Calling the AI Witness in Second Requests and Merger Clearance
By Sean McDermott, FTI Technology

The next major witness in a Second Request investigation might not be a human. It could be an algorithm. Generative artificial intelligence has now taken its place in the enterprise stack, from Microsoft Copilot drafting emails, to Google Gemini summarizing chat threads, meeting notes and internal discussions. These systems are producing discoverable content at scale and much of it can appear to be human-authored. Organizations that expect to face a Second Request or regulatory inquiry in the coming years must begin to prepare for the need to collect AI artifacts as part of discovery and distinguish between human-generated and machine-generated materials.

Often, in Second Requests or other merger clearance reviews from the U.S. Department of Justice, Federal Trade Commission, European Commission, U.K. Competition and Markets Authority and other agencies, large volumes of data come into scope for analysis and production. Data from hundreds of custodians are collected, petabytes of communications are reviewed and millions of documents are analyzed for insight into organizations' competition strategies and market positions. These matters are already considered to be among the most challenging and high-pressure discovery and review exercises. Now, with the impending addition of AI-generated artifacts into the realm of digital evidence, forensic experts face a new frontier: distinguishing what was authentically authored versus what was AI-assisted, AI-summarized or synthetically generated.

As agencies sharpen their emphasis on "ordinary-course" business records in Hart-Scott-Rodino filings, the introduction of AI-generated content creates an evidentiary vulnerability that did not exist even a year ago. Contemporary merger reviews increasingly hinge on how executives describe competitive dynamics in their day-to-day communications. Yet when strategic assessments, draft emails, Teams summaries or even off-the-cuff market characterizations are generated or heavily shaped by an AI assistant, the resulting materials may not accurately reflect any human executive views.

This is not a theoretical problem. For example, a Microsoft Copilot-generated summary describing a rival as "dominant," could be misinterpreted by regulators as a contemporaneous admissions despite never having been authored, reviewed or endorsed by the relevant decision-maker. In opposition, an overly benign AI-drafted description of competitive pressure could mask legitimate antitrust red flags. In both cases, synthetic content risks distorting the evidentiary record that agencies rely upon to evaluate competitive intent, assess internal deliberations and test theories of harm. As AI becomes embedded in enterprise communication flows, organizations must anticipate this distortion effect and ensure that model-generated artifacts do not inadvertently reshape the narrative of how their executives understand the market.

From a preservation standpoint, organizations should also expect that the next 12 months will bring new obligations as Microsoft 365 Copilot, Google Gemini and other enterprise AI platforms begin surfacing richer audit trails and model-level metadata. Beyond traditional mailboxes and chat repositories, enterprises may need to preserve prompt histories, AI-

generated outputs, agent or model identifiers and usage telemetry that explains how an AI system produced a given communication. In practical terms, AI platforms will need to be treated as new sources of electronically stored information, with retention settings and legal holds updated accordingly to prevent default system policies from automatically purging critical AI artifacts before they can be collected in a Second Request.

Another emerging risk comes from AI hallucinations: instances where tools like Copilot or Gemini generate statements, summaries or characterizations that have no basis in any underlying human communication. Once embedded in an email thread or summary pane, these hallucinated assertions become discoverable records that can be mistaken for real executive viewpoints.

Several specific challenges that may impact Second Request investigations as AI-generated materials become more pervasive across enterprises include:

- **Attribution and intent**: If an executive's AI assistant sent a scheduling message or wrote an email that was not reviewed before it was sent, the lines that traditionally define whether a document is responsive or privileged become blurred. The more executives rely on this kind of AI-assisted automation, the less clearly the intent to communicate will map to a human custodian. From a digital forensics standpoint, this creates a significant challenge in a Second Request review: AI-authored content can carry metadata indistinguishable from a human-created record, making it difficult to interpret chain-of-thought, privilege triggers or whether a document reflects actual executive deliberation. This ambiguity elevates privilege-waiver risk, increases the likelihood of misclassification during responsiveness review and can force the producing party to defend evidentiary decisions that hinge not on human conduct but on algorithmic behavior, something current Second Request protocols are not yet designed to evaluate or certify.

- **Chain-of-custody implications**: Second Request protocols often rely heavily on date stamps, sender/recipient metadata and application logs to validate evidence. AI systems introduce model IDs, prompt histories and output logs to validate evidence. These are new forensic layers that aren't yet part of standard preservation orders, which could create inadvertent risks or confusion when compiling and producing a response to a Second Request or other regulatory inquiry. Complicating matters further, AI model updates can occur silently and automatically, meaning the same prompt issued at different times may produce materially different outputs. This "model version drift" breaks long-standing forensic assumptions around reproducibility and may impact the reliability of reconstructed timelines or intent analysis. Traditional custodian questionnaires rarely ask whether AI assistants were used to draft, summarize or auto-generate communications, or what devices and applications were linked to those assistants.

- **Defensibility risks**: How a producing party certifies the authenticity and completeness of records when some content originates from non-human agents has not yet been established in standard practices. For example, if an AI tool drafted part

of a communication chain, there is no set of rules to guide whether that text is still considered a record "of" the associated custodian. Organizations may struggle to demonstrate defensibility in their processes and production sets if AI-generated messages become conflated with custodian communications.

- **AI artifacts in discovery**: Microsoft 365 Copilot outputs, Teams summaries or even auto-generated replies can appear in user mailboxes and audit logs. These can become part of responsive data sets. Without clear attribution, these artifacts may confuse or conflict with information provided from other human-generated sources.

In the year ahead, it's likely for organizations to begin to see model-authentication language and provenance attestations appear in collection protocols for Second Requests and other merger clearance and competition enforcement inquiries. This new facet of discovery should be expected to develop similar to how technical details like metadata (e.g., "FileCreated" and "LastModifiedBy") previously became standard in these types of matters.

The question in merger clearance is no longer only, "Who said it?" With increasing generative AI adoption, that question is set to take on additional layers: "Which model said it?" and "Under what prompt, what version and with what audit trail?" In merger control matters and beyond, the next evolution in defensible collection will hinge on proving AI provenance as rigorously as proof of data authenticity has become expected as a standard practice. Ultimately, organizations may need to treat AI engines themselves as quasi-custodians with discoverable model states, version histories, prompt logs and provenance artifacts. This shift represents a fundamental evolution in evidentiary practice and will require new protocols and new certifications.