Discovery of AI: Prompts, Training Data, and the New ESI

Speakers:

- **Daniel Garrie,** Founder, Law & Forensics; Neutral, JAMS; Adjunct Faculty, Harvard
- **Scot G. Doyen,** Shareholder, Clausen Miller P.C.
- **Karen Silverman,** CEO and Founder, The Cantellus Group
- **Mari Grace,** Counsel, Debevoise & Plimpton LLP
- **Patricia Antezana,** Counsel, Reed Smith

Disclaimer



This is not legal advice, nor should it be considered legal advice.



This presentation and the comments contained therein represent only the personal views of the participants, as spoken, and do not reflect those of their employers or clients.



This presentation is offered for educational and informational uses only.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Speakers



Daniel B. Garrie

Founder,
Law & Forensics
Neutral, JAMS
Adjunct Faculty,
Harvard



Scot G. Doyen

Shareholder,
Clausen Miller P.C.



Karen Silverman

CEO and Founder, The
Cantellus Group



Mari Grace

Counsel,
Debevoise & Plimpton
LLP



Patricia Antezana

Counsel, Reed Smith

(c) 2026 Lexeprint Inc. All Rights Reserved.

Agenda

- **Overview of AI**
- **AI Data as ESI**
- **Requesting AI Artifacts**
- **Recent Decisions**
- **Best Practices**

(c) 2026 Lexeprint Inc. All Rights Reserved.

Overview of Artificial Intelligence



(c) 2026 Lexeprint Inc. All Rights Reserved.

What Is Generative AI?

- **A system that produces new content rather than retrieving it.** Traditional software searches a database and returns what is already there. A generative model composes an answer that did not exist before — which is why there is often no stored 'document' behind an output.
- **It works by predicting, not by reasoning.** A large language model is trained on vast quantities of text and learns statistical patterns. Given a prompt, it predicts the most likely next word, then the next — fluently, and sometimes confidently wrong.
- **Two separate phases, two separate records.** Training builds the model from a fixed corpus. Inference is each individual use. The discovery questions, custodians, and retention periods differ entirely between the two.
- **The user supplies the prompt; the vendor supplies much of the rest.** Hidden system instructions, model version, temperature settings, and retrieved context all shape the output, and most of that lives with a third party.
- **Why this matters for discovery.** Every one of these mechanics creates a distinct, discoverable artifact — and each has its own default retention window, often measured in days.

(c) 2026 Lexeprint Inc. All Rights Reserved.

What a Generative AI System Actually Is

- **A statistical model, not a database.** A large language model does not store and retrieve documents. It predicts the next unit of text from patterns learned during training, which is why it cannot reliably tell you what it 'knows' or where that came from.
- **Training is distinct from use.** Training data shapes the model once, up front. Inference is each individual use — a prompt in, an output back. These generate entirely different evidentiary records.
- **Outputs are not reproducible by default.** Model version, temperature settings, retrieved context, and system instructions all change the answer. Re-running a prompt produces a new event, not proof of the original one.
- **Most enterprise AI is layered.** A vendor platform calls a third-party model, which may retrieve from the company's own documents. Evidence is scattered across all three layers and across multiple contracts.

(c) 2026 Lexeprint Inc. All Rights Reserved.

The Vocabulary That Belongs in Your Requests

User-Facing Terms

- **Prompt.** The instruction a user types. Discoverable as the party's own statement.
- **Output.** What the model generates in response, including drafts never sent.
- **Session or conversation history.** The full multi-turn exchange, including abandoned branches.
- **Activity log.** Who used which tool, when, and with what settings.

System and Model Terms

- **System prompt.** The hidden standing instruction governing every response. Often the most probative artifact.
- **Training and fine-tuning data.** The corpus that shaped the model's behavior.
- **Model version or identifier.** A discrete, load-bearing fact — behavior is not reproducible without it.
- **Retrieval context and embeddings.** The documents actually fed to the model at run time.

(c) 2026 Lexeprint Inc. All Rights Reserved.

AI Data as Electronically Stored Information

(c) 2026 Lexeprint Inc. All Rights Reserved.

The Governing Principle: Old Rules, New Data

- **Rule 26(b)(1) governs without modification.** Parties may obtain discovery of any non-privileged matter relevant to a claim or defense and proportional to the needs of the case. Nothing in the rule carves out machine-generated content.
- **Rule 34 is medium-agnostic.** It reaches writings, drawings, graphs, and 'other data or data compilations stored in any medium' — language drafted broadly for exactly this situation.
- **Courts have consistently declined to create an AI exception.** The emerging federal consensus is that relevant generative AI data is discoverable and must be treated like any other potentially relevant ESI.
- **The real fights are practical, not doctrinal.** Disputes center on proportionality, retention, privilege, and third-party possession — not on whether AI data is discoverable at all.

(c) 2026 Lexeprint Inc. All Rights Reserved.

What Actually Exists to Be Discovered

Inference-Layer Artifacts

- **Prompts and outputs,** including deleted and regenerated turns.
- **System prompts and configuration files** governing tool behavior.
- **Tool-call and agent logs** showing what an autonomous system was permitted to do and what it did.
- **Retrieval records** identifying which internal documents were surfaced to the model.
- **Usage telemetry —** user identity, timestamp, token counts, model version.

Model-Layer Artifacts

- **Training and fine-tuning datasets,** and their provenance documentation.
- **Model cards and evaluation reports** describing known limitations and error rates.
- **Version history** showing how a model changed over time.
- **Validation and testing records,** including bias and accuracy audits.
- **Governance materials —** AI use policies, approval records, and internal risk assessments.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Why AI Evidence Disappears Before You Ask for It

- **Retention windows are measured in days.** Prompt and tool-call logs are commonly retained for thirty days or less by default — and many inference providers retain nothing on the customer's behalf at all.
- **The data sits with vendors, not custodians.** A traditional hold reaching mailboxes and shared drives will not touch an orchestration platform or a model API provider.
- **Contracts rarely require preservation.** Standard AI vendor agreements impose no litigation-hold obligation, so the duty to preserve and the ability to preserve can be held by different entities.
- **Model versions are deprecated on the provider's schedule.** Once a version is retired, the behavior at issue may be permanently unreproducible, regardless of any hold.
- **Consumer accounts are the blind spot.** Employees using personal AI accounts for work generate relevant data entirely outside enterprise retention systems.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Who Actually Possesses the AI Data?

Mapping the Custody Chain

- **The employee's endpoint.** Browser history, local caches, and copied-out drafts — often the only trace of a consumer-account session.
- **The enterprise platform.** Copilot, an internal chatbot, or a review tool, typically with admin-configurable logging and retention.
- **The orchestration layer.** The middleware that assembles prompts, calls tools, and retrieves documents; frequently a separate vendor.
- **The model API provider.** The entity actually running inference, often retaining nothing on the customer's behalf.

The Rule 34 Control Question

- **The standard is possession, custody, or control.** A party must produce responsive ESI it has the legal right or practical ability to obtain — not only what sits on its own servers.
- **Contract terms are the evidence.** Whether a party controls vendor-held AI data usually turns on the service agreement and its data-access provisions.
- **Request the agreements early.** The AI vendor contract tells you what exists, who keeps it, and for how long.
- **Subpoena where control fails.** A Rule 45 subpoena to the platform or provider is the fallback when the party genuinely cannot compel production.

(c) 2026 Lexeprint Inc. All Rights Reserved.

When the AI Is the Actor: Agentic Systems

- **Autonomous agents leave no human author to depose.** When a system takes an action on its own, there is no email thread and no decision-maker who can explain why — and the engineers who built it frequently cannot say either.
- **The evidence is configuration, not correspondence.** What the agent was instructed to do, what tools it was permitted to call, and what it retrieved matter more than any single output.
- **Request the system prompt and the tool-call log together.** The standing instruction explains intent; the tool-call log shows what the system actually did, in sequence, with timestamps.
- **The model version is load-bearing.** Behavior cannot be reconstructed without it, and providers deprecate versions on their own schedule.
- **Re-running the agent proves nothing.** A re-run is a new event producing a new output. Only the contemporaneous log establishes what happened — which makes preservation, not reproduction, the whole ballgame.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Requesting AI Artifacts in Discovery

(c) 2026 Lexeprint Inc. All Rights Reserved.

Building the Request

Open With Foundation Discovery

- **Interrogatory:** identify every AI or machine-learning tool used in connection with the matters at issue, by vendor, product, and version.
- **Interrogatory:** identify the custodians authorized to use each tool and the period of use.
- **Request:** produce AI governance policies, approval records, and acceptable-use documentation.
- **Rule 30(b)(6):** designate a witness on AI systems, logging architecture, and retention settings.

Then Request the Artifacts

- **Specify the artifact by name —** prompts, outputs, system prompts, tool-call logs, model identifiers, retrieval context.
- **Define the time period and custodian set** narrowly enough to survive a proportionality objection.
- **Request native format with metadata** intact; exported transcripts routinely strip model version and timestamps.
- **Address third-party possession** expressly, invoking the party's practical ability to obtain the data.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Use the Rule 26(f) Conference and the ESI Protocol

- **Put AI on the agenda explicitly.** Rule 26(f)(3)(C) requires the parties to discuss disclosure and production of ESI. Raise AI tools by name, or risk waiving the issue until the data is gone.
- **Negotiate retention before you negotiate production.** An agreed preservation stipulation covering prompt and log data is the single highest-value early win.
- **Address AI-assisted review directly.** If either side will use generative AI for responsiveness review, the protocol should state what disclosure of prompts, validation metrics, and hosting environment is required.
- **Build in a model-version freeze.** Where behavior of a specific model is at issue, seek an order preserving that version and its configuration.
- **Set an authentication framework early.** Agree on how AI-generated material will be authenticated to avoid a Rule 901 fight on the eve of trial.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Privilege and Work Product: A Narrower Shield Than Counsel Expect

- **A prompt is not privileged merely because a lawyer suggested the tool.** Business and technical content does not become privileged because counsel recommended its creation; the communication itself must convey legal advice.
- **Counsel-directed litigation prompts stand on stronger ground.** Prompts reflecting an attorney's mental impressions in anticipation of litigation present a genuine work-product claim — but the claim must be preserved and logged.
- **Partial disclosure waives.** Courts have ordered production of a full AI transcript where a party disclosed a portion of it or placed the AI-assisted investigation at issue.
- **Expert use is methodology, not work product.** Where an expert used AI in forming an opinion, the prompts are fair ground for discovery as part of the methodology.
- **Log AI materials on the privilege log.** Unlogged prompt data is an invitation to a waiver motion.

(c) 2026 Lexeprint Inc. All Rights Reserved.

(c) 2026 Lexprint Inc. All Rights Reserved.

Recent Decisions Shaping AI Discovery

Prompts and Logs Are Discoverable

In re OpenAI Copyright Litigation

- **S.D.N.Y., MDL No. 25-MD-3143 (Dec. 2, 2025).** Magistrate Judge Ona Wang compelled production of millions of generative AI logs — user prompts and model responses — on condition that user identifiers be anonymized.
- **Holding:** the logs were relevant and proportional to claims that the AI systems reproduced copyrighted works in their outputs.
- **Takeaway:** volume alone does not defeat production. Anonymization and protective-order terms are the court's preferred tools over outright denial.

Conservation Law Foundation v. Shell Oil

- **D. Conn., No. 3:21-cv-00933 (May 18, 2026).** Magistrate Judge Thomas Farrish ordered the plaintiff to produce the generative AI prompts its expert used in preparing her report.
- **Holding:** 'An expert witness's methodology is fair ground for discovery.' Reportedly the first federal order compelling expert AI prompt disclosure.
- **Note:** the district court stayed the order pending objection — the issue remains live, but the reasoning is already being cited.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Protocols, Monitors, and the Expanding Preservation Map

James v. Cerebras Systems

- **N.D. Cal., No. 4:25-cv-09361 (July 7, 2026).** Magistrate Judge Robert Illman entered a negotiated ESI protocol and companion protective order creating a distinct 'AI responsiveness review' regime, separate from traditional technology-assisted review.
- **Requires disclosure of** the prompts, templates, and iteration history used to determine responsiveness, plus hosting environment, training data, and validation metrics.
- **Significance:** the first federal codification of up-front, mandatory prompt disclosure for AI-driven document review.

United States v. RealPage

- **Consent decree entered Nov. 24, 2025.** The court-appointed monitor may demand source code, model training documentation, and runtime logic — and retain outside experts at the company's expense to interpret it.
- **Effect:** model artifacts, configuration files, and version history are treated as discoverable records in an antitrust enforcement posture.
- **Takeaway:** preservation protocols built around custodians and mailboxes do not reach any of these categories.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Best Practices

(c) 2026 Lexeprint Inc. All Rights Reserved.

Practical Checklists

For the Requesting Party

- **Send an AI-specific preservation letter immediately,** naming prompts, outputs, logs, system prompts, model versions, and vendor-held data.
- **Ask the foundation questions first** so your artifact requests are targeted and defensible.
- **Pin the model version in writing** before the provider deprecates it.
- **Treat vendors as separate sources** and subpoena where the party lacks practical control.
- **Test authenticity, not just content —** re-running a prompt verifies nothing.

For the Producing Party

- **Map your AI footprint now,** including shadow use of consumer accounts by employees.
- **Extend the litigation hold** to AI platforms, logging systems, and vendor-held data on day one.
- **Change default retention settings** before a duty to preserve attaches, not after.
- **Negotiate preservation terms** into AI vendor contracts as a standard provision.
- **Log privilege claims contemporaneously** and avoid partial disclosure that waives the whole exchange.

(c) 2026 Lexeprint Inc. All Rights Reserved.

Key Takeaways

- **There is no AI exception to the Federal Rules.** Relevant prompts, outputs, and logs are discoverable ESI, and courts have said so repeatedly.
- **Preservation is the emergency; production is the negotiation.** Default retention of days, vendor-held data, and deprecated model versions mean the record is lost before most holds are issued.
- **Specificity wins.** Request artifacts by technical name and sequence foundation discovery ahead of artifact requests.
- **The ESI protocol is the leverage point.** Courts are increasingly willing to enter AI-specific protocols — ask for one.
- **Privilege is narrower than counsel assume.** Expert prompts are methodology; partial disclosure waives; business prompts are not protected by a lawyer's recommendation.

(c) 2026 Lexeprint Inc. All Rights Reserved.



(c) 2026 Lexeprint Inc. All Rights Reserved.



Daniel B. Garrie, Esq.
Law & Forensics LLC – Founder
JAMS – Neutral
Harvard – Adjunct Faculty

Contact:
E: daniel@lawandforensics.com
URL: https://www.lawandforensics.com/

B.A., Computer Science, Brandeis Uni.
M.A., Computer Science Brandies Uni.
J.D., Rutgers School of Law

Daniel Garrie, Esq. is the Co-Founder of Law & Forensics LLC, where he heads the Computer Forensics and digital discovery Cybersecurity teams. Daniel has been a dominant voice in the computer forensic and cybersecurity space for the past 20 years, as an attorney and technologist. He is an adjunct professor at Harvard for Computer Forensics and prior to Law & Forensics, he successfully built and sold several technology start-up companies.

Since co-founding Law & Forensics LLC in 2008, Daniel has built it into one of the leading boutique cybersecurity forensic engineering firms in the industry. He is a mediator, arbitrator, and e-discovery special master for JAMS and is a partner and head of the Cybersecurity at Zeichner, Ellman & Krause LLP. Daniel has both a Bachelor's and a Master's degree in computer science from Brandeis University, as well as a J.D. from Rutgers Law School.

Daniel has led cyber and forensic teams in some of the most visible and sensitive cyber incidents in the United States. He and his team have worked for two of the top five banks in the globe, and dozens of the largest private and public companies in the world. In addition, Daniel has been awarded several patents for advanced cybersecurity and forensic platform he built with his team that is currently used in the industry, TableTop.AI, CustodyTrack.IO, and Forensic Scan.

Daniel is also well-published in the cybersecurity space and has authored more than 200 articles and books. His work is cited by Black's Law Dictionary 10th Ed. In defining the terms "software", "internet", and "algorithm." Lastly, he has been recognized by several United States Supreme Court Justices for his legal scholarship and is a trusted source and thought leader for cybersecurity articles and opinions, being cited over 500 times to date.

(c) 2026 Lexeprint Inc. All Rights Reserved.



Scot G. Doyen
Clausen Miller P.C. – Shareholder

Contact:
E: sdoyen@clausen.com
W: https://www.clausen.com/attorney/scot-g-doyen/

B.S., Sociology, Lamar Uni.
J.D., South Texas College of Law Houston

Scot Doyen is a Partner at Clause Miller P.C. and is Board Certified in Civil Trial Law by the Texas Board of Legal Specialization, a distinction held by fewer than 2% of attorneys licensed to practice law in Texas. He has built a distinguished trial practice representing insurers and businesses in complex litigation, insurance coverage disputes, and bad faith claims.

Mr. Doyen has maintained a broad and varied trial practice for more than 20 years, trying more than 60 significant cases and handling numerous civil and criminal appeals. In recognition of his work, he was selected to the Texas Super Lawyers list in 2011 for Personal Injury Defense Litigation, an honor awarded to fewer than 5% of eligible Texas attorneys.

Throughout his career, Mr. Doyen has routinely represented insurance carriers in both routine and complex coverage matters, providing strategic coverage analysis, legal opinions, and defense of bad faith claims involving personal and commercial lines of insurance. He played a significant role in defending hundreds of Hurricane Ike lawsuits and continues to represent insurers throughout Texas in a wide range of first-party property and casualty litigation.

Mr. Doyen also advises insurance carriers on umbrella and excess liability coverage issues, including drafting coverage opinions, monitoring litigation, and assuming the defense of insureds when appropriate. His experience spans a broad range of complex insurance matters, including bad faith and extra-contractual claims, interpretation of excess, umbrella, and manuscript policies, products liability and commercial litigation coverage disputes, employment practices liability claims, advertising injury and Directors & Officers liability issues, and professional liability coverage involving lawyers, accountants, insurance agents, and real estate professionals.



Mari Grace
Counsel, Debevoise & Plimpton

Contact:
E: mgrace@debevoise.com
W: https://www.debevoise.com/marigrace

Mari Grace is counsel and a member of the firm's Data Strategy & Security Group. Her practice focuses on advising clients on artificial intelligence governance, regulatory compliance and litigation risk. She works with organizations to assess AI use cases, develop governance frameworks, negotiate AI-related agreements, create and update policies and procedures, train legal departments on ethics and privilege issues, and navigate the rapidly evolving legal and regulatory landscape surrounding artificial intelligence.

Drawing on more than a decade of litigation experience, Ms. Grace helps clients evaluate and manage the legal risks associated with emerging technologies, novel business practices and developing regulatory frameworks. She advises clients on AI-related litigation preparedness, regulatory exposure and enterprise risk management.

Ms. Grace has represented major energy, telecommunications, hospitality and retail companies, as well as financial institutions, in antitrust litigation, securities class actions, derivative litigation, M&A disputes and other complex commercial litigation in federal and state courts across the country.

Prior to joining Debevoise, Ms. Grace practiced at another prominent international law firm, where she represented clients in high-stakes commercial litigation and regulatory matters.

She received her J.D. from Fordham University School of Law in 2011 and her B.A. from George Mason University in 2007. She is admitted to practice in New York and the District of Columbia

(c) 2026 Lexeprint Inc. All Rights Reserved.



Karen Silverman
CEO and Founder, The Cantellus Group

Contract:
E: karen.silverman@cantellusgroup.com
W: https://www.cantellusgroup.com/

Karen is a leading global expert in practical governance strategies for AI and other frontier technologies. As the CEO and Founder of The Cantellus Group, she advises Fortune 50 companies, startups, consortia, and governments on how to govern cutting-edge technologies in a rapidly changing policy environment. Her expertise is informed by more than 20 years of practice and management leadership at Latham & Watkins, LLP where she advised global businesses in complex antitrust matters, M&A, governance, ESG, and crisis management.

Karen is a World Economic Forum Global Innovator and sits on its Global AI Council where she has contributed to the Board Toolkit for AI and ongoing work on AI, data, and cybersecurity issues. She has been named one of the Top Ten Legal Innovators (The Financial Times, 2019), a Top AI Lawyer in California (California Daily Journal, 2019), and one of the 100 Brilliant Women in AI Ethics (Women in AI Ethics, 2020).

As a leading voice in the governance of AI and other frontier technologies, she is a regular speaker at conferences and forums including CogX, RSA, HIMSS, the Athens Roundtable, National Judicial College, Aspen Institute Technology Fellows and Berkeley Law. Her thoughts on the governance, oversight, and real-world applications of AI, AR, VR, and other nascent technologies are featured in The World Economic Forum's Agenda and the AI Journal. Karen currently chairs the board of Krunam, a public benefit corporation developing complex content moderation tools, and sits on the Board of Directors and board committees of multiple nonprofits.

Prior to founding The Cantellus Group, Karen spent 20 years at Latham & Watkins, LLP in global strategic and tactical management roles, including Office Managing Partner for the San Francisco office. She specialized in global merger control and helped to build the firm's leading global M&A competition practice. For this work, she has been recognized as a top antitrust practitioner by the GCR, Law 360, Legal 500, Who's Who, the Daily Journal, The Recorder, and SF Business Times.

Karen earned her J.D. from the University of California at Berkeley and her B.A. from Wellesley College. She held early career roles at McCutchen LLP, Jones Day LLP, Paul Hastings LLP, the Department of Justice's Antitrust Division, and the Securities and Exchange Commission.

(c) 2026 Learning Inc. All Rights Reserved



Patricia Antezana
Reed Smith LLP – Counsel

Contact:
E: PAntezana@ReedSmith.com
W: https://www.reedsmith.com/

Patricia specializes in managing e-discovery. With a background in general litigation, including commercial matters, business torts and product liability defense, Patricia brings a wealth of experience resolving a wide range of discovery issues. She advises clients regarding concerns that arise prior to litigation, including legal holds and preservation, and throughout litigation, including early case assessment, collections, review, and productions.

Patricia has been involved with all phases of litigation, including management of large-scale discovery, dispositive motions, trial preparation and alternative dispute resolution. She has coordinated electronic and hard copy record collection and managed large-scale review projects in numerous complex litigation matters.

Additionally, Patricia has written and spoken extensively on the topic of e-discovery and recently founded a Pittsburgh chapter of the national organization Women in E-Discovery (WiE).

(c) 2026 Lexeprint Inc. All Rights Reserved.