Unraveling Digital Mysteries: How AI Copilots can Revolutionize Digital Forensic Investigations[1]

By Hans Henseler, Professor of Digital Forensics & E-Discovery, University of Leiden Applied Sciences, and Senior Digital Forensic Scientist at the Netherlands Forensic Institute.

Introduction

In hindsight, 2021 was a significant inflection point in the world of artificial intelligence, characterized by remarkable developments in deep learning, manifesting in models such as DALL·E, CLIP and in models that were surpassing GPT-3 in size and ability. These advancements hinted at a future not limited to machines performing computational tasks but also emulating intricate human-like activities. However, it was November 2022, with the emergence of ChatGPT, that the world glimpsed a truly transformative tool, suggesting potential applications even in niches like digital forensics [1,2].

Yet, the digital forensics community, by and large, has yet to fully embrace or debate the profound implications of these advancements. Every case in digital forensics presents its own universe of data, sourced from confiscated devices, cloud accounts, and other digital touchpoints. Parsing through this enormity, especially with looming backlogs in forensic labs and the aspirations to involve detectives without specialized forensic training, demands a radical rethinking of our tools and approaches.

Instead of merely focusing on the limitations or potential pitfalls of Large Language Models (LLMs), we ought to explore their promise. Retrieval-Augmented Generation (RAG) is one such promising frontier. By coupling real-time data retrieval with the robust capabilities of generative models, RAG offers a compelling case for the next evolutionary step in digital forensics. This article emphasizes not just the

[1] Published by eForensics Magazine https://eforensicsmag.com/unraveling-digital-mysteries-how-ai-copilots-can-revolutionize-digital-forensic-investigations/

challenges but also the transformative potential of AI for forensic experts and investigative detectives alike.

Understanding the gaps in LLMs

GPT-4 and ChatGPT are heralded for their unparalleled capacity to understand and generate content with a human touch. However, as much as these tools have revolutionized the AI landscape, they aren't without shortcomings. Central to their limitations is their fixed knowledge base; after their training phase, they cannot easily update or expand their knowledge, e.g., on seized data.

Furthermore, while LLMs sometimes produce information not strictly based on their training data—a phenomenon known as 'hallucination'—this can be a critical concern in digital forensics. While such creative outputs might be advantageous in brainstorming or strategic simulations, it's a potential hazard in digital investigations where precision and evidence integrity are paramount.

Harnessing the Power of RAG

While RAG was introduced as early as 2020, its recent integration with LLMs has been transformative, especially for accessing new information or details not publicly available. Consider a scenario where an investigator seeks evidence in a new case. A RAG-enhanced AI system, instead of just relying on its preset knowledge, would actively "search" through case data looking for relevant facts, integrating this with its foundational understanding to present a comprehensive, informed solution.

RAG isn't about diminishing the expertise of digital forensic experts but about amplifying their tools. Imagine RAG as a bridge between LLMs and real-time digital case data. The outcome? A forensic tool that reduces potential inaccuracies, and aids digital forensic experts in navigating the ever increasing volume of digital data from electronic devices with amplified precision and assurance.

Digital forensic applications

LLMs have already demonstrated their ability to assist in tasks like semantic search, interpreting different digital traces, software development and summarizing intricate datasets. How can RAG further increase the potential of these models for digital forensic professionals? Here are a few examples:

- Enhanced Semantic and Explorative Search: With RAG, when presented with a forensic query, the system doesn’t solely depend on its base knowledge. It can pull information from digital traces, such as emails, chat records, documents, geo-locations, timestamps, pictures, and browser history, that can be found in electronic devices from suspects.
- Digital Trace Interpretation: RAG can rapidly contextualize different types of digital traces, providing a more holistic view of potential evidence and helping professionals draw connections across varied data points.
- Assistance and Direction: Drawing inspiration from the concept of the Windows 11 copilot, imagine a forensic-specific AI tool that guides experts through intricate digital forensic software and that assists non-technical investigators with using review platforms, delivering real-time, context-sensitive assistance.

In summary, RAG promises to reshape how digital forensic professionals engage with data, analysis, and technology. It amplifies the capabilities of LLMs, weaving their profound foundational understanding with the agility to incorporate and contextualize the latest, relevant information.

Concerns

Incorporating AI solutions, especially within the sensitive realm of digital forensics, presents a trio of significant challenges: reliability, ethics, and privacy.

- Reliability: While RAG's capability to fetch external data is commendable, it's crucial that it taps into reliable and unbiased databases. Mistakes stemming from flawed data can significantly skew forensic conclusions.
- Ethical: Digital forensics intertwines with ethical decision-making. AI might decipher data, but humans must grapple with the ethical implications, ensuring AI insights don't overshadow vital moral judgments.
- Privacy: Sharing digital evidence with cloud-based AI stirs concerns about data breaches and evidence integrity. However, the rise of on-premise open-source LLMs in 2023 offers a promising development.

By tackling these challenges head-on, we pave the way for RAG's strategic integration into digital forensics, enhancing experts' capabilities while upholding core forensic principles.

Conclusion

In a field overwhelmed by digital evidence and growing backlogs, the introduction of AI "copilots" offers a new hope for digital forensics. The term "copilot" aptly describes these AI tools' role: not to replace but to reliably aid forensic experts in their investigative pursuits.

The suite of copilots introduced by Microsoft in partnership with OpenAI underscores the vast potential inherent in such AI advancements. The vision here is twofold: First, for non-technical investigators, the copilot serves as a guide, simplifying the process and ensuring informed conclusions. This approach democratizes basic forensic investigations, enabling a broader range of professionals to decipher digital trails. Second, seasoned digital forensic experts can utilize these copilots for advanced tasks, such as tool development and software reverse engineering, directing their expertise where it's most valuable.

The implications of this two-pronged vision are profound. For digital forensic experts and non-technical investigators to collaboratively tackle cases and truly democratize basic forensic investigations, a centralized approach becomes indispensable. The traditional model of isolated workstation investigations doesn't align with the computational demands of modern tools, especially powerful systems like on-premise large language models. This is where the concept of 'Digital Forensics as a Service' enters the picture, offering centralized processing pipelines and creating an environment where AI copilots can thrive. By embracing this shift, the digital forensic field stands to undergo a transformative evolution.

[1] ChatGPT as a Copilot for Investigating Digital Evidence (2023) by Hans Henseler and Harm van Beek. https://ceur-ws.org/Vol-3423/paper6.pdf

[2] ChatGPT for Digital Forensic Investigation: The Good, The Bad, and The Unknown (2023) by Mark Scanlon et al., https://arxiv.org/abs/2307.10195