PERSPECTIVES

INTEGRATING AI IN EDISCOVERY AND DIGITAL FORENSICS FOR CORPORATE INVESTIGATIONS

JS|HELD

Our perspectives feature the viewpoints of our subject matter experts on current topics and emerging trends.

Introduction

When most people think of Artificial Intelligence (AI), they often envision Generative AI (Gen-AI) tools that create content (such as images, videos, or text) or imagine futuristic scenarios involving intelligent robots or dystopian worlds where AI becomes self-aware, like Skynet and the T-800 from Terminator. While these ideas are fascinating, the real-world impact of AI is far more practical and increasingly relevant, especially in the legal technology space.

Although AI has been around for a while, its practicality and rapid advancement have become especially evident in recent years. Today, we're seeing a growing adoption of AI-driven processes across the legal industry, particularly in areas like eDiscovery and digital forensics.

The integration of AI with legal and analytical technologies is transforming how legal professionals manage and analyze data. AI enables faster, more accurate research, review, and analysis, reducing human error and providing deeper insights into complex and unstructured datasets.

In investigations requiring eDiscovery and digital forensics, AI can be a powerful complementary tool. It enables the analysis of digital information in ways that are both efficient and insightful. This article explores practical recommendations and processes for implementing AI in digital investigations and their workflows.

Modern Data Sources in Digital Investigations

Today's digital investigations often involve a wide array of modern and disparate data sources, each with unique formats and challenges. Some of these sources include:

- Messaging apps (e.g., WhatsApp, Signal, Telegram)
- Social Media Platforms (e.g., Discord, Reddit, TikTok)
- Cloud-based platforms (e.g., Slack, Microsoft 365, Google Workspace)
- Online backups and storage (e.g., iCloud, Google Drive, Dropbox)

It is important to understand the nature and structure of these data types, as it informs how investigators identify, collect, preserve, analyze, and validate digital evidence in a forensically sound manner.

A key aspect of digital analysis involves examining the contextual information surrounding data, as well as the attributes that enrich its context, such as metadata (which describes the who, what, when, and where of a file). Contextual data can include elements like chat messages, user-generated documents, or system logs that record user or system activity. Metadata, on the other hand, might indicate when a document was created, who authored it, or who the participants of a message correspondence were.

These data sources often contain vast amounts of information, making manual review both time-consuming and potentially prone to error. However, when metadata and contextual data are analyzed together, they can reveal and tell a story, helping investigators reconstruct timelines, understand user behavior, and identify potential threats or anomalies. By leveraging AI, digital investigators can quickly process this information together, corroborate findings, and uncover potential hidden information.

Security and Privacy Risks When AI is Used in Investigations

Some of the main concerns when utilizing an AI utility are security and privacy. Users may wonder about the security of data transmission and data retention, as well as the privacy regarding the sources of information gathered by the AI utility.

There is a potential risk of data exposure when using AI in a business setting. AI analysis of proprietary information should be used in a closed, secure environment and not on public platforms.

This limits the exposure of data to outside sources.

To run AI in a secure and enclosed environment, it's recommended that the environment is SOC 2 compliant as well as ISO 27001 certified and maintains strict access controls and audit trails. This ensures data is managed securely, remains private and available, and follows established protocols for protection and accountability.

When using public platforms, the above security controls may not be met. Privacy also becomes an issue as most public platforms collect user information, which may include:

- Geolocation
- AI Text Input / Searches
- Contact details
- Device and browser cookies
- IP Address
- Personally Identifiable Information (PII) (email address, name, contact information)

Even if an account is deleted, the above information may be sold to other third parties. AI in a closed environment, such as in conjunction with document review platforms (i.e., an eDiscovery tool), does conform to many of the security and privacy concerns outlined above.

AI-Driven Digital Investigations Workflows

When conducting an investigation, the volume of data for analysis may get immense and tend to become unmanageable. Even after using culling techniques within a forensic environment – which may include keywords, file types, file paths, etc. – the data can still be very cumbersome to analyze.

The most straightforward and sensible solution is to have the data in an eDiscovery environment where AI can be deployed on the data and analysis can be conducted.

This data can consist of the following:

- Emails
- Electronic Documents (e.g., MS Office, PDFs)
- Chats
- Internet History Files
- Carved Files from Unallocated Space
- Recovered/Deleted Files

Some think AI pertains primarily to traditional analytics (e.g., TAR and CAL). With AI, the technology builds a large language model of the documents available and provides answers to questions asked in real time and in layman's terms. The need to guess keywords and the application of other culling criteria is unnecessary. AI is able to rapidly process large datasets in a matter of moments, whereas humans would need days, or perhaps even weeks, for the same task. As AI digests information, it is able to create relationships and find hidden connections with additional data points. This enables examiners to find relevant information through direct references without the application of any culling.

One of the most attractive features of AI is the ability to use natural language search to search for documents and data of interest. Querying for data within an AI environment is as simple as conducting a search in Google using layman's terms. There is no need for complex syntax or understanding a particular language model. Common questions asked are:

» Has *Party A* been in contact with *Party B*?
» What knowledge or involvement did *Party A* have regarding *Issue B*?
» Which other parties had knowledge of *Issue B*?

The result is not only a summary of what the AI tool found to be the answer but also supporting documentation based on the data provided. In this sense, not only does AI answer the examiner's question but also provides supporting evidence as the basis for its answer. Examiners can then review the supporting evidence and determine if further queries need to be made or go back to the forensic environment and conduct further queries based on the results of AI. Within the forensic environment, examiners can delve deeper into logs, reports, and forensic artifacts (i.e., registry files and unallocated space) to see if there is any further evidence.

In the above scenario, AI proves to be a good starting point for an investigation. Key questions related to a matter are queried first with the use of AI. Then, based on the responses further analysis can be conducted either within the eDiscovery environment or the forensic environment.

The most notable drawback of AI is likely to be the cost of using this technology. Most pricing models are priced on a per-document basis. A document fee is usually incurred each time AI executes a query. Another point to consider is that the answer from AI will depend on the population of data. If the data is biased towards one opinion or party, the answer will be reflective of this.

For instance, if questioned, "Is Michael Jordan the greatest basketball player of all time?" and within the data population, there are no documents suggesting otherwise, the most likely response from AI will be: "Yes, Michael Jordan is the greatest player of all time," and it will provide supporting evidence from the document population to support this answer.

The AI technology is only as knowledgeable as the documents provided for it to gain knowledge. Therefore, when using this technology, examiners should be mindful of the dataset and the content of the dataset.

Conclusion

Recognizing how integrating eDiscovery and AI with the support of digital forensics can impact on how investigations are conducted. It can lead to enabling quicker analysis of large, complex datasets, enhancing the efficiency, accuracy, and depth of a forensic investigation. It removes the need for traditional, labor-intensive methods like keyword culling and supervised training, allowing investigators to ask natural language questions and receive both answers and supporting evidence in real time.

However, the use of AI also brings important considerations, particularly around data security, privacy, and cost. Deploying AI in secure, closed environments is essential to protect sensitive information. Additionally, investigators must remain aware of potential data bias, as AI's insights are only as reliable as the data it analyzes.

Ultimately, AI can act as a powerful starting point in digital investigations by accelerating information discovery and uncovering hidden connections that help guide a deeper forensic analysis.

Acknowledgments

We would like to thank our colleagues Richard Chung and Kiran Patel for providing insights and expertise that greatly assisted this research.

More About J.S. Held's Contributor

Richard Chung is a Managing Director providing Digital Investigations & Discovery services within J.S. Held's Global Investigations practice. Based in New York, he has more than 18 years of experience in technology consulting. His expertise is in digital forensics and eDiscovery, in connection with all aspects of investigations, litigation, and antitrust and merger matters. Richard has worked across a wide variety of industries, including financial services, pharmaceuticals, healthcare, construction, energy, and technology. His extensive experience and decisive expertise make him a trusted advisor on a variety of technology consulting engagements involving data preservation, data analysis, document review, and document production.

Richard can be reached at richard.chung@jsheld.com or +1 347 786 8958.