



AN FTI TECHNOLOGY REPORT – JULY 2026

The State of Emerging Data 2026: AI Expands the Digital Universe

How traditional approaches to discovery and
governance are buckling under the pressures of
modern data

As artificial intelligence transforms enterprise operations, it is simultaneously creating new categories of disputes and investigations and new data complexities in litigation, regulatory inquiries and compliance matters. Over the past year, dozens of cases have emerged wherein data from AI systems is considered critical evidence. The result is a new class of discoverable information: AI prompts, system responses, embedded summaries and AI-generated outputs requiring entirely new approaches to discovery and investigation.

AI artifacts now intersect with core business decisions, communications and workflows, appearing in regulatory inquiries, litigation and internal investigations. This is occurring in real time, ahead of settled case law, creating analysis challenges at the core of defensible discovery.

AI data's role in discovery is still taking shape. To date, many of the initial cases are high-profile disputes directly involving AI-first companies, where evidentiary issues center on AI prompt-response interactions and AI-generated content rather than traditional documents. Legal teams are facing the complexities involved with identifying, preserving and contextualizing distributed AI artifacts across complex systems containing massive data volumes. They must conduct these exercises to find facts, demonstrate how AI generated or influenced materials are at issue, and parse the differences between AI-generated content, human-generated content, intellectual property, copyrighted content and more, all while avoiding discovery failures.

Traditional E-Discovery for Emerging AI Organizations

Fast-growing AI companies face the same regulatory and legal challenges as any enterprise, but often lack mature response capabilities. These organizations also rely heavily on emerging cloud-first collaboration technologies and are at a heightened risk of experiencing trade secret disputes or IP theft. FTI Technology leverages emerging data expertise, solutions such as FTI Technology Connect, cloud-first forensics capabilities and core e-discovery experience to deliver immediate response support while building defensible frameworks that strengthen organizational capabilities for future matters.



AI as Evidence

AI data can now be considered as the fifth element in the emerging data sources paradigm shift, joining shared access, chat messaging, linked content and dynamic document versions as a significant disruptor to traditional processes.

Current legal workflows and e-discovery don't typically account for preserving, collecting and analyzing AI artifacts such as prompt-response logs, file access records and AI-generated outputs for evidence review. Just as chat and collaboration tools have required legal teams to evolve their approaches, organizations should anticipate the need for another significant adjustment to address AI artifacts in their e-discovery matters.

Common challenges include:

- Distinguishing what was authentically authored versus what was AI-assisted, AI-summarized or synthetically generated.
- Requests to include model-authentication language and provenance attestations in collection protocols.
- The need for new retention settings and legal hold updates to preserve prompt histories, AI-generated outputs, agent or model identifiers, and usage telemetry that explains how an AI system produced a given communication.
- Preventing default system policies from automatically purging critical AI artifacts before they can be collected.
- Analyzing for AI hallucinations, because once embedded in an email thread or summary pane, hallucinated assertions become discoverable records that can be mistaken for real viewpoints.

Addressing Authenticity and Forensic Analysis

When generative AI outputs are admitted as evidence, they must be scrutinized and assessed in order to determine legitimacy, particularly where deepfakes or other fraudulent materials may be in play. Rigorous metadata analysis can provide valuable information in this context, but experts must also contextualize the information in the real world to help distinguish whether a video, picture, audio, etc., is legitimate or originated from AI tools.

“AI will soon dominate emerging evidential trends. Enterprise copilots and standalone AI models are already generating relevant artifacts — prompts, autonomous actions and code output — that are landing inside investigative workflows. Combined with cloud-native event markers, mobile activity artifacts and short-lived collaboration logs, AI introduces new dimensions in provenance, attribution and intent analysis which can augment traditional communication review.”

— Craig Hendley, FTI Technology Managing Director

Ambiguous Guidance

To date, case law that defines the relevance of AI-generated content as evidence has been limited, creating a degree of ambiguity and practical challenges. In one recent litigation in the U.S. District Court in Connecticut, the court granted the defendant's motion to compel production of AI prompts used by the plaintiff's expert witness. A summary of the case states that the judge found that characterizing AI prompts as "notes" was not sufficient to satisfy the requirement limiting otherwise-relevant discovery and rejected the plaintiff's representation that no prompts existed, distinguishing them from the "search terms" the plaintiff claimed to have already produced. The court ordered the expert witness to disclose any AI prompts or queries and warned that sanctions would be available if the requirements weren't met.

Additionally, the Advisory Committee on the Rules of Evidence has responded to AI-related evidence challenges with a proposed new rule, Rule 707 (Machine-Generated Evidence). This proposal is designed to address concerns about the reliability of computer technologies that generate predictions or inferences from data; and in its current draft, states that machine-generated evidence produced without an expert witness will be held to the same reliability standards used to qualify expert testimony.

Overcoming Discovery Obstacles

An ongoing case involving an AI startup provides an example of the practical and technical challenges that often arise when handling AI system data as evidence. The company is facing litigation and allegations that the database it uses to train its models and generate responses for users contains information that has been used improperly. To respond to the litigation, the company needed to analyze its underlying systems, parse through numerous emerging data sources and AI databases, and determine whether there were any evidence or indicia of wrongdoing.

At the outset of the case, the e-discovery scope spanned a range of emerging data sources including Google Workspace, Slack and Notion environments, in addition to traditional data sources and numerous versions of documents. Beyond these various structured and unstructured data sources, the legal team also needed to capture and analyze its large retrieval augmented generation (RAG) database, which served as the foundation for its products and their inherent functionality. The database included logs of how users engaged with the company's AI tools and what underlying data was used to generate responses to users. Company source code was also in scope, requiring careful handling and review environments to protect sensitive IP.

According to an Onna survey, organizations over-collect or are forced to re-collect data in more than one-third of their matters.



Due to the volume and types of data involved, typical data migration processes and document review systems were not feasible, and alternative workflows were required to access and examine the data. Early on in the discovery process, a team of FTI Technology experts worked closely with technical support at AWS — the cloud infrastructure behind the client’s systems — to determine the tooling available to access, migrate and query the RAG database. Using capabilities embedded in AWS, the data was transferred directly into FTI Technology’s secure environment, where the RAG database could be mapped and interrogated using cloud native tools — functionality that was not possible using traditional e-discovery platforms. By pushing the analytics to the data in its native environment, rather than within an e-discovery platform, the team was able to query the data faster, reduce the burden of data migration and run meaningful analytics on a novel data source.

In the ongoing case, these workflows and specialized data and AI expertise have provided the client — which did not have mature infrastructure to address complex legal and risk issues — with the guidance it needed to effectively respond to this high stakes litigation. Without an innovative, unconventional approach, defensible discovery for this matter would not have been able to progress.

Next Generation Collaboration

The Onna survey found that the average organization manages data from 10 collaboration platforms per matter. Many of these matters include the latest wave of project management and workflow tools like Notion, Asana, Basecamp, Trello, Airtable and Monday.com, which sit at the center of all key emerging data sources complexities: linked content, multiple live versions of documents, shared access and chat. Long-standing tools are also adding AI capabilities to their toolbelt, and newer players are built AI-first. While many would have been considered tertiary or IT-specific in the past, their proliferation, integration with communication platforms and strong fit for cloud-enabled businesses have made them key sources of evidentiary value.

As an example, several challenges that arise with many of these platforms include:

- A lack of robust native e-discovery capabilities leaves the need for manual or custom workflows for preservation, search and collection.
- Data resides in shared, non custodial workspaces, making it difficult to define ownership, scope collections and ensure defensibility.
- Continually changing pages, version history and real time collaboration complicate reconstruction of a reliable point in time record.
- Heavy reliance on linked, embedded and integrated content increases the risk of incomplete or inconsistent collections.
- Granular permissions and cross platform integrations make it challenging to identify, access and validate all relevant data.

“Platform sprawl is accelerating and no single tool provides end-to-end visibility. Effective investigations now frequently require correlation across security tooling, identity platforms, collaboration logs and device artifacts. Evidence increasingly spans chat, social media and specialist cloud tools, each leaving its own telemetry footprint requiring specialist analysis.”

– Hendley

Data Spans a Wide Horizon

The recent challenges introduced by AI data and the newest collaboration functionality are compounded by those within the wider landscape of emerging data sources. Legal teams are still grappling with the pace of change within systems like Google Workspace, Microsoft 365, Slack and others, in addition to the volume, variety and veracity challenges across chat, linked content, shared access and versions, which continue to evolve and impact information governance and discovery in unique ways. These challenges are so stark and persistent that organizations collect the right amount of data in only a [reported](#) 2% of matters and the volume of data collected from collaboration applications meaningfully drives costs in litigation and investigations for two-thirds of organizations.

In the latest edition of [The General Counsel Report](#), 63% of legal department leaders said they were either very or extremely concerned about the impact of emerging data sources on their legal risk. One noted, “*We are very disciplined, but it is something we are concerned about and must continuously monitor.*” Several chief information officers were also interviewed for the report, and these participants — who tend to be closer to the inner workings of their organization’s data challenges — expressed deeper worries over emerging data sources, with 100% categorizing their feelings as either very or extremely concerned.

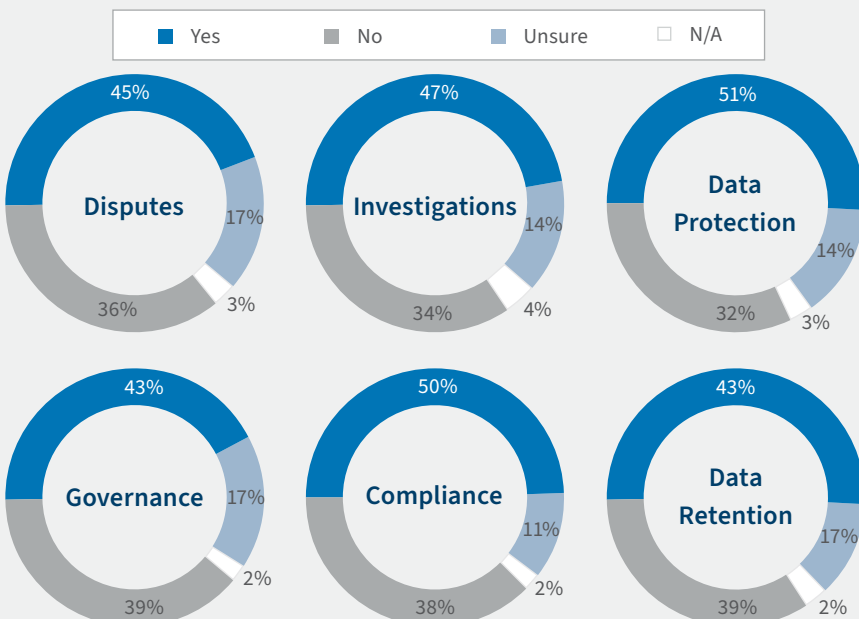
Source Code Complexities

In the cloud-first workplace, IP can walk out the door in seconds. A reported 37% of organizations face IP loss when an employee departs and more than one-third of IP theft incidents now involve source code and algorithms.

Reviewing code — which often resides in source control management systems like Git, Subversion or Perforce — for sensitive trade secret litigation is often urgent, sensitive, costly and complicated, requiring specialized digital forensic processes. Organizations concerned about software code theft or facing a related dispute can implement compliance monitoring across software development tools to monitor activity and document development cycles that may be examined alongside evidence in litigation. Litigation readiness playbooks that account for complex data and the unique attributes of source code discovery and analysis can also help.

Emerging Data Sources Challenges Across the Legal Spectrum

Has your organization experienced new challenges associated with collaboration applications, linked documents and cloud productivity platforms?



“Our concern is that the sheer volume of data requires more intentional organization and protection.”

Staying Focused on a Moving Target

With emerging data sources and AI system data, legal teams have the dual difficulty of credibly handling the technical complexity and explaining AI systems to regulators, courts and boards with comparable depth. Success in the face of these issues requires emerging data expertise, AI sophistication, forensic rigor and deep disputes experience.

37% of organizations believe their collaboration data is more than doubling in volume year over year.

Key steps to improve readiness for these issues include:

- **Revising playbooks** to account for emerging data sources nuances including the potential need for treating AI system artifacts as sources of electronic evidence.
- **Educating** the legal team on the potential need for additional metadata analysis to validate and distinguish human generated content from AI generated content.
- Assessing retention settings to understand how and where AI prompts and conversation logs are stored.
- **Preparing** proportionality arguments applicable to the burden of discovery related to AI system data and other emerging data sources.
- **Engaging** emerging data and digital forensics experts early on to navigate the technical complexities involved with analyzing unconventional formats and new types of metadata related to AI artifacts.
- **Data mapping** and assessing the readiness of collaboration, project management and specialized software.



Conclusion

Traditional workflows and tools no longer hold up to the modern data demands in discovery. While AI tools are providing relief for certain e-discovery workflows, the presence of AI system data as potential evidence exacerbates these challenges. When complex data is in play, technical knowledge is essential to provide a semblance of predictability to the discovery process, and to inform the legal team of the options for handling each unique data source. Moreover, expertise informs an appropriate balance between satisfying the courts and managing risk tolerance regarding the scope of discovery.

Organizations must be aware of the constantly changing realities of emerging data sources in legal and compliance, and develop an understanding of how to avoid common pitfalls that may quickly derail a case. FTI Technology's emerging data sources experts bring the technical adaptability and experience to help clients navigate litigation readiness, digital forensics, e-discovery and fact finding in numerous cases involving data from complex AI systems and other emerging data sources.



Tim Anderson

Senior Managing Director
+1 (415) 293 4483
tim.anderson@fticonsulting.com

Craig Hendley

Managing Director
+44 20 3727 1242
craig.hendley@fticonsulting.com

Grace Hougey

Senior Director
+1 (661) 600-2393
grace.hougey@fticonsulting.com

Rohit Kelkar

Managing Director
+1 (872) 228-3568
rohit.kelkar@fticonsulting.com

The views expressed herein are those of the author(s) and not necessarily the views of FTI Consulting, Inc., its management, its subsidiaries, its affiliates, or its other professionals. FTI Consulting, Inc., including its subsidiaries and affiliates, is a consulting firm and is not a certified public accounting firm or a law firm.

FTI Consulting is the leading global expert firm for organizations facing crisis and transformation.

FTI Consulting is dedicated to helping organizations manage change, mitigate risk and resolve disputes: financial, legal, operational, political and regulatory, reputational and transactional. FTI Consulting professionals, located in all major business centres throughout the world, work closely with clients to anticipate, illuminate and overcome complex business challenges and opportunities.

© 2026 FTI Consulting, Inc. All rights reserved. [fticonsulting.com](https://www.fticonsulting.com)