logs of large language model (LLM) agent prompts that can be used as evidence (Yin et al., 2025).

Our paper aims to explore these two themes of the role of artificial intelligence in digital forensics:

(1) How LLMs and other AI tools are enhancing and creating new digital forensic tools across storage, network, and volatile memory analysis

And

(2) How AI-generated artifacts constitute a new realm of digital forensic evidence, with challenges emerging in their reliability in an investigation

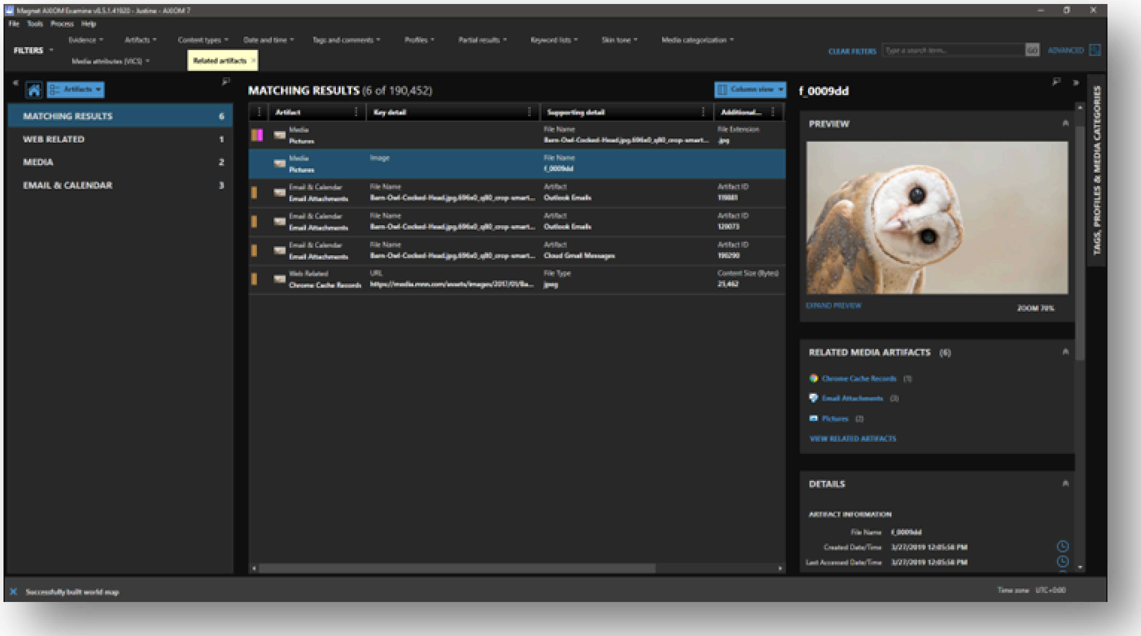
By examining widely used tools like Magnet AXIOM, Darktrace, and Cellebrite Pathfinder as well as conducting a review of recent scholarly papers regarding the topic, we aim to highlight the capabilities that the AI boom has added to digital forensics, as well as identify the risks associated with these innovations.

## AI-Enabled Forensic Tools and Workflows

Across the industry, modern forensic software suites have started to integrate AI into their tools, especially in order to automate incident identification, evidence discovery, and analysis. Below, we present key examples of such tools and their capabilities:

- **Magnet AXIOM,** as part of the **Magnet Forensics Suite,** mobile chat log evidence discovery
    - As of 2020, Magnet AXIOM is an extremely popular discovery tool used as part of a forensic investigator's toolbox to "recover, analyze, and report digital evidence from mobile devices and cloud-based applications". In terms of US investigations, AXIOM has been supplied to "dozens of federal agencies", with "Both ICE and CBP" having "renewed their license every year since 2014" (AFSC Investigate, 2021)

- *Magnet.AI* has been a part of the Magnet AXIOM software since April of 2018, using "machine learning to automatically categorize chat and pictures for things like drugs, guns, nudity, abuse, luring, or sexual conversations" (Maxiom, 2018). Adopting AI technology relatively early, Magnet has continually revised its software over the years to keep up with advancements in the industry.



MATCHING RESULTS 6
WEB RELATED 1
MEDIA 2
EMAIL & CALENDAR 3

MATCHING RESULTS (6 of 190,452)

| Artifact | Key detail | Supporting detail | Additional... |
|---|---|---|---|
| Media<br>Pictures | | File Name<br>Barn-Owl-Cocked-Head.jpg.696x0_q80_crop-smart... | File Extension<br>jpg |
| Media<br>Pictures | Image | File Name<br>f_0009dd | |
| Email & Calendar<br>Email Attachments | File Name<br>Barn-Owl-Cocked-Head.jpg.696x0_q80_crop-smart... | Artifact<br>Outlook Emails | Artifact ID<br>119881 |
| Email & Calendar<br>Email Attachments | File Name<br>Barn-Owl-Cocked-Head.jpg.696x0_q80_crop-smart... | Artifact<br>Outlook Emails | Artifact ID<br>120073 |
| Email & Calendar<br>Email Attachments | File Name<br>Barn-Owl-Cocked-Head.jpg.696x0_q80_crop-smart... | Artifact<br>Cloud Gmail Messages | Artifact ID<br>190290 |
| Web Related<br>Chrome Cache Records | URL<br>https://media.mnn.com/assets/images/2017/01/Ba... | File Type<br>jpeg | Content Size (Bytes)<br>21,462 |

f_0009dd

RELATED MEDIA ARTIFACTS (6)

Chrome Cache Records (1)
Email Attachments (3)
Pictures (2)

| | |
|---|---|
| File Name | f_0009dd |
| Created Date/Time | 3/27/2019 12:05:58 PM |
| Last Accessed Date/Time | 3/27/2019 12:05:58 PM |

Figure 1: Example of *Magnet Copilot* automatically identifying and categorizing evidence from a device image in the *Artifact Explorer* of Magnet AXIOM

- The latest versions of the software (circa Mid 2024) include *Magnet Copilot*, an AI assistant that provides offline data recognition and media authenticity capabilities using computer vision models like R-CNNs to provide notifications to the investigator when related media is found after scanning the device image. (Ciligot, C., 2025) By providing these automatic categorization and recognition capabilities, examiners no longer have to peruse device images to search for evidence, reducing the time of discovery

- **Cellebrite Pathfinder,** as part of the **Cellebrite** suite, digital evidence collection and analysis tool
  - Founded in 1999, Cellebrite and its suite of tools are largely popular across law enforcement agencies across the globe as part of their forensic investigation process. Cellebrite provides multiple key pieces of software, including Insights for mobile device data extraction, Guardian for cloud-based evidence sharing and review, as well as Pathfinder for data collection and correlation. They claim that their technology is in the hands of "6,900 public safety agencies, enterprises in over 100 countries, and

used in more than 5 million cases” in federal, enterprise, state, and local contexts. (Cellebrite, 2025)

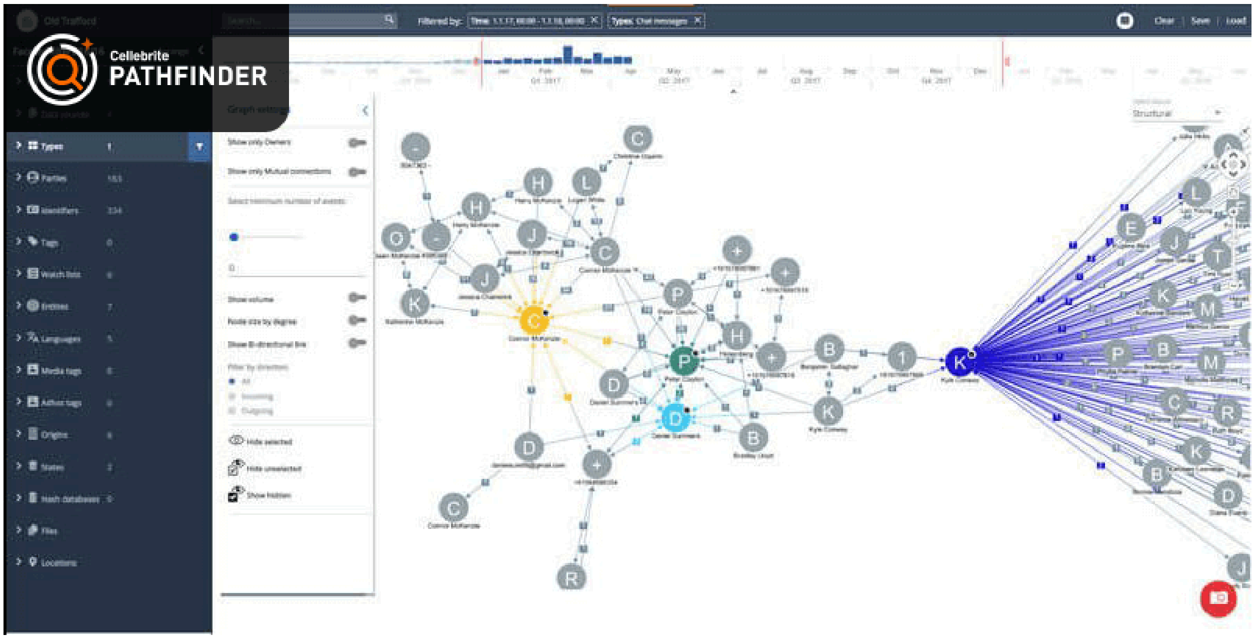


Figure 2: Example of Cellebrite Pathfinder’s graph interface that automatically collects and draws connections between data points from multiple data sources to create a case-web of involved parties

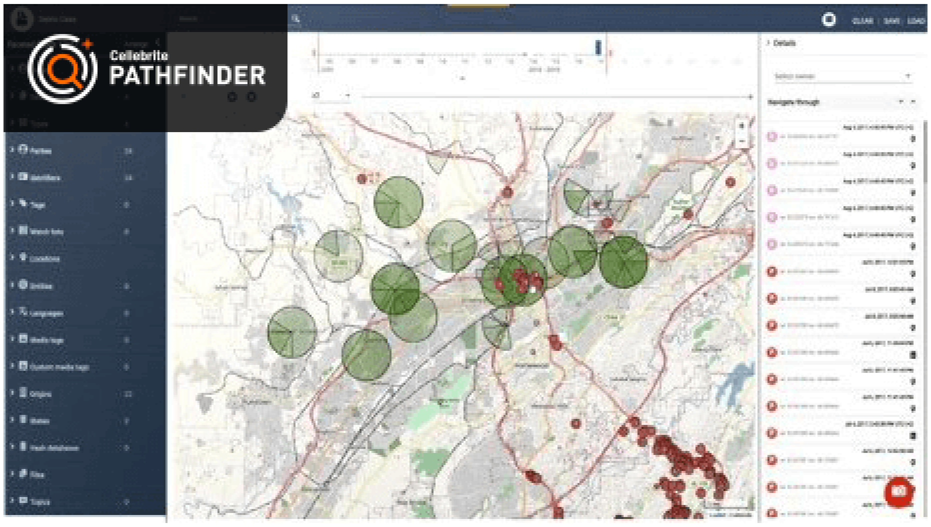


Figure 3: Automatic geo-location identification of where investigative data points occurred in Cellebrite Pathfinder

- ○ Since May of 2021, Cellebrite has provided its Pathfinder software as a platform to “automate the ingestion and analysis of digital evidence” (Cellebrite, 2021). Working on both mobile and desktop devices, Pathfinder uses a “multilingual transformer-based language model” to analyze chat communication across various platforms as well as on-device media. (Cellebrite, 2021). Additionally, Pathfinder leverages facial recognition technologies to “identify people, not numbers” (Cellebrite, 2021). to consolidate data from “an unlimited amount of sources” (Cellebrite, 2025b) into a dashboard-like interface with graph visualization, connecting key investigative points from the suspect and victim together. It automatically links these data points to GPS data from photo geo-tags, cell tower dumps, or Wi-Fi hotspots to construct a full picture of when and where an incident occurred. (Cellebrite, 2021) Similar to Magnet AXIOM, this AI-enabled software works to streamline the storage forensic process, cutting down the need for manual review by perusing large amounts of data in fractions of the time it would take a team of investigators to do by hand

- **Darktrace**, an AI cybersecurity platform, offers network forensics.

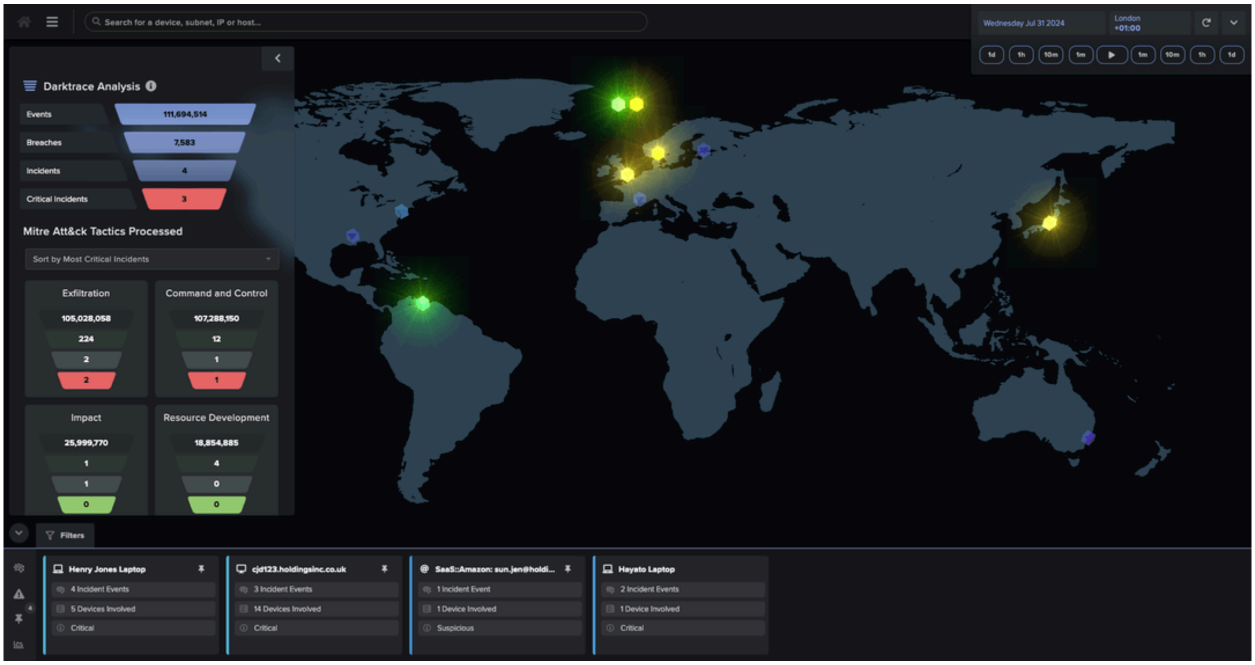


Figure 4: Example of the Darktrace network tool visualization interface that provides a map of network nodes, as well as their state and log of previous incidents

  - Being a newer company (2013), Darktrace is an example of an AI-enabled startup that offers various services not previously possible before the modern advent of LLMs. Although targeted toward cybersecurity, their network detection & response solution, powered by “*self-learning AI*” through reinforcement learning strategies (Darktrace, 2025), proves useful in the realm of network forensic analysis.
  - Their AI-powered platform provides a visualization interface for network incidents, automatically analyzing “every connection, device identity and attack path” within the network to uncover blind spots, and “save you the hassle of manual tuning” (Darktrace, 2025). In particular, whenever an incident occurs, the *self-learning AI/Cyber AI* automatically triages, interprets, and provides a report of the incident. They provide the example of a ransomware outbreak stating how Darktrace’s

software can trace the spread across the network node to provide an investigator with a list of affected machines and a complete attack timeline, useful in a digital forensics context where this volatile network data can often be lost if not captured in real-time. (Darktrace, 2020)

- **Large-Language Model Assisted Analysis**
  - According to a paper published by Cornell University in 2025 (Yin et al., 2025), other than particular AI features integrated into digital forensics targeted tools, LLMs have become widely used across the field across all steps of the investigative process (Yin et al., 2025). At the time of the paper, the authors mentioned how LLMs like ChatGPT-3.5, 4.5 and Google Gemini “process enormous volumes of text-based evidence” like chat logs, documents and emails much faster than humans could ever do, sifting through these pieces of evidence to categorize text, recognize patterns, and connect names with timestamps or address to “identify complex interrelationships” (Yin et al., 2025).
  - The researchers further present an example of LLM-driven Mobile Evidence Contextual Analysis, where they showed how mobile chat logs can be fed to an LLM to perform contextual analysis by finding criminal intent through pattern identification or coded language in messages that go beyond just simple keyword matching.

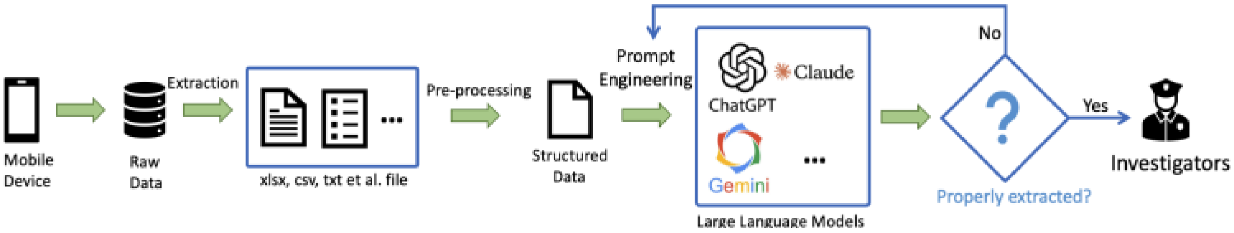


Figure 5: Overview of LLM-driven Mobile Evidence Contextual Analysis Framework

  - These LLM approaches to digital forensics have led to the creation of some open-source technologies that showcase the potential of these technologies. Such an example is *volGPT*, the “first prompt-based large language model for memory forensics, providing analysts with triage automation and explanations for triage reasons” using GPT-like reasoning models as a base (Oh et al., 2024). According to the software’s proposal, *volGPT* proved very effective on ransomware infected memory dumps from Windows systems, achieving a detection accuracy of “up to 99% and a minimum of 87% in five ransomware families” with a high accuracy and tendency to avoid misidentifying benign processes as suspicious (avoid false positives) (Oh et al., 2024). This was even in the face of such malware hiding itself through process masquerading. The software would take a memory image, filter down the list of suspicious processes, and provide text explanations for each software as to why it was flagged as suspicious. Although still new, the example framework set by *volGPT* suggests that AI can not only assist investigators in static storage or network forensics, but also in volatile memory forensics, being much faster than current memory heuristic analysis tools that rely on deep operating system analysis to be effective. This could potentially open the door to memory forensics to be more easily accessible and useful in forensic investigations, speeding up the process and simplifying the complex task.

As we found above, the integration of AI in digital forensics tools has fundamentally changed the investigative workflow, potentially freeing up time in an investigation. Tasks like data discovery and analysis across all types of digital evidence can be automated and accelerated by AI & ML, allowing analysts to focus on the interpretation of said data. These AI tools even give investigators various hypotheses and explanations as to what and how something occurred, providing leads that they can further chase.

However, the benefits of this age of AI-driven forensics are not necessarily without risks or challenges. A recent paper published by Cornell University in December of 2024 on the Robustness of AI-Driven Tools in Digital Forensics found that (Sanna et al., 2024) these AI models do not guarantee perfect accuracy in their findings as they can overlook evidence or give false insight (ie, flag things as false negatives or false positives). Such is an inherent drawback of all AI tools, and is one that is continually advanced upon as time goes on and machine learning practices improve, but in the context of forensics, these risks are very dangerous. The authors of this paper focus on an experimental study on how AI-based digital forensic tools behave when tasked to recognize nudity in images, or deep fakes. In their findings, they discussed how these tools are “not robust enough” as they reported a high number of misclassifications in terms of missing some nude images, and misclassifying deep fakes as the real person, even if obvious to a human (Sanna et al., 2024).

Some of these errors could be due to the bias of the dataset the algorithms are trained on (WebAsha, 2025), for example, if a certain model was trained on more images of a certain ethnicity, it might perform poorly on others. In terms of text conversations, bias could show itself in terms of the AI being more likely to classify text conversations of certain groups of people as dangerous due to unawareness of context or cultural references (WebAsha, 2025). Even then, as shown by the previously mentioned paper, smart adversaries could always alter images and text to avoid AI detection in “anti-forensic attacks” so that they are altered in a way that the classification algorithms in these models do not detect potentially harmful content (Sanna et al., 2024). The inherent error and bias that come with training a lot of these machine learning models and AI tools showcase how, even despite rigorous testing and validation, these errors in classification models, investigators cannot simply take a piece of data flagged as relevant and use it as proof as part of their investigation. It is part of their job to make sure that every piece of evidence they present is the truth and has no holes, something of which current AI & ML algorithms do not guarantee 100% of the time, especially as many pieces of industry software act as closed source black boxes that do not disclose their evidence classification process (Yin et al., 2025). As these tools become more and more commonplace, it would be naive to think that in the legal context of digital forensics AI tools could completely replace a human investigator at each step of the forensics process – ultimately they are still a tool and one whose insight might need to be further vetted than that of non-AI enabled tools.

### LLM Prompt Logs/AI Interaction Data

The emergence of LLM technology has opened the door for new forensic considerations. Much like a disk image having information such as time stamps, file system information, etc. large language model technologies have introduced their own ways to have markers for investigative purposes.

- As a means to brace LLMs for possible digital forensic investigations, invocation logs- the logs of the user interactions with the model- may be collected. Given an application service that may use a model like GPT or Gemini, user prompting of the service can be collected and fed back into the model to be summarized, analyzed, etc.

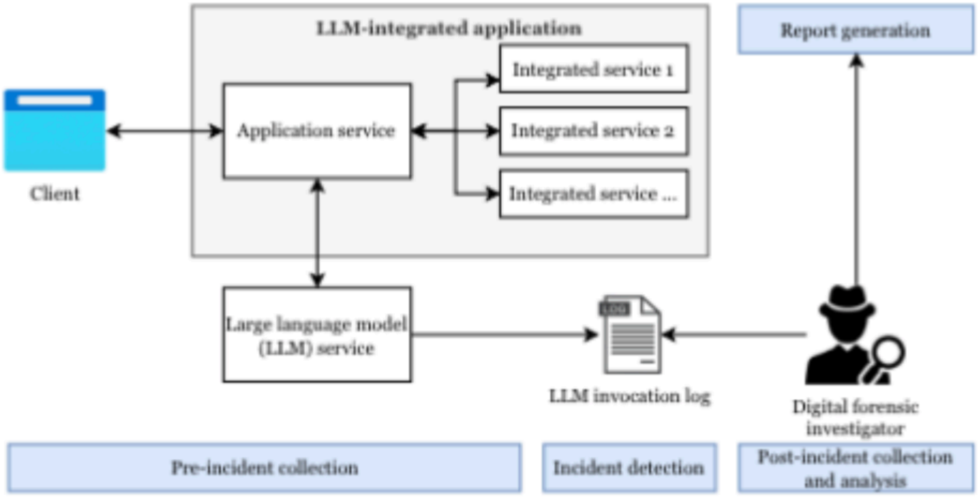


Figure 6: Analysis of LLM-generated Invocation Logs

LLM invocation logs, as outlined in the image above, can be seen as a measure taken to ensure that a digital forensic investigation could be done if needed. The generated logs are evidence that can be taken before any sort of investigation is prompted to take place, and can have applications in limiting attacks such as prompt injections or denial-of-service attacks. Through analysis of invocation logs, whether it be through a human or ML/AI-based approach, forensics professionals can find malicious patterns in the collected evidence relevant to the investigation. Take, for instance, the idea of a prompt injection, which aims to “trick” an LLM into deviating from its system behavior, potentially revealing sensitive information it may have been trained on. Invocation logs can help illustrate the types of prompting that would be used to “inject” these systems and trick them. It is important to note that there is the added benefit of the fact of real-time collection of these logs as the malicious prompting is happening. (Chernyshev et al., 2025)

### Deepfake Technology, Authenticity, and Anti-Forensics

Artificial intelligence and similar technologies yield not only more sources of digital evidence, but also more avenues for anti-forensic and/or blatantly malicious digital activity. Deepfakes and AI-generated media are becoming increasingly more prevalent in the modern day and are a foremost example of how these technologies have come to be forensic evidence themselves. Additionally, the growing significance of this type of media brings to light the forensic challenge of authenticating it. Below, we present some examples of AI media technology and generative anti-forensic technology.

- **Deepfake technology** has emerged as one of the most popular and controversial AI-generated media phenomena. The technology, which gets its name from “deep learning” and “fake,”- can manipulate trained audio, picture, and video data to synthesize the likeness of a human being. Deepfake technology works particularly well on individuals who have a significant amount of publicly available auditory and visual data, like celebrities, public figures, and world leaders. It can be limited, however, particularly in video media that involve a more full-body depictions of someone. Deepfake technology has a relatively “static” way of only depicting full-frontal facial features and movements (take, for instance, a speech by a public figure), while features like the back of one’s head, or their body are harder to manipulate through deepfake technology. One workaround has been the “face-swapping” of a deepfaked face onto a lookalike’s body. (Mirsky, Lee 2022)

- The uptick in AI-generated media has brought forth concern for whether presenting media as evidence is enough in itself. **Authenticity** is a primary concern when looking at digital media evidence. There are many different ways in which the authenticity of generated media can be examined. Pixel-level anomalies and unnatural lighting patterns, constitute some of the spatial considerations to be made when examining digital media evidence, while biological markers may include abnormal eye movement and blood flow changes in the face. (Armerini et al., 2025)

- **Generative adversarial networks (GAN's)-** which are often the building blocks of deepfakes themselves- have uses in both forensics and anti-forensics. GAN's are machine learning frameworks that aim to create realistic, synthetic data. GAN's, which can be used to manipulate images (as seen through deepfake) can exploit forensic classifiers by design and manipulate in a way that attacks the forensic algorithm itself. GAN attacks have been proposed to remove tools such as median trace filters over an image through a GAN that was trained to remove the median filter traces and then discriminate between the altered and unaltered images. In 2018, a GAN-based attack, MISLGAN, forged source camera models of images by modifying traces in the images source with a high success rate. The trained generator was able to even forge camera models of images that were not trained on a pre-trained forensic camera model CNN (convolutional neural network). (Stamm et al., 2022)

## Conclusion

The deep learning/artificial intelligence boom of the 2010's and 2020's has led to many developments that increase productivity and efficiency. Such improvements lend themselves to digital forensics, digital forensic evidence and anti-forensics. High-level deepfake technology yields hyperrealistic depictions of one's likeness and can be used to deceive and influence, particularly in the advent of large-scale social media platforms. Invocation logs encompass both AI tools as evidence and vessels for augmented analysis. Whether it be improvements to existing tools like Magnet AXIOM or the malicious manipulation of images to blur forensic evidence, these technologies add a level of complication to an already multifaceted field. As this technology continues to develop and expand, forensics professionals must learn to adapt their tools and methods. Similarly, perhaps beyond the scope of this paper, the legislation of artificial intelligence is ambiguous and slow-moving. The problems with lawmaking regarding these issues are often representative of a greater issue in digital forensics and computing in general- access to education and know-how of the tools and technologies being legislated. The future of this technology in the cybersecurity, forensic, and greater digital realm will rely on the joint effort of practitioners and lawmakers alike.

## References

1. Amerini, I.; Barni, M.; Battiato, S.; Bestagini, P.; Boato, G.; Bruni, V.; Caldelli, R.; De Natale, F.; De Nicola, R.; Guarnera, L.; et al. Deepfake Media Forensics: Status and Future Challenges. *J. Imaging* **2025**, *11*, 73. https://doi.org/10.3390/jimaging11030073
2. Chernyshev, M., Baig, Z., Doss, R.R.M.: Towards large language model (LLM) forensics using llm- based invocation log analysis. In: Proceedings of the 1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis, pp. 89–96 (2023)
3. Dunsin, D., Ghanem, M. C., Ouazzane, K., & Vassilev, V. (2024). A comprehensive analysis of the role of Artificial Intelligence and machine learning in modern digital forensics and incident response. *Forensic Science International: Digital Investigation*, *48*, 301675. https://doi.org/10.1016/j.fsidi.2023.301675
4. Hagan, Dennis. (2021). Digital Forensics -The Processes and Some Forensic Tools Digital Forensics. *Webster University*. 10.13140/RG.2.2.23879.10402.
5. Yin, Z., Wang, Z., Xu, W., Zhuang, J., Mozumder, P., Smith, A., & Zhang, W. (2025). Digital Forensics in the Age of Large Language Models. *ArXiv*. https://arxiv.org/abs/2504.02963
6. AFSC Investigate. (2021). *Magnet Forensics Inc*. Magnet Forensics Inc | AFSC Investigate. https://investigate.afsc.org/company/magnet-forensics
7. Magnet axiom 2.0 - magnet.AI. Magnet Forensics. (2018). https://www.magnetforensics.com/resources/magnet-axiom-2-0-magnet-ai/#:~:text=Magnet%20AXIOM%202
8. Ciligot, C. (2025, March 20). *Magnet axiom 8.6: Magnet Copilot Offline AI capabilities and more*. Magnet Forensics. https://www.magnetforensics.com/blog/magnet-axiom-8-6-magnet-copilot-offline-ai-capabilities-and-more/#:~:text=While%20AI%20can%20be%20used,giving%20you%20the%20ability%20to
9. *About - cellebrite*. Cellebrite Forensics. (2025). https://cellebrite.com/en/about/
10. *Cellebrite unveils New Pathfinder release, designed to securely integrate, manage, and drive actionable insights from data to meet Mission & Investigation Objectives - Cellebrite*. Cellebrite Forensics. (2021). https://cellebrite.com/en/cellebrite-unveils-new-pathfinder-release-designed-to-securely-integrate-manage-and-drive-actionable-insights-from-data-to-meet-mission-investigation-objectives/

11. *Accelerate justice with cellebrite*. Cellebrite Forensics. (2025b). https://cellebrite.com/en/home/
12. Darktrace. (2025). *Network security management: AI Network Security Protection*. Network Security Management | AI Network Security Protection. https://www.darktrace.com/products/network
13. *Darktrace's AI analyst: Closing the Cyber Skills Gap*. Darktrace. (2020, February 25). https://www.darktrace.com/de/blog/bridging-the-cyber-skills-gap-cyber-ai-analyst-for-ot#:~:text=Darktrace%27s%20AI%20Analyst%3A%20Closing%20the,related%20alerts%20and%20useful
14. Oh, D. B., Kim, D., Kim, D., & Kim, H. K. (2024). Volgpt: Evaluation on triaging ransomware process in memory forensics with large language model. *Forensic Science International: Digital Investigation*, *49*, 301756. https://doi.org/10.1016/j.fsidi.2024.301756
15. Sanna, S. L., Regano, L., Maiorca, D., & Giacinto, G. (2024). Exploring the Robustness of AI-Driven Tools in Digital Forensics: A Preliminary Study. *ArXiv*. https://arxiv.org/abs/2412.01363
16. Stamm, M.C., Zhao, X. (2022). Anti-Forensic Attacks Using Generative Adversarial Networks. In: Sencar, H.T., Verdoliva, L., Memon, N. (eds) Multimedia Forensics. Advances in Computer Vision and Pattern Recognition. Springer, Singapore. https://doi.org/10.1007/978-981-16-7621-5_17
17. WebAsha Technologies. (2025, February 28). *AI in Digital Forensics: A revolutionary breakthrough or a risky gamble?* https://www.webasha.com/blog/ai-in-digital-forensics-a-revolutionary-breakthrough-or-a-risky-gamble#:~:text=,Biased
18. Yisroel Mirsky and Wenke Lee. 2021. The Creation and Detection of Deepfakes: A Survey. ACM Comput. Surv. 54, 1, Article 7 (January 2022), 41 pages. https://doi.org/10.1145/3425780