4TH EDITION

GENERATIVE AI

FOR SMART DISCOVERY PROFESSIONALS



JOHN TREDENNICK, J.D. DR. WILLIAM WEBBER

4TH EDITION

GENERATIVE AI

FOR SMART DISCOVERY PROFESSIONALS

MERLIN

A NOTE FROM THE AUTHORS



John Tredennick, J.D.



Dr. William Webber

Introduction to the Fourth Edition

When ChatGPT launched in November 2022, it signaled the arrival of a new technological era. Now, several years into the Generative AI revolution, we can assess its impact with clearer eyes. What once seemed like science fiction has become standard practice in many legal workflows. Discovery professionals who dismissed AI as hype now find themselves working alongside these systems daily. The question is no longer whether AI will transform legal work, but how to harness its capabilities effectively while maintaining professional standards and ethical obligations.

This transformation extends far beyond simple document search or basic summarization. Today's AI systems can plan multi-step investigations, analyze patterns across thousands of transcripts, extract structured data from unstructured documents, and even collaborate with other AI systems to tackle complex discovery challenges. These "agentic" systems represent a fundamental shift in how we approach information analysis. Rather than simply responding to our prompts, they can pursue goals, iterate on their approaches, and refine their work through multiple passes.

For discovery professionals, this evolution matters immensely. The volume and complexity of data in modern litigation continues to grow exponentially. Traditional approaches struggle to keep pace. A single matter might involve millions of documents, hundreds of custodians, dozens of languages, and terabytes of data spanning email, chat, social media, and proprietary systems. Manual review remains prohibitively expensive. Keyword search misses critical context. Even traditional predictive coding has limitations when dealing with nuanced questions that require synthesis across documents.

Generative AI offers something different. These systems can read documents much as humans do, understanding context and nuance. They can follow complex instructions, synthesize information from multiple sources, and present findings in clear narrative form. More importantly, when properly designed and deployed, they can do this at scale while maintaining audit trails

and providing citations to source material.

But this power comes with responsibility. Discovery professionals must understand not just what these systems can do, but how they work, where they fail, and what safeguards are necessary. We must grapple with ethical questions about confidentiality, accuracy, bias, and professional competence. We must develop new workflows that combine AI efficiency with human judgment. And we must stay current as technology continues its rapid evolution.

Our Perspective

We write from direct experience building and deploying AI systems for discovery and investigations. We have seen these systems succeed brilliantly and fail unexpectedly. We have worked with clients who embraced AI early and others who approached it cautiously. We have learned from both successes and mistakes.

This practical experience shapes our perspective. We are neither AI evangelists who see it as a panacea nor skeptics who dismiss its potential. We recognize that AI represents genuine innovation in legal technology while acknowledging its current limitations. We believe discovery professionals can and should leverage these capabilities, but only with proper understanding, appropriate safeguards, and clear-eyed assessment of risks.

Our goal is to equip you with the knowledge needed to make informed decisions about AI in your own practice. We want you to understand the technology well enough to evaluate vendor claims critically, implement AI systems effectively, and use them ethically. We want you to see opportunities others might miss while avoiding pitfalls others might stumble into.

Most importantly, we want you to approach AI as a tool that enhances your professional capabilities rather than replaces them. These systems amplify human expertise—they don't substitute for it. Your judgment, experience, and ethical grounding remain essential. AI changes how you work, not whether your work matters.

Welcome to the evolving world of AI-powered discovery. Let's explore it together with the rigor, skepticism, and intellectual curiosity the subject deserves.

John Tredennick and William Webber

November 1, 2025

© 2025 Merlin Search Technologies, Inc.

TABLE OF CONTENTS

01 **Understanding Large Language Models**
The core technology behind today's AI revolution in discovery.

02 **Context Windows and Conversation Management**
How AI maintains context and coherence without true memory.

03 **Hallucinations and Accuracy**
Recognizing and reducing false or fabricated AI outputs.

04 **Data Security, Privacy, and Confidentiality**
Ensuring safe and compliant use of client data.

05 **Ethical and Professional Responsibilities**
Applying ethical rules to AI-driven legal work.

06 **Understanding Agentic AI Systems**
Exploring autonomous systems that plan, act, and reason.

07 **Agentic Workflows in Discovery**
Coordinating AI agents to manage complex discovery tasks.

08 **Implementing and Controlling Agentic Systems**
Balancing automation with human oversight and control.

© 2025 Merlin Search Technologies, Inc.

TABLE OF CONTENTS

09 **Ethical Considerations for Agentic Systems**
Addressing accountability and transparency in autonomous AI.

10 **Document Analysis and Review Workflows**
Enhancing document review with scalable AI support.

11 **Transcript and Testimony Analysis**
Extracting insights and contradictions from deposition records.

12 **Timeline and Chronology Construction**
Transforming data into clear, defensible case timelines.

13 **Relationship Mapping and Communication Analysis**
Revealing hidden networks and communication patterns.

14 **Production Review and Gap Analysis**
Detecting missing data and production inconsistencies.

15 **AI-Powered Argumentation and Advocacy**
Building persuasive trial narratives from complex evidence.

16 **Building AI-Augmented Discovery Practice**
Integrating AI into everyday workflows for lasting advantage.

© 2025 Merlin Search Technologies, Inc.

MERLIN

PART ONE

FOUNDATIONS OF GENERATIVE AI FOR DISCOVERY PROFESSIONALS

© 2025 Merlin Search Technologies, Inc.

PART ONE

Foundations of Generative AI for Discovery Professionals

Before you can deploy Generative AI effectively in discovery and investigations, you need to understand what these systems are, how they work, and what constraints govern their use. Part One provides this essential foundation.

Many legal professionals approach AI with a mix of curiosity and concern. The technology seems powerful but opaque. Headlines tout revolutionary capabilities while also warning about hallucinations and security risks. Marketing materials promise dramatic efficiency gains, but ethical guidance counsels caution. This combination often leaves discovery professionals uncertain about how to proceed.

Part One cuts through this confusion by providing clear, accurate explanations of the technology underlying Generative AI. We explain not just what Large Language Models can do, but how they do it and why that matters for your work. We address the most serious concerns about accuracy and security directly, examining both genuine risks and overblown fears. And we provide the ethical framework needed to use these tools responsibly within the bounds of professional obligations.

What You'll Learn

Part One establishes the technical foundation you need by explaining how Large Language Models work, what context windows are and why they matter for discovery workflows, why hallucinations occur and how to mitigate them, whether you can safely send client information to AI systems, and what ethical obligations govern AI use in legal practice. You'll understand not just what these systems can do, but their fundamental limitations, security characteristics, and the professional responsibilities that constrain their deployment.

Why This Foundation Matters

You might be tempted to skip directly to Part Three's practical applications and examples. We understand the impulse. Discovery professionals face immediate challenges requiring practical solutions. Reading about transformer architectures and training cutoffs might seem academic when you need to review a million documents by next month.

But experience teaching legal professionals about AI has taught us that understanding fundamentals pays dividends. Those who grasp how LLMs work make better decisions about when and how to deploy them. They design more effective workflows. They catch errors others miss. They explain their methodologies more convincingly to clients, courts, and opposing counsel. And they adapt more successfully as the technology evolves.

Consider hallucinations. If you understand why they occur—that LLMs predict probable next words rather than retrieve verified facts—you'll naturally implement appropriate verification.

© 2025 Merlin Search Technologies, Inc.

If you view hallucinations as mysterious malfunctions, you might either avoid AI entirely or trust it naively. Neither serves your clients well.

Similarly, understanding context windows explains why some discovery tasks work beautifully with AI while others require different approaches. Knowing that LLMs have no memory clarifies both security characteristics and practical limitations. Grasping the ethical framework allows you to adopt AI confidently rather than hesitantly, knowing you're meeting professional obligations.

How to Approach Part One

We've written Part One to be accessible regardless of technical background. We explain concepts clearly without assuming prior AI knowledge. But we don't oversimplify. Discovery professionals deserve accurate understanding, not vague analogies that break down under scrutiny.

If you find certain sections challenging, persist through them. The investment pays off. If particular topics interest you deeply, use them as starting points for further exploration. We provide frameworks for understanding AI, not exhaustive technical documentation.

Most importantly, read Part One with your practice in mind. As you learn about context windows, consider your typical document volumes. When we discuss hallucinations, think about your verification workflows. While exploring ethics, reflect on your client relationships and professional obligations. The goal isn't abstract knowledge but practical understanding that improves your work.

Moving Forward

Part One establishes the foundation for everything that follows. Part Two will build on this foundation, exploring agentic AI systems that can plan, execute, and refine complex multi-step workflows. Part Three will demonstrate these capabilities through detailed discovery applications, showing how the principles from Part One translate into practical tools that transform legal work.

But those advanced capabilities and practical applications rest entirely on the fundamentals we cover here. Understanding what LLMs are, how they process information, where they fail, how to protect confidential data, and what ethics requires provides the knowledge base for responsible, effective AI adoption.

Let's begin with the most basic question: What exactly is a Large Language Model, and how does it work?

© 2025 Merlin Search Technologies, Inc.

MERLIN

01

Understanding Large Language Models

© 2025 Merlin Search Technologies, Inc.

CHAPTER ONE

Understanding Large Language Models

The technology behind Generative AI represents one of the most significant developments in computing history. For discovery professionals, understanding how these systems work is not merely academic curiosity. It provides the foundation for using them effectively, recognizing their limitations, and making informed decisions about when and how to deploy them in legal work.

This chapter introduces the fundamental concepts underlying Large Language Models. We explain what these systems are, how they're built, and why they matter for legal practice. Our goal is not to make you a machine learning engineer, but to give you sufficient understanding to use these tools confidently and critically.

What is Generative AI and How Does it Work?

Generative AI is artificial intelligence capable of creating new content—text, images, videos, audio, code, and other data forms. Unlike earlier AI systems that could only classify or analyze existing information, generative systems produce original content with similar characteristics to their training data.



This capability matters tremendously for discovery work. Traditional software can search documents for keywords, count their occurrences, and organize results. But it cannot read a document and explain what it means in context. It cannot synthesize information across hundreds of sources to answer a complex question. It cannot draft a coherent summary that captures the essence of testimony spanning multiple depositions.

Generative AI can do all of these things. The technology achieves this through advanced neural networks called Large Language Models, typically trained on billions of text examples. Through this training, these models learn to recognize and replicate patterns and structures in language, enabling them to create content that closely resembles human-generated work.

The key insight is that language follows patterns. Words don't appear randomly in sequence—they follow grammatical rules, semantic relationships, and contextual conventions. "The judge issued a" will likely be followed by words like "ruling," "decision," or "order," not "banana" or "combustion." Large Language Models learn these patterns at a scale that allows them to generate remarkably coherent and contextually appropriate text.

© 2025 Merlin Search Technologies, Inc.

At their core, LLMs perform one deceptively simple task: predicting what word should come next in a sequence based on the words that came before. This prediction process, when repeated across trillions of examples during training, produces systems capable of sophisticated language understanding and generation.

What are LLMs and Why They Matter?

Because of the extensive training and expansive computational resources required, these generative AI systems are called Large Language Models (LLMs). The term captures two essential characteristics: they are large both in terms of their neural network architecture and the volume of data they process, and they specifically focus on language understanding and generation.

GPT (Generative Pre-trained Transformer) is an LLM, but many others now exist. The current landscape includes Anthropic's Claude, Google's Gemini, Meta's Llama, Mistral, and hundreds of other models. Some are proprietary and available only from their publishers through cloud services. Others are available through open-source licenses, allowing organizations to host them independently.

For discovery professionals, several characteristics distinguish LLMs from earlier legal technology:

Contextual understanding: LLMs don't just match keywords—they understand context, nuance, and relationship between concepts. A search for "apple" in traditional systems returns every document containing that word, regardless of whether it discusses fruit, technology companies, or the "apple of someone's eye." An LLM understands which meaning applies based on surrounding context.

Natural language interaction: You can communicate with LLMs using plain English questions and instructions rather than Boolean search syntax or specialized query languages. This dramatically lowers the technical barrier for sophisticated document analysis.

Synthesis across sources: LLMs can read multiple documents and synthesize their content into coherent narratives, identifying themes, contradictions, and relationships that might take human reviewers days to assemble.

Adaptability: The same LLM can perform document summarization, extract structured data, answer questions, draft correspondence, or analyze contracts. This versatility makes LLMs fundamentally different from purpose-built legal software that handles only specific tasks.

Scalability: Once deployed, LLMs can analyze thousands of documents as easily as one, maintaining consistent quality across large document populations.

These capabilities make LLMs particularly valuable for discovery work, where professionals routinely confront massive document volumes, tight deadlines, and complex questions requiring synthesis across sources.

© 2025 Merlin Search Technologies, Inc.



LLM models come in different sizes with different skills and intelligence levels.

GPT and the Transformer Architecture

The term “GPT” stands for Generative Pre-trained Transformer. OpenAI coined this name for its groundbreaking language model, but the acronym has become somewhat generic, describing the architectural approach that most modern LLMs employ.

Let’s unpack what each component means:

Generative: The model creates original text rather than simply selecting from pre-existing options. When you ask a question, the system generates its response word by word, choosing each successive word based on probability distributions learned during training.

Pre-trained: Before the model ever interacts with end users, it undergoes extensive training on vast text collections. This pre-training phase is where the model learns language patterns, factual information, and reasoning capabilities. The “pre” distinguishes this foundational training from later fine-tuning.

Transformer: This refers to the neural network architecture underlying the model. Transformers, introduced in a 2017 research paper titled “Attention Is All You Need,” revolutionized natural language processing through a mechanism called “attention.” This allows the model to weigh the importance of different words in context, understanding that in the phrase “the bank of the river,” the word “bank” relates more closely to “river” than to financial institutions.

Different versions of these models are identified by numbers and names. OpenAI’s progression includes GPT-3, GPT-3.5, GPT-4, GPT-4 Turbo, and GPT-4o. The numbers indicate major architectural improvements and expanded capabilities. Letters like “o” (in GPT-4o, standing for “omni”) often signify specific enhancements—in this case, multimodal capabilities handling

© 2025 Merlin Search Technologies, Inc.

text, images, and audio.

Competing models follow similar naming conventions. Anthropic's Claude progressed from Claude 1 through Claude 2 to Claude 3 (with variants Opus, Sonnet, and Haiku indicating different capability-cost trade-offs), and now Claude Sonnet 4 and 4.5. Google's Gemini, Meta's Llama, and others maintain their own versioning schemes.

For practical purposes, you need not master these technical distinctions. What matters is understanding that different models have different strengths, and newer versions generally outperform their predecessors. When evaluating AI tools for discovery work, pay attention to which underlying models they use and how recently those models were released.

Training Process: From Data to Intelligence

Understanding how LLMs are trained helps explain both their remarkable capabilities and their notable limitations. The training process typically involves two major stages: pre-training and fine-tuning.

Pre-training: Learning from Massive Text Collections

The first stage involves exposing the model to enormous amounts of text—typically hundreds of billions of words drawn from books, articles, websites, code repositories, and other sources. During this unsupervised learning phase, the model doesn't receive explicit instruction about what different texts mean. Instead, it learns by repeatedly predicting the next word in sequences and adjusting its internal parameters when predictions prove incorrect.

Imagine showing someone millions of legal documents and asking them to guess the next word in each sentence, providing feedback on each guess. Over time, that person would develop intuitions about legal language patterns: what terms typically follow "motion to," how contract provisions are structured, which arguments courts find persuasive. LLMs learn similarly, but at a scale impossible for humans—processing trillions of word predictions during training.

This process is extraordinarily expensive. Training GPT-4 reportedly cost around $100 million, including cloud computing expenses for the specialized hardware clusters required. The training consumed months of computing time on systems employing tens of thousands of powerful GPUs (graphics processing units) originally designed for rendering video but repurposed for AI calculations.

The result is a model that has internalized vast amounts of information and learned sophisticated language patterns. It "knows" (in a statistical sense) facts about history, science, law, culture, and countless other domains. It can complete sentences naturally, answer questions, and generate coherent text on diverse topics.

Fine-tuning: Refining Through Human Feedback

Pre-training alone produces models that understand language but don't always behave helpfully. They might generate toxic content, refuse straightforward requests, or provide responses that are technically accurate but practically useless. Fine-tuning addresses these issues.

© 2025 Merlin Search Technologies, Inc.

During fine-tuning, humans interact extensively with the model, asking questions and providing feedback on responses. This process, often called Reinforcement Learning from Human Feedback (RLHF), teaches the model to generate responses that are helpful, harmless, and honest. Thousands of hours of human interaction guide the model toward more useful behaviors.

For legal applications, fine-tuning proves particularly important. A model trained only on general internet text might not understand professional conventions, confidentiality concerns, or the precision required in legal contexts. Fine-tuning helps models learn these domain-specific requirements.

Some vendors create specialized legal LLMs through additional fine-tuning on legal texts. Others use general-purpose models but carefully craft prompts to invoke appropriate behavior for legal tasks. Both approaches can produce effective tools, though they involve different trade-offs in cost, customization, and maintenance.

Training Cutoff: A Critical Limitation

One crucial aspect of LLM training is the cutoff date. During training, the model learns from data available up to a specific point in time. Once training completes, the model's parameters are fixed. It cannot learn from new information or update its knowledge base.

For example, if an LLM's training data cutoff is January 2025, it will have no knowledge of events, developments, or information emerging after that date. This creates several implications for legal work:

Current events: The model cannot answer questions about recent news, new regulations, or fresh case law without additional mechanisms to access current information.

Document analysis: The model has never seen your client's documents, internal communications, or case-specific materials, regardless of when they were created. Even documents from 2020 are unknown to the model if they weren't part of its training data.

Legal research: For recent cases or statutory changes, the model lacks native knowledge and may confidently provide outdated information.

This limitation is why most practical legal AI applications employ Retrieval Augmented Generation (RAG), which we explore in Chapter 3. RAG systems provide the model with current information from external sources, overcoming the training cutoff limitation.

Understanding training cutoff also helps explain another characteristic: LLMs have no memory of conversations. The trained model is static, like a "brain in a jar" that cannot form new memories or learn from interactions. When you submit a question, the model processes it and generates a response, but it retains nothing once that response is complete. This has important security implications we address in Chapter 4.

© 2025 Merlin Search Technologies, Inc.

LLMs as Modern Supercomputers

The computational requirements for running LLMs rival those of traditional supercomputers. This matters for understanding both the capabilities and the economics of these systems.

LLMs require specialized hardware called GPUs (Graphics Processing Units). Originally designed for rendering video game graphics, these chips excel at the parallel mathematical operations required for neural network computations. A single high-end GPU suitable for LLM work costs over $100,000. The largest models require clusters of tens of thousands of these chips operating in concert.

Microsoft reportedly deployed 25,000 NVIDIA A100 GPUs for its GPT-4 implementation. The hardware alone represents hundreds of millions of dollars in capital investment. Operating costs add substantially to this figure—running an advanced LLM like GPT-4o reportedly costs over a million dollars daily in electricity and infrastructure expenses.

Why such massive requirements? The prediction process that makes LLMs valuable involves extraordinarily complex mathematical operations. Each word prediction might involve trillions of calculations. Generating a single page of text can require more computational work than running traditional database queries across millions of records.

This computational intensity has several practical implications:

Cloud delivery: Most legal professionals will access LLMs through cloud services rather than running them locally. The infrastructure costs make on-premise deployment impractical for all but the largest organizations.

Latency considerations: Even with massive computing power, generating lengthy responses takes time. When planning workflows, account for the seconds or minutes required for complex analyses.

Cost structures: AI vendors typically charge based on the amount of text processed (measured in tokens). Understanding these costs matters when designing discovery workflows that might analyze millions of documents.

Environmental impact: The energy consumption of large AI systems raises sustainability questions. Some organizations factor this into technology decisions.

Efforts continue to develop more efficient models that can run on smaller infrastructure. Apple reportedly works on models that can operate partially on mobile phones. Some open-source models can run on single high-end servers. But building and training the most capable models still requires supercomputer-scale resources.
For practical purposes, think of these LLMs as extraordinarily powerful computational resources accessible through simple interfaces. You interact with them through text, but behind that interface operates computing infrastructure of unprecedented sophistication and scale.

The Current Model Landscape

As of this writing, several major LLM families dominate the legal technology landscape. Understanding their characteristics helps you evaluate AI tools and platforms.

© 2025 Merlin Search Technologies, Inc.

OpenAI GPT Series: OpenAI's models, particularly GPT-4o and its variants, remain widely used in legal applications. These models excel at complex reasoning and instruction-following. They power numerous legal AI tools through OpenAI's API. GPT-4o offers strong performance across diverse tasks with reasonable cost-effectiveness. The "o" designation indicates omni-modal capability—handling text, images, and eventually audio in integrated fashion.

Anthropic Claude Series: Claude Sonnet 4 and 4.5 have earned strong reputations in legal circles for careful, nuanced responses. Anthropic emphasizes safety and reliability, appealing to legal professionals concerned about accuracy. Claude models generally produce fewer hallucinations than some alternatives, though no model is immune to this issue. Claude Sonnet 4.5 in particular has gained adoption for complex legal analysis tasks. Anthropic also offers extended context windows, allowing analysis of very long documents in single sessions.

Google Gemini: Google's entry into the LLM space brings substantial resources and integration with Google's ecosystem. Gemini models offer competitive performance and particularly strong multimodal capabilities. For organizations already using Google Workspace, Gemini integration provides convenient access. Gemini also offers extensive context windows approaching or exceeding one million tokens in some configurations.

Meta Llama Series: Meta releases its Llama models under open-source licenses, making them available for self-hosting and customization. This appeals to organizations with data sovereignty concerns or specific fine-tuning needs. While requiring more technical expertise to deploy, Llama models offer flexibility unavailable with purely cloud-based alternatives. Several legal tech vendors build on Llama foundations, adding legal-specific fine-tuning.

Specialized Legal Models: Some vendors have developed models fine-tuned specifically for legal work. These typically start with base models like those above but undergo additional training on legal texts and tasks. Examples include Harvey's legal-specific models and various court-specific tools. Whether specialized models outperform general-purpose LLMs with good prompting remains debated.

Model Characteristics to Consider:

When evaluating LLMs for discovery applications, several factors matter:

Reasoning capability: How well does the model handle complex logical tasks, multi-step problems, and nuanced interpretation? This varies significantly across models and versions.

Accuracy and reliability: What is the hallucination rate? How often does the model confidently assert incorrect information? This proves particularly critical for legal work.

Context window size: How much text can the model process simultaneously? Larger windows enable analyzing longer documents or more sources together, but don't assume bigger is always better—there are trade-offs we explore in Chapter 2.

Speed and cost: How quickly does the model generate responses, and what does it cost per token processed? These factors affect practical scalability for large discovery projects.

© 2025 Merlin Search Technologies, Inc.

Multimodal capabilities: Can the model handle images, charts, and other non-text content? This matters when analyzing document productions containing diverse file types.

API reliability and support: For production legal use, consistent availability and responsive support matter. Downtime during critical deadlines is unacceptable.

Data handling and security: Does the vendor commit to zero-data retention? Are there adequate security certifications? We address these questions thoroughly in Chapter 4.

The Rapid Pace of Change:

The model landscape evolves faster than book publishing allows. New models emerge monthly. Capabilities improve continuously. Costs decline as efficiency increases. By the time you read this, newer models will likely have superseded those we describe.

Rather than memorizing current model names and specifications, focus on understanding the evaluation framework. When new models arrive, assess them using the criteria above. Ask vendors specific questions about reasoning capability, accuracy, security, and costs. Test models on representative tasks from your actual workflows.

The fundamental concepts underlying LLMs remain relatively stable even as specific implementations advance. Understanding how these systems work, what they can and cannot do, and how to evaluate them provides durable knowledge that survives the next generation of models.

Looking Ahead

This chapter established the foundation: LLMs are sophisticated statistical models trained on vast text collections to predict and generate language. They achieve remarkable capabilities through massive computational resources and careful training processes. Multiple competing models exist, each with distinct characteristics and trade-offs.

With this foundation, we can now explore how these systems actually interact with users and handle extended conversations. Chapter 2 examines context windows—the mechanism through which LLMs receive information and maintain apparent memory across multi-turn interactions. Understanding context windows proves essential for designing effective discovery workflows and recognizing the practical limits of what LLMs can accomplish in single sessions.

The technology we've described in this chapter enables the practical applications we explore later. But between understanding how LLMs work and successfully deploying them in discovery lies important knowledge about their operational characteristics, limitations, and risks. The next chapters build that knowledge systematically, preparing you to use these powerful tools effectively and responsibly.

© 2025 Merlin Search Technologies, Inc.

MERLIN

02

Context Windows and Conversation Management

© 2025 Merlin Search Technologies, Inc.

CHAPTER TWO

Context Windows and Conversation Management

In Chapter 1, we described Large Language Models as having no memory—unable to learn from interactions or retain information after processing. Yet when you use ChatGPT or similar interfaces, you can carry on extended conversations where the AI seems to remember everything you discussed. These observations appear contradictory. How can a system with no memory maintain coherent multi-turn dialogues?

The answer lies in understanding context windows and the relationship between LLMs and the interfaces through which we access them. This chapter explores these mechanisms, explaining how AI systems manage conversations, what limitations these mechanisms impose, and what implications they hold for discovery work.

The Context Window Metaphor: The Whiteboard Outside the Brain

Recall our description of a trained LLM as a "brain in a jar"—a powerful analytical engine that processes information but retains nothing. This metaphor captures an essential truth about how LLMs operate, but it requires extension to explain how conversations work.

Imagine that the brain in a jar has access to a whiteboard positioned outside it. The brain can read what appears on the whiteboard and write responses back, but the whiteboard itself sits in external memory, separate from the brain's internal structure. Once the brain finishes processing information and writing its response, that response is erased from its awareness. The brain returns to its baseline state, ready for the next request, with no memory of what just happened.

This whiteboard is the context window. It serves as the communication interface between the external world and the LLM's processing core. Everything the LLM considers when generating a response must appear on this whiteboard. Conversely, the LLM can only write its response to this same space.

The metaphor helps explain several important characteristics:

Limited space: Whiteboards have a finite area. Similarly, context windows can hold only a limited amount of text. We measure this capacity in tokens—roughly equivalent to words, though the precise mapping is more complex.

Temporary storage: When you erase a whiteboard, the information disappears. When an LLM finishes processing a request, the context window contents are cleared, ready for the next interaction.

External to the core: The whiteboard exists outside the brain in the jar. Similarly, context windows sit in temporary memory outside the LLM's trained parameters. This separation explains

© 2025 Merlin Search Technologies, Inc.

why conversations don't alter the model itself.

Single view: The brain can only process what currently appears on the whiteboard. If earlier conversation exists, it must be rewritten on the whiteboard for the brain to consider it. The brain cannot reach back to remember—it can only read what's presently visible.

This architecture has profound implications for how we design discovery workflows. Everything you want the LLM to consider—your instructions, the documents to analyze, the questions to answer, and the conversation history—must fit within this context window simultaneously.

How ChatGPT and Other Interfaces Work

Understanding the distinction between interfaces like ChatGPT and the underlying LLMs clarifies how conversations are maintained despite the LLM's lack of memory.

The Interface Layer

ChatGPT is not GPT. This seemingly obvious point bears emphasizing because many users conflate the two. ChatGPT is a software application that provides a browser-based interface for interacting with GPT. The distinction matters because the interface handles several functions beyond what the LLM itself does:

Conversation storage: ChatGPT saves your complete conversation history in its database. You can close your browser, return days later, and continue exactly where you left off. The LLM itself retains nothing—all memory resides in the interface.

Context window management: When you submit a new message in an existing conversation, ChatGPT retrieves the prior conversation history and sends it to GPT along with your new message. This happens automatically and invisibly. From your perspective, you're just continuing a conversation. From GPT's perspective, it's receiving a completely new request that happens to include contextual information.

Session management: The interface tracks multiple conversations, allows you to organize them, search through them, and delete them. It manages authentication, billing, and usage limits. None of this involves the LLM directly.

User experience: The interface formats responses for readability, provides convenient copying and export options, and creates the familiar chat-like interaction pattern. The LLM just generates text—the interface makes that text useful.

Other AI interfaces work similarly. Claude's web interface, Microsoft's Copilot, Google's Gemini interface, and countless specialized legal AI tools all follow this pattern: an interface layer manages conversation state and user experience while routing requests to underlying LLMs.

How a Conversation Actually Works

When you start a conversation by typing a question, here's what happens:

1. **You submit a prompt:** Your message goes to the interface (ChatGPT, Claude, etc.).

© 2025 Merlin Search Technologies, Inc.

2. **The interface prepares the context window:** The interface takes your prompt and writes it to the context window, which will be sent to the LLM.

3. **The LLM processes:** GPT (or Claude, or Gemini) reads the context window, processes the information, and generates a response. It writes this response back to the context window.

4. **The interface retrieves and displays:** The interface reads the LLM's response from the context window, displays it to you, and stores both your prompt and the response in its database.

5. **Context window is cleared:** Once the interface retrieves the response, the context window is cleared, ready for the next request.

Now suppose you want to continue the conversation. You type a follow-up question. Here's what happens:

1. **You submit a follow-up:** Your new message goes to the interface.

2. **The interface reconstructs context:** The interface retrieves your entire conversation history from its database—all previous messages from both you and the AI.

3. **The interface prepares the context window:** It writes the full conversation history plus your new message onto the context window.

4. **The LLM processes as if new:** GPT receives what appears to it as a completely new request. It has no memory of responding to your earlier questions. But because the context window contains the full conversation history, it can generate a response that appropriately continues the dialogue.

5. **The cycle repeats:** Response generated, displayed, stored, context window cleared.

This explains how LLMs can "remember" conversations despite having no memory. They don't remember—they're simply provided with the complete conversation history every time they generate a response. The illusion of memory comes entirely from the interface's storage and context window management.

Important Implications

This architecture has several implications worth noting:

Each request is independent: From the LLM's perspective, every request is brand new. It doesn't gradually learn about your case or remember your preferences (though we'll discuss the recent "memory" feature later). Each response is generated fresh from the context window contents.

Different compute nodes: Your second request in a conversation might be handled by a completely different instance of GPT running on different hardware. The interface routes requests to available compute resources. This works seamlessly because everything needed sits in the context window, not in any persistent state.

© 2025 Merlin Search Technologies, Inc.

Interface differences matter: Different interfaces to the same LLM can produce different user experiences. One might summarize long conversations before sending them to the LLM. Another might send complete verbatim history. These choices affect both functionality and cost.

Data persistence: While the LLM retains nothing, the interface stores your complete conversation history. For security and confidentiality, understanding where and how interfaces store data matters greatly. We explore this in Chapter 4.

Context Window Evolution: From 4k to 1M+ Tokens

The size of context windows has expanded dramatically since ChatGPT's 2022 release. Understanding this evolution helps contextualize current capabilities and anticipate how growing windows might affect discovery workflows.

The Early Days: 4k Tokens

When GPT-3.5 first powered ChatGPT, it offered a 4,096-token context window. This translates to roughly 3,000 words or about six single-spaced pages. While revolutionary for its time, this limitation significantly constrained practical applications.

Consider discovery implications: You couldn't ask the system to analyze even a moderately long deposition transcript. A typical single-day deposition might run 150-300 pages, far exceeding the available space. Contract review faced similar constraints. Most commercial contracts exceed a few pages, and comparative analysis across multiple contracts was impossible.

The conversation history consumed context window space alongside any documents you wanted analyzed. A lengthy back-and-forth would exhaust the available capacity, forcing the interface to truncate early conversation turns. The LLM would begin "forgetting" the start of your discussion.

Rapid Expansion: 8k, 16k, 32k

Recognition of these limitations drove rapid development. Within months, context windows expanded to 8k, then 16k, then 32k tokens. Each doubling enabled new use cases.

At 32k tokens (approximately 24,000 words), you could analyze substantial documents—perhaps a short deposition or several contracts simultaneously. This opened practical applications in legal review that simply weren't feasible at 4k.

The Jump to 128k and Beyond

In late 2023 and early 2024, another leap occurred. GPT-4 Turbo arrived with a 128k token context window. Anthropic released Claude 3 with 200k tokens. These capacities changed the game substantially.

At 128k tokens (roughly 96,000 words), you can analyze:

© 2025 Merlin Search Technologies, Inc.

- Most single-day deposition transcripts completely
- Several average-length contracts simultaneously
- Lengthy reports or legal memoranda
- Substantial portions of document productions

At 200k tokens, capacity grows to approximately 150,000 words—enough for multiple transcripts, extensive contract portfolios, or comprehensive document collections.

Both Anthropic and Google announced even larger experimental windows, approaching or exceeding one million tokens. Anthropic demonstrated Claude analyzing "The Great Gatsby" complete with scripts from all nine Star Wars films—pure marketing, but illustrative of scale. Google's Gemini similarly claimed a million-token capacity.

Current State: 128k-200k as Standard

As of this writing, most advanced LLMs offer context windows in the 128k-200k token range as their standard configuration. This represents roughly 100,000-150,000 words, or 200-300 pages of text. For many discovery tasks, this proves sufficient to analyze substantial documents or collections.

The million-token windows remain somewhat experimental. They're available in limited configurations, often with substantial cost premiums, and questions persist about how effectively LLMs use such extensive context.

The "Lost in the Middle" Problem

Larger context windows solve some problems but create others. Research has revealed a concerning pattern: LLMs often struggle to use information effectively when it appears in the middle of very long contexts.

What the Research Shows

Studies testing LLM performance on question-answering tasks across varying context lengths found consistent patterns. When relevant information appears at the beginning or end of the context window, models retrieve and use it effectively. But when the same information appears in the middle—particularly in very long contexts—performance degrades significantly.

Imagine asking an LLM to find a specific contractual provision buried in page 47 of a 100-page context window containing multiple contracts. Research suggests the model might miss it, even though the information technically sits within its context window. The same provision appearing in the first few pages or final few pages stands much higher chance of being noticed and used.

This "lost in the middle" phenomenon appears to worsen as context windows grow. Paradoxically, adding more capacity sometimes reduces effective performance if critical information gets buried in the expanded middle section.

Why This Happens

© 2025 Merlin Search Technologies, Inc.

The technical explanation involves how transformer architectures process sequence information and how attention mechanisms distribute across very long contexts. Without diving into neural network details, the essential point is that LLMs don't maintain truly uniform awareness across their entire context window. Certain positions—beginnings and ends especially—receive disproportionate attention.

Think of it as analogous to human reading comprehension. If you skim a very long document quickly, you'll likely remember points from the introduction and conclusion better than random facts from the middle chapters. LLMs exhibit similar patterns, though for different underlying reasons.

Implications for Discovery Work

This limitation has practical consequences for designing discovery workflows:

Document placement matters: When analyzing multiple documents, consider placing the most important ones at the beginning or end of the context window rather than in the middle.

Explicit references help: Rather than asking "What does the discovery plan say about privilege review?" and hoping the LLM finds the relevant section buried mid-document, you might say "Review the section on page 12 discussing privilege review and summarize it." Explicit direction helps overcome position bias.

Chunking strategies remain relevant: Despite larger windows, breaking long documents into focused chunks and analyzing them separately sometimes produces better results than feeding everything to a massive context window at once.

Verification becomes critical: The lost-in-the-middle problem means you cannot assume the LLM considered all information in a large context window. Verification of key findings against source documents remains essential.

Window size optimization: Bigger isn't always better. A focused analysis of carefully selected documents in a moderate context window might outperform a kitchen-sink approach that dumps everything into a million-token window.

Practical Implications for Discovery Work

Understanding context windows helps you design more effective discovery workflows. Here we explore when different approaches work best and how to work within context window constraints.

When Larger Windows Help

Single-document analysis: When you need comprehensive analysis of one substantial document—a lengthy contract, complete deposition, or detailed report—larger context windows allow the LLM to consider the entire document at once. This preserves context that might be lost if you analyzed the document in chunks.

Comparative analysis: Comparing provisions across multiple contracts, finding contradic-

© 2025 Merlin Search Technologies, Inc.

tions between witness testimonies, or identifying patterns across documents all benefit from having everything visible simultaneously in the context window.

Conversation continuity: In complex investigations requiring extensive back-and-forth with the AI, larger windows allow more conversation history to remain accessible without truncation. This maintains continuity across dozens of exchanges.

Reduced API calls: Analyzing everything in one large context window rather than multiple smaller ones reduces the number of API calls needed, potentially lowering costs and simplifying workflow architecture.

When RAG/Chunking Strategies Work Better

Very large document collections: No context window, no matter how large, can hold millions of discovery documents. When analyzing substantial document populations, retrieval-based strategies remain essential. The system searches to find relevant documents, then analyzes those specifically rather than attempting to cram everything into the context window.

Precision over breadth: Sometimes you don't need everything in view simultaneously. If you're searching for specific types of clauses across a contract portfolio, retrieving and analyzing just those sections might prove more accurate than presenting all contracts together and asking the LLM to find relevant provisions.

Cost management: Larger context windows cost more to use. Each token sent to the LLM incurs charges. For large-scale review, the costs of repeatedly sending huge contexts can become prohibitive. Targeted retrieval followed by focused analysis often proves more economical.

Avoiding lost-in-the-middle: As discussed earlier, extremely large contexts can reduce effectiveness. Strategic chunking that presents information in digestible pieces sometimes produces better results than maximum-size windows.

Iterative investigation: Discovery investigations often proceed iteratively. You find some information, which raises new questions, leading to new searches. This natural workflow pattern doesn't require everything in view at once. Instead, you benefit from targeted retrieval in each phase, analyzing focused sets of relevant documents as you progress.

Hybrid Approaches

Many effective discovery workflows combine approaches. Consider this pattern:

1. **Initial broad search:** Use semantic search or AI classification to identify potentially relevant documents from a large collection.

2. **Cluster and prioritize:** Group similar documents and rank by likely relevance.

3. **Focused analysis:** Analyze high-priority document clusters using appropriately-sized context windows—large enough for meaningful synthesis, small enough to avoid lost-in-the-middle problems.

© 2025 Merlin Search Technologies, Inc.

4. **Iterative refinement:** Based on findings, conduct targeted searches for specific information, analyzing results in focused contexts.

5. **Synthesis:** Combine insights from multiple focused analyses into comprehensive understanding.

This hybrid approach leverages both retrieval methods (for finding documents) and context window capacity (for analyzing them), avoiding the weaknesses of either extreme.

Cost Considerations with Large Contexts

Context window size directly affects cost. AI providers typically charge based on tokens processed, with separate rates for input tokens (what you send) and output tokens (what the AI generates). Larger contexts mean more input tokens, increasing costs.

Consider the economics:

Small context (8k tokens): Analyzing a short document costs perhaps a few cents. Conducting thousands of such analyses across a document collection remains affordable.

Medium context (128k tokens): Analyzing a substantial document or small collection might cost several dollars per analysis. Scaling this across thousands of analyses becomes expensive but potentially justifiable for complex matters.

Large context (1M tokens): A single analysis might cost tens of dollars. This makes sense for critical documents requiring comprehensive analysis but proves impractical for routine review across large populations.

Conversation history: Remember that every message in an ongoing conversation includes all prior messages as context. A conversation with 50 exchanges might accumulate substantial token costs as the context window fills with history. Some interfaces implement strategies to manage this—summarizing old conversation turns, pruning less relevant history, or using separate sessions for distinct topics.

For discovery budgets, these cost structures suggest tailoring context window usage to task importance. Reserve large context windows for critical analyses where comprehensive synthesis matters. Use smaller, focused contexts for routine review tasks. Design workflows that minimize redundant token usage—don't resend the same documents repeatedly if you can cache analyses.

Why LLMs Have No Memory (And Why That Matters)

We've explained that LLMs don't remember conversations—the interface handles that through context window management. But this characteristic deserves deeper exploration because it has security and practical implications.

Technical Reality: Fixed Parameters

When an LLM completes training, its neural network parameters are set. These parameters—

essentially the “weights” that determine how the model processes input and generates output—remain fixed during inference (the process of generating responses to prompts).

Your prompts don’t modify these parameters. The model reads your input, processes it through its fixed neural network, generates output, and returns to exactly the same state afterward. It’s genuinely stateless between requests.

This differs fundamentally from how humans learn through conversation. When you explain something to a colleague, they retain that information. Their understanding evolves. An LLM doesn’t. Each interaction is independent, processed through the identical neural network structure.

The “Memory” Feature Caveat

Recent versions of ChatGPT and some other interfaces offer a “memory” feature that might seem to contradict what we’ve said. This feature allows the interface to maintain persistent facts about you across conversations—your preferences, your role, recurring topics you discuss.

However, this “memory” isn’t LLM learning. Instead, the interface maintains a separate database of facts about you. When you start a new conversation, the interface includes relevant facts from this database in the context window sent to the LLM. From the LLM’s perspective, these are just more instructions in the current context window, not actual memories.

This distinction matters because:

The LLM still knows nothing about you: Different users querying the same LLM with identical prompts get identical responses (accounting for randomness in text generation). Your history doesn’t affect the underlying model.

The interface controls memory: Interface providers decide what to save, how long to retain it, and when to include it in contexts. The LLM has no agency in this process.

Security implications remain unchanged: The LLM cannot leak your information to other users because it never stored that information in the first place. We explore this crucial security point in Chapter 4.

Practical Implications

Understanding that LLMs lack memory helps you use them effectively:

Provide complete context: Don’t assume the LLM remembers earlier statements unless they appear in the current conversation. If you discussed a case yesterday in one conversation and start a new conversation today, you must reintroduce the case context.

Conversation organization: Since conversations maintain independent context, you might deliberately use separate conversations for separate matters. This prevents unrelated context from consuming token space and keeps discussions focused.

© 2025 Merlin Search Technologies, Inc.

Session management: Some discovery workflows benefit from long-running sessions where extensive context accumulates. Others work better with frequent fresh starts that clear context and begin from clean states. Understanding the stateless nature helps you choose appropriately.

Reproducibility: Because the LLM has no state beyond what's in the current context window, providing identical inputs produces consistent outputs (within the bounds of generation randomness). This supports validation and verification workflows.

Training doesn't happen during use: You cannot "teach" an LLM through conversation. Fine-tuning requires returning to the training process with new data. For legal applications, this means domain knowledge must come from careful prompting, retrieval of relevant information, or using models specifically fine-tuned for legal work before deployment.

Looking Ahead

This chapter examined how LLMs interact with users through context windows managed by interface layers. We explored the evolution of context window sizes, the lost-in-the-middle problem, and practical implications for discovery workflows. Understanding these mechanisms allows you to design more effective applications of AI in legal work.

But even with this knowledge, a critical concern remains: accuracy. LLMs can generate impressively fluent text that's factually wrong, a phenomenon called hallucination. For legal professionals, where precision and reliability are paramount, this represents perhaps the most serious challenge in AI adoption.

Chapter 3 addresses hallucinations comprehensively—what they are, why they occur, how to detect them, and most importantly, how to structure workflows that minimize their risk while preserving the efficiency gains AI offers. Understanding hallucinations isn't optional for responsible AI use in discovery. It's foundational to maintaining professional standards while leveraging these powerful tools.

© 2025 Merlin Search Technologies, Inc.

MERLIN

03

Hallucinations and Accuracy

© 2025 Merlin Search Technologies, Inc.

CHAPTER THREE

Hallucinations and Accuracy

The most serious concern about using Large Language Models in legal work isn't their cost, complexity, or learning curve. It's their tendency to generate plausible-sounding content that is simply wrong. This phenomenon, called "hallucination," represents the fundamental challenge that legal professionals must understand and address before deploying AI systems in practice.

Unlike human mistakes, which typically involve misunderstanding, oversight, or carelessness, LLM hallucinations often appear confident and detailed. The system doesn't indicate uncertainty or flag potential errors. It generates fabricated case citations, non-existent statutes, or invented facts with the same fluency it uses for accurate information. For legal professionals accustomed to associates who acknowledge when they don't know something, this behavior can be dangerously misleading.

This chapter examines hallucinations comprehensively: what they are, why they occur, how to recognize them, and most importantly, how to structure workflows that preserve AI's efficiency benefits while minimizing accuracy risks. Understanding hallucinations isn't merely advisable for legal AI use—it's essential to maintaining professional competence and avoiding the career-damaging mistakes several lawyers have already made.

What Are Hallucinations?

A hallucination occurs when an LLM confidently generates detailed information that is factually incorrect, fabricated, or cannot be verified in its training data or provided context. The term "hallucination" captures the phenomenon well: the system appears to perceive information that doesn't exist, much as a hallucinating person might see things that aren't there.



Hallucinations occur when the LLM makes up the answer to your question.

Key Characteristics of Hallucinations

Plausibility: Hallucinated content isn't random nonsense. It follows grammatical rules, fits contextual expectations, and often incorporates real elements. A hallucinated case citation includes proper formatting, realistic party names, and appropriate court identifiers. This plausibility makes hallucinations dangerous—they don't trigger immediate skepticism.

Confidence: LLMs generate hallucinated content with the same apparent certainty they use for accurate information. There's no hedging, no indication of uncertainty, no disclaimer that the information might be fabricated. The system presents invented facts as definitively as ver-

ified ones.

Detail: Hallucinations often include specific, concrete details that enhance their apparent credibility. Rather than vaguely mentioning "a federal case," the system might cite "Johnson v. Peterson, 847 F.3d 592 (9th Cir. 2019)," complete with plausible docket numbers, judge names, and holding descriptions—none of which exist.
Internal consistency: Hallucinated information often maintains internal logical consistency. If the system invents a case, subsequent references to that case will remain consistent with the initial fabrication. This consistency across multiple references reinforces the false impression of accuracy.

Why Hallucinations Occur

Understanding why LLMs hallucinate requires remembering what they fundamentally do: predict probable next words in sequences based on training patterns. They aren't searching databases of verified facts. They're generating statistically plausible continuations of text.

When you ask an LLM about a real case it encountered during training, it generates text based on patterns it learned from thousands of case discussions. The output might accurately reflect that training data. But when you ask about something not in its training data, the system doesn't respond "I don't know." Instead, it generates plausible-sounding text following the patterns it learned, effectively inventing information that fits the expected format.

Consider this analogy: If you ask someone to describe a Supreme Court case they've never heard of, an honest person says they don't know. But imagine someone trained to always provide answers, never admit ignorance, and construct responses from general knowledge of how legal writing works. They might generate a plausible-sounding description that follows all the right patterns—procedural posture, legal issues, holding, reasoning—while inventing every substantive detail. That's essentially what LLMs do when hallucinating.

The training process teaches LLMs to generate fluent, contextually appropriate text. It doesn't teach them to distinguish between "I learned this specific fact during training" and "I can generate plausible text fitting this context." The model has no internal mechanism for evaluating whether generated content corresponds to reality. It only evaluates whether the text flows naturally and fits the pattern.

Fine-tuning through human feedback reduces hallucination frequency by teaching models to be more careful, acknowledge uncertainty, and provide better-grounded responses. But fine-tuning doesn't eliminate the fundamental tendency. The model architecture still generates text through statistical prediction, not fact verification.

Hallucination vs. Mistaken Inference

It's important to distinguish hallucinations from legitimate analytical errors. LLMs can make mistaken inferences—drawing incorrect conclusions from ambiguous text, misinterpreting nuanced arguments, or applying faulty logic to genuine information. These are analytical failures, similar to mistakes human lawyers make when reasoning from complex sources.

Hallucinations are different. They involve fabricating information wholesale. The distinction

© 2025 Merlin Search Technologies, Inc.

matters because different mitigation strategies apply:

Mistaken inferences require better prompting, clearer instructions, human review of reasoning, and verification of analytical conclusions. You might address them by providing more context, asking the system to explain its reasoning, or requesting alternative interpretations.

Hallucinations require ensuring the system bases outputs on verified source material rather than generating from training memory. You address them through retrieval-based architectures, mandatory citation requirements, and systematic verification of factual claims against source documents.

A system that reads a contract and misinterprets an ambiguous clause made an analytical error. A system that invents contract provisions that don't exist in the source document hallucinated. The first is a reasoning problem; the second is a reliability problem.

The Fake Cases Crisis: Lessons from Recent Sanctions

The legal profession's introduction to AI hallucinations came through a series of high-profile cases where lawyers submitted court filings citing non-existent cases generated by ChatGPT. These incidents provide sobering lessons about hallucination risks and the consequences of insufficient verification.

The Mata v. Avianca Case

In May 2023, lawyers Steven A. Schwartz and Peter LoDuca filed a brief in federal court in New York opposing a motion to dismiss. The brief cited multiple cases supporting their arguments, including Vargas v. China Southern Airlines Co., Shaboon v. Egyptair, and Petersen v. Iran Air. The citations appeared properly formatted with case numbers, court identifiers, and holding descriptions.

Opposing counsel couldn't locate these cases in legal databases. When questioned, the lawyers provided additional details—quotes from the purported decisions, internal citations to other cases, and explanations of the holdings. All fabricated.

The lawyers had used ChatGPT to conduct legal research. When ChatGPT generated the fake citations, they didn't verify them in actual legal databases. When opposing counsel questioned the cases, they returned to ChatGPT asking for confirmation. ChatGPT, unable to acknowledge its earlier hallucinations, provided additional fabricated details supporting the fake cases' existence.

The court was not amused. Judge Kevin Castel sanctioned both lawyers, fining each $5,000 and referring them to grievance committees for potential disciplinary action. In his opinion, Judge Castel emphasized that while using AI for legal research wasn't improper per se, lawyers maintain responsibility for ensuring accuracy of their filings. The opinion noted that even a cursory verification effort would have revealed the cases didn't exist.

The Park v. Kim Case

Shortly after the Mata sanctions, attorney Jae Lee faced similar problems in a federal appeal.

© 2025 Merlin Search Technologies, Inc.

Lee cited a non-existent state court decision in briefing to the Second Circuit Court of Appeals. The citation came from ChatGPT, and again, basic verification would have revealed the fabrication.

The Second Circuit referred Lee to a grievance panel for potential sanctions, noting the concerning pattern of lawyers submitting AI-generated fake cases without verification.

Additional Incidents

These weren't isolated incidents. Several other cases emerged involving fake citations from AI systems:

In Colorado, an attorney submitted a brief citing cases that didn't exist, later explaining he had used an AI tool without proper verification.

In Texas, another lawyer faced sanctions for similar conduct, with the court emphasizing that technological tools don't diminish attorneys' responsibility for accuracy.

The pattern across these cases is remarkably consistent: lawyers used ChatGPT or similar tools for legal research, received hallucinated cases, failed to verify them in legal databases, submitted them to courts, and faced professional consequences when the fabrications were discovered.

What Went Wrong

These incidents reveal multiple failures in understanding and using AI:

Misunderstanding AI capabilities: The lawyers apparently believed ChatGPT functioned like a legal research database—searching and retrieving actual cases. In reality, it was generating plausible-sounding text based on patterns, not retrieving verified citations.

Absence of verification: In traditional legal practice, if an associate provides case citations, prudent partners verify key cases before filing. These lawyers skipped verification, apparently trusting AI output implicitly.

Compounding errors: When questioned, several lawyers returned to ChatGPT for confirmation rather than checking actual legal databases. ChatGPT, unable to admit error, provided additional fabricated details, deepening the problem.

Inadequate understanding of hallucinations: The lawyers likely didn't understand that AI systems could generate confident, detailed, completely fabricated legal citations. They learned this expensive lesson publicly.

Professional responsibility failures: The sanctions opinions emphasized that lawyers cannot delegate their professional duty of competence to technology. Using AI tools doesn't excuse submitting false information to courts.

Lessons for Discovery Professionals

© 2025 Merlin Search Technologies, Inc.

While these cases involved litigation briefs rather than discovery, the lessons apply directly to discovery work:

Verification is mandatory: AI-generated information requires verification against source documents, regardless of how confident or detailed the output appears.

Understand tool limitations: Know what your AI tools actually do. Are they searching databases and retrieving documents? Or generating text based on patterns? The distinction is critical.

Design verification into workflows: Don't treat verification as optional or reserve it for suspicious outputs. Build systematic verification into every AI-assisted process.

Human responsibility remains: AI assistance doesn't diminish professional obligations. Discovery professionals remain responsible for accuracy, regardless of what tools they employ.

Training is essential: Organizations deploying AI tools must ensure users understand both capabilities and limitations. The sanctioned lawyers might have avoided their predicament with proper training about hallucinations.

Hallucination Risk Factors

Not all AI applications present equal hallucination risk. Understanding factors that increase or decrease risk helps you assess when to use AI, what verification strategies to employ, and where extra caution is warranted.

High-Risk Scenarios

Relying on training memory: When you ask LLMs questions that require them to recall information from training data—"What does the Jones Act say about seaman status?" or "Summarize the holding in Iqbal"—hallucination risk is high. The system might have seen relevant information during training, or it might generate plausible-sounding responses based on general patterns.

Generating citations: Any task requiring case citations, statutory references, or specific source attributions presents substantial hallucination risk. LLMs can easily generate properly formatted citations to non-existent sources.

Precise factual claims: Questions requiring specific facts, dates, numbers, or technical details invite hallucinations. "When was the EPA regulation adopted?" might receive a confidently stated but fabricated year.

Obscure or specialized topics: The less common a topic in training data, the more likely hallucinations become. Asking about a rarely cited case, an obscure regulation, or a novel legal theory increases risk that the system will generate plausible-sounding fabrications.

Complex multi-step reasoning: Tasks requiring the system to combine multiple facts, perform calculations, or execute multi-step logic without source documents present elevated hallucination risk. Each reasoning step introduces additional opportunity for error or fabrication.

© 2025 Merlin Search Technologies, Inc.

Lower-Risk Scenarios

Document summarization: When you provide source documents and ask the system to summarize them, hallucination risk decreases substantially. The system bases its response on the provided text rather than generating from training memory.

Pattern identification: Tasks like identifying document types, classifying content, or recognizing themes in provided documents present lower hallucination risk because they operate on visible input rather than recalled information.

Draft generation: Using AI to draft initial versions of documents—demand letters, deposition outlines, privilege logs—carries lower hallucination risk in terms of factual fabrication. The drafts may need substantial revision for quality, but they typically don't invent facts from nothing.

Question answering from context: When you provide documents and ask specific questions answered in those documents, hallucination risk is lower. The system can extract information from visible context rather than generating from memory.

Comparative analysis: Asking the system to compare provisions across provided contracts, identify differences between documents, or analyze patterns in visible texts reduces hallucination risk. The comparison operates on provided source material.

Mitigation Through Design

Understanding risk factors allows designing workflows that minimize exposure:

Emphasize extraction over generation: Tasks that extract information from provided documents carry less risk than tasks requiring the system to generate information from training memory.

Require citations: Mandate that every factual claim include citations to source documents. This allows verification and reduces pure fabrication.

Limit scope to provided materials: Instruct systems explicitly to base responses only on provided documents, not on training knowledge. This dramatically reduces hallucination risk. Implement verification steps: Build verification into workflows rather than treating it as optional. Have humans or secondary AI systems check key facts against source documents.

Use specialized tools appropriately: For tasks like legal research requiring verified citations, use databases designed for that purpose rather than general-purpose LLMs.

Mitigation Strategies: RAG, REG, and Verification

The discovery profession has developed several approaches to minimize hallucination while preserving AI's efficiency benefits. Understanding these strategies and their appropriate applications is essential for responsible AI deployment.

Retrieval Augmented Generation (RAG)

© 2025 Merlin Search Technologies, Inc.

RAG represents a fundamental shift in how we use LLMs. Rather than asking the system to answer questions from its training memory, RAG systems first search document collections to find relevant passages, then provide those passages to the LLM as context for generating responses.

How RAG Works:

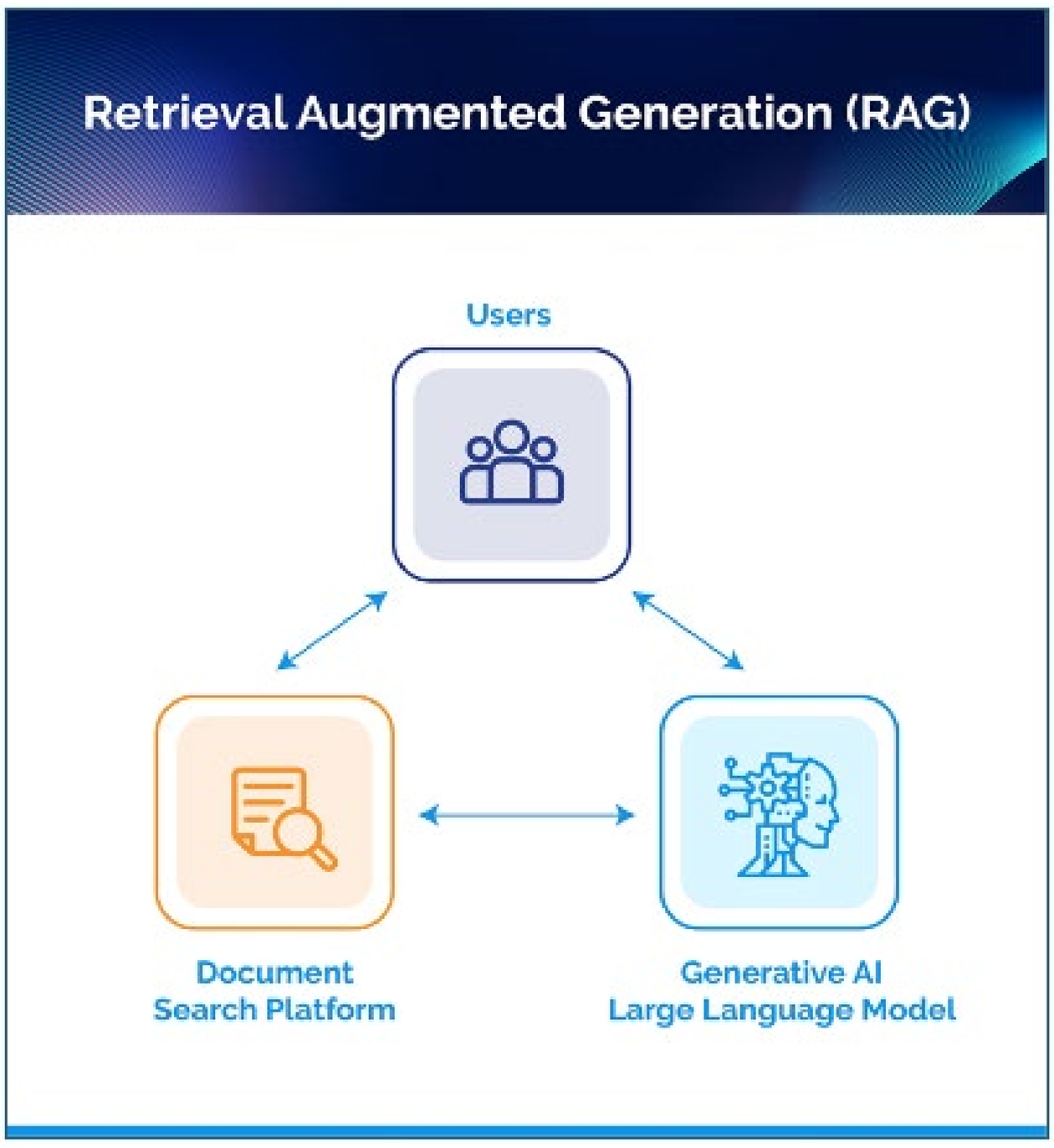
1. **User query:** You submit a question or request.

2. **Document retrieval:** The system searches a document collection using semantic search, keyword matching, or hybrid approaches to find relevant passages.

3. **Context assembly:** Retrieved passages are formatted and placed in the LLM's context window along with your query.

4. **Prompted response:** The LLM receives explicit instructions to base its response on the provided passages rather than its training memory.

5. ***Generated output:*** The system generates a response grounded in the retrieved docu-



A typical RAG architecture.

© 2025 Merlin Search Technologies, Inc.

ments.

Benefits of RAG:

Reduced hallucination: By constraining the LLM to provided source material, RAG dramatically reduces fabrication risk. The system can't invent information not present in retrieved documents.

Source attribution: RAG systems can cite specific retrieved passages, allowing verification of outputs against source material.

Current information: RAG overcomes training cutoff limitations. The document collection can include material from after the LLM's training date, making the system capable of addressing current information.

Domain specificity: RAG allows you to create domain-specific AI applications by controlling what documents the system searches. A contracts AI might search only your organization's contract repository, ensuring responses reflect your specific agreements rather than generic training data.

Challenges with RAG:

Retrieval quality matters: If the search step fails to find relevant documents, the LLM lacks information needed to answer accurately. Poor retrieval undermines the entire approach.

Context window limits: Only top-ranked retrieved documents fit in the context window. Important information in lower-ranked results might be missed.

Complexity: RAG systems require more infrastructure than simple LLM queries—document collections, search capabilities, orchestration logic, and careful prompt engineering.

Retrieval Exclusive Generation (REG)

Building on RAG principles, we developed an even more stringent approach called Retrieval Exclusive Generation. REG emphasizes that systems should use exclusively retrieved information, completely eliminating reliance on the LLM's pre-trained knowledge.

How REG Differs from RAG:

Explicit exclusivity instructions: We instruct LLMs twice—once in system-level instructions and again in user prompts—to base analysis solely on provided documents, explicitly prohibiting use of training knowledge.

Mandatory source citations: Every substantive claim in the output must include direct citations to source documents with specific identifiers allowing immediate verification.

Verification protocols: REG implementations include systematic verification steps where humans or secondary systems check that cited sources actually support claimed information.

© 2025 Merlin Search Technologies, Inc.

Rejection of outside knowledge: If the system cannot answer a question based on provided documents, it must acknowledge this rather than supplementing with training knowledge.

Discovery Applications:

The book's earlier examples demonstrate REG in practice. In the Bacardi trademark investigation example, the system searched a large document collection, retrieved relevant documents, analyzed only those documents, and generated a report with citations to specific document IDs. Every claim in the report could be traced directly to source material.

Similarly, in the transcript analysis example, the system searched across deposition transcripts, retrieved relevant sections, and generated answers citing specific transcript sections and page numbers. Nothing in the report came from the LLM's general training—everything derived from the provided transcripts.

Benefits of REG:

Maximum accuracy: By eliminating training memory as a source, REG minimizes hallucination risk to near-zero for factual claims.

Complete verifiability: Every claim includes citations allowing immediate verification against source documents.

Defensibility: When questioned about AI-assisted work product, you can demonstrate that outputs derive entirely from case materials, not from AI training data of uncertain provenance.

Professional standards: REG aligns with legal professional responsibilities to base work on verified sources rather than unverifiable AI-generated content.

Citation Requirements

Regardless of whether you implement full RAG or REG, requiring citations in AI outputs provides crucial benefits:

Verification capability: Citations allow you to check whether source documents actually support claimed information. Without citations, verification requires re-searching source materials, often proving impractical.

Hallucination detection: Hallucinations often become apparent when you attempt to verify citations. The cited document doesn't exist, doesn't say what's claimed, or doesn't support the conclusion drawn.

Quality signal: Systems that provide detailed citations tend to generate more accurate outputs. The act of citing sources appears to encourage more careful generation.

Professional documentation: Citations create the audit trail needed to demonstrate that work product derives from case materials rather than AI fabrication.

© 2025 Merlin Search Technologies, Inc.

Implementation approaches:

Mandatory citation prompts: Explicitly instruct the system to cite sources for every factual claim. “Provide citations to source documents for all substantive statements” should appear in your prompts.

Citation formats: Specify citation formats that allow easy verification. Document IDs, page numbers, section references, and file names all help locate source material quickly.

Citation validation: Build workflows that check cited sources. This might involve human review of key citations or automated systems that verify cited documents exist and contain claimed information.

Rejection of uncited claims: Treat any factual claim without citation as suspect. Either require the system to provide citations or remove uncited claims from outputs.

Verification Protocols

Even with RAG, REG, and citation requirements, human verification remains essential. No technical architecture eliminates the need for professional oversight.

What to verify:

Key facts: Any fact critical to your analysis, argument, or decision requires verification. Check that source documents actually support claimed information.

Citations: Verify that cited sources exist and contain the claimed content. This proves especially important for any citations that will appear in filings or reports to clients.

Numbers and dates: AI systems often struggle with numerical precision. Any dates, dollar amounts, percentages, or quantitative claims warrant careful verification.

Logical inferences: Check that conclusions drawn from source materials are logical and supported. Even if individual facts are accurate, inferences might be flawed.

Legal standards: Verify that any statements about legal requirements, standards, or precedents accurately reflect controlling law.

Verification strategies:

Sampling: For large volumes of AI-generated output, verify a statistical sample. Use findings to assess overall reliability and identify areas needing closer review.

Risk-based verification: Verify high-stakes claims more thoroughly than routine matters. Allocate verification effort according to potential impact of errors.

Cross-checking: Compare AI outputs against other sources—traditional research, human analysis, or different AI systems. Consistency across sources increases confidence.

Expert review: Have subject matter experts review AI-generated content in their areas. A tax

© 2025 Merlin Search Technologies, Inc.

specialist reviewing AI-generated tax analysis can spot errors general reviewers might miss.

Adversarial review: Consider how opposing counsel might challenge AI-generated work product. Verify elements that could become points of contention.

Quantifying Risk: When to Trust AI Outputs

Understanding hallucination risks helps you make informed decisions about when to rely on AI outputs with less verification versus when to treat outputs as starting points requiring extensive validation.

Factors Affecting Reliability

Specificity of sources: Outputs based on specific provided documents are more reliable than outputs based on training memory. If the system analyzed your case documents and cited specific sections, reliability is higher than if it generated responses from general knowledge.

Citation completeness: Outputs with comprehensive citations are more reliable than those with vague references or no citations. “According to Document ID 45892, page 7, paragraph 3” is more trustworthy than “based on the contract terms.”

Consistency across sources: Claims supported by multiple independent sources are more reliable than claims appearing in single sources, especially if those sources come from parties with aligned interests.

Domain complexity: In straightforward domains with clear factual content, AI outputs may be more reliable. In complex areas requiring nuanced interpretation or specialized knowledge, expect lower reliability without expert verification.

Output type: Summaries of provided documents are generally more reliable than analytical conclusions. Extracted facts are more reliable than legal interpretations. Direct quotes are more reliable than paraphrases.

Model quality: More recent, larger, better-fine-tuned models generally produce more reliable outputs. GPT-4o or Claude Sonnet 4.5 outputs are typically more reliable than outputs from earlier or smaller models.

Creating Reliability Frameworks

Organizations should develop frameworks for assessing output reliability and determining appropriate verification levels:

High-reliability outputs (minimal verification needed):

- Summaries of provided documents with comprehensive citations
- Extracted structured data from clear sources
- Categorization of provided documents
- Timeline generation from documents with date verification

© 2025 Merlin Search Technologies, Inc.

Medium-reliability outputs (moderate verification needed):

- Analytical conclusions from provided documents
- Comparative analyses across multiple sources
- Identification of themes and patterns
- Draft work product requiring professional review

Low-reliability outputs (extensive verification required):

- Anything based on training memory without source documents
- Complex legal analysis requiring specialized expertise
- Outputs intended for court filings or client deliverables
- Novel or unusual conclusions not obviously supported by sources

No-reliability outputs (treat as purely generative without evidentiary value):

- Legal research without verified citations
- Factual claims without source attribution
- Technical or specialized content outside verified sources
- Predictive or speculative content

Cost-Benefit Analysis

Verification requires time and resources. Decisions about verification intensity should weigh reliability needs against practical constraints:

For document review in discovery, you might tolerate higher error rates than for court filings. If you're identifying potentially relevant documents from millions of records, and 5% classification errors are acceptable given manual review follows, extensive verification of AI classifications might be unnecessary.

Conversely, for deposition summaries that attorneys will rely on for trial preparation, verification intensity should be high. Missed contradictions or inaccurate attributions could undermine case strategy.

The key is matching verification effort to stakes. Don't treat all AI outputs identically. Invest verification resources where errors matter most.

The Human-in-the-Loop Imperative

Throughout this chapter, we've emphasized verification, oversight, and human judgment. This reflects a fundamental principle: AI should augment human expertise, not replace it. The phrase "human-in-the-loop" captures this principle—AI generates outputs, but humans remain actively engaged in reviewing, verifying, and taking responsibility for work product.

Why Humans Remain Essential

Professional judgment: AI can analyze documents and identify patterns, but professional judgment determines strategy, assesses significance, and makes tactical decisions. These require experience, wisdom, and contextual understanding that AI systems lack.

© 2025 Merlin Search Technologies, Inc.

Contextual understanding: Humans understand stakes, relationships, politics, and nuances that AI systems cannot grasp. A document might be technically non-privileged but strategically sensitive. AI sees only technical analysis.

Error detection: Humans can recognize when AI outputs don't make sense, conflict with known facts, or seem anomalous. This common-sense check catches errors that might otherwise propagate through workflows.

Ethical responsibility: Professional obligations rest on humans, not tools. Lawyers and discovery professionals cannot delegate ethical responsibility to AI systems. They remain accountable for work product quality and ethical compliance.

Adaptation and learning: Humans learn from errors, adapt strategies when approaches fail, and apply creativity to novel situations. AI systems lack this adaptive capability within deployed applications.

Effective Human-AI Collaboration

The goal isn't minimizing human involvement—it's optimizing the division of labor between humans and AI:

AI handles volume: AI excels at processing large volumes quickly. Use it to analyze thousands of documents, generating initial summaries or classifications that humans would find prohibitively time-consuming.

Humans handle judgment: Reserve human attention for decisions requiring professional judgment, strategic thinking, or ethical considerations. Don't waste human capacity on tasks AI handles well.

AI generates drafts: Use AI to create initial versions of work product—summaries, outlines, analyses—that humans then refine, verify, and finalize.

Humans verify key elements: Focus human verification on critical facts, important conclusions, and high-stakes elements where errors would be most consequential.

Iterative refinement: Establish workflows where AI generates outputs, humans review and provide feedback, AI refines based on that feedback, and the cycle continues until quality standards are met.

Looking Ahead

This chapter examined the most serious challenge in legal AI adoption—the tendency of LLMs to generate plausible but fabricated content. We explored why hallucinations occur, what makes them dangerous, and how to structure workflows that minimize their impact while preserving efficiency benefits.

The strategies we discussed—RAG and REG architectures, mandatory citations, systematic verification, and human-in-the-loop workflows—represent current best practices. But they don't

© 2025 Merlin Search Technologies, Inc.

eliminate all risks. Responsible AI use in discovery requires ongoing vigilance, healthy skepticism toward AI outputs, and commitment to verification.

Understanding hallucinations prepares you to use AI effectively. But another concern often dominates initial discussions about AI in legal practice: security and confidentiality. Can you safely send client information to AI systems? Do you risk waiving privilege or breaching confidentiality obligations? These questions deserve thorough examination.

Chapter 4 addresses data security, privacy, and confidentiality comprehensively. We explore where your data goes when you use AI systems, what protections exist, what risks you should understand, and how to assess whether AI use aligns with your professional obligations to protect client information. Security concerns are legitimate, but as you'll see, they may be more manageable than many legal professionals initially assume.

© 2025 Merlin Search Technologies, Inc.

MERLIN

04

Data Security, Privacy, and Confidentiality

© 2025 Merlin Search Technologies, Inc.

CHAPTER FOUR

Data Security, Privacy, and Confidentiality

When ChatGPT burst onto the scene in late 2022, one question dominated discussions among legal professionals: "Can I send client data to this system without breaching confidentiality or waiving privilege?" The concern was understandable. Lawyers face strict ethical obligations to protect client information, and the prospect of transmitting sensitive documents to cloud-based AI systems raised legitimate questions about security, data retention, and inadvertent disclosure.

These concerns haven't disappeared, but our understanding has matured considerably. This chapter addresses data security and confidentiality comprehensively, examining how AI systems handle data, what protections exist, what risks remain, and how to assess whether AI use aligns with professional obligations.

The conclusion many discovery professionals have reached may surprise those still operating from initial fears: using commercial AI services through proper channels presents security and confidentiality risks comparable to other cloud services legal professionals routinely employ, such as email, document management systems, and ediscovery platforms. In some respects, AI systems may be more secure than traditional alternatives.

Understanding AI Data Flow

Before assessing security, we must understand what happens to data you send to AI systems. The answer depends entirely on which service you use and how you access it.

The Consumer vs. Commercial Divide

This distinction is fundamental and often misunderstood. Consumer AI services and commercial AI services handle data very differently, with profound implications for legal use.

Consumer Services:

Consumer versions of AI platforms—the free ChatGPT website, free Claude interface, or free Gemini access—typically reserve rights to use your inputs for various purposes. When OpenAI first released ChatGPT as a free public beta, its terms explicitly allowed using conversation data to improve the service and train future models.

For AI companies, these free offerings serve multiple purposes. They allow widespread testing and feedback collection. They generate publicity and user adoption. And critically, they provide data that helps improve models through the fine-tuning process we discussed in Chapter 1.

The trade-off seemed clear: users received free access to powerful technology, and in exchange, their inputs might be used to improve the service. For casual personal use, many found this acceptable. For professional legal work involving confidential client information, it

© 2025 Merlin Search Technologies, Inc.

obviously wasn't.

Commercial Services:

Recognizing that many potential customers couldn't accept consumer terms, AI providers developed commercial offerings with fundamentally different data handling policies. These commercial services, accessed through paid subscriptions or API agreements, include strict contractual provisions prohibiting use of customer data for training, improvement, or any purpose beyond processing the specific user request.

The distinction isn't merely one of cost. It's a fundamental difference in the relationship between user and provider. Consumer services involve data exchange—you provide data that helps improve the service. Commercial services are pure service provision—you pay money, and the provider processes your data according to strict limitations.

Where Your Data Goes

When you submit information to a commercial AI service, here's the actual data flow:

1. **Transmission:** Your prompt travels from your device to the AI provider's servers through encrypted connections (typically TLS/SSL encryption, the same used for banking and email). The data is encrypted in transit, making interception during transmission extremely difficult.

2. **Processing:** The AI provider's servers receive your encrypted data, decrypt it in their secure environment, and route it to an available LLM instance. This might involve load balancing across data centers, but the data remains within the provider's secure infrastructure.

3. **Context Window:** Your prompt is placed in the context window (the "whiteboard" from Chapter 2) along with any conversation history or additional context. The LLM reads this information, processes it through its neural network, and generates a response.

4. **Response Generation:** The generated response is written back to the context window. The interface layer retrieves this response to return to you.

5. **Temporary Storage:** Throughout this process, your data exists in temporary memory—the context window and associated processing buffers. This is volatile storage that's cleared once processing completes.

6. **Context Window Clearing:** After the response is returned, the context window is cleared. The LLM instance returns to its baseline state, retaining no information about your request.

7. **Return Transmission:** Your response travels back to your device through encrypted connections.

8. **Interface Storage:** If you're using an interface like ChatGPT or Claude's web application, your conversation is stored in the interface provider's database so you can return to it later. This storage is separate from the LLM itself and subject to the interface's data re-

© 2025 Merlin Search Technologies, Inc.

tention policies.

What Doesn't Happen

Equally important is understanding what doesn't occur with commercial AI services:

The LLM doesn't learn from your data: Your prompts don't modify the model's neural network parameters. The model processes your request and returns to exactly the same state afterward. This is the "brain in a jar" characteristic from Chapter 2.

Your data doesn't enter training: Commercial agreements explicitly prohibit using customer data for training future models. Your confidential documents won't appear in the next version of GPT or Claude.

Other users can't access your data: Each request is processed independently. User A's prompts never appear in User B's context window. The LLM has no mechanism to leak information between users because it doesn't retain information after processing.

Data doesn't persist indefinitely in processing systems: Once your request is completed and the response returned, the processing systems clear your data from temporary memory. It doesn't sit in databases waiting for potential breach.

Providers don't review your prompts (typically): Commercial agreements usually specify that providers won't access or review customer data except in limited circumstances like security monitoring, abuse prevention, or when required by law.

Commercial Agreements and Data Protection

The contractual terms governing commercial AI services provide the framework for data protection. Understanding these agreements helps assess whether using specific services aligns with professional obligations.

Standard Commercial Provisions

Major AI providers offer commercial terms that include several key provisions:

Data Non-Use Clauses:

These provisions explicitly state that the provider will not use customer data for training, improvement, or any purpose other than providing the service. Here's Microsoft's language for Azure OpenAI Service as a representative example:

> "Your prompts (inputs) and completions (outputs), your embeddings, and your training data:
>
> - are NOT available to other customers
> - are NOT available to OpenAI
> - are NOT used to improve OpenAI models
> - are NOT used to improve any Microsoft or 3rd party products or services

- are NOT used for automatically improving Azure OpenAI models for your use in your resource"

OpenAI's own commercial terms include similar provisions:

> "We do not use Content that you provide to or receive from our API ('API Content') to develop or improve our Services."

Anthropic, Google, and other providers include comparable language in their commercial agreements.

Data Retention Limits:

Commercial agreements typically specify how long customer data is retained and in what circumstances. Many specify:

- **Processing data:** Retained only as long as necessary to complete the request (typically seconds or minutes)
- **Logs:** May be retained for limited periods for security and operational purposes (30-90 days commonly)
- **Conversation history:** Retained in interface storage as long as user maintains their account, but deletable by user
- **Backups:** May exist in backup systems for limited periods but subject to same use restrictions

Zero Data Retention Options:

Some providers offer zero data retention configurations where even logs are minimized or eliminated. These configurations provide maximum security for highly sensitive applications, though they may limit troubleshooting capabilities if issues arise.

Encryption Requirements:

Commercial agreements specify encryption standards for data in transit and at rest. Industry-standard encryption (AES-256 for data at rest, TLS 1.2+ for transmission) is typical.

Compliance Certifications:

Enterprise AI providers typically maintain compliance with various security and privacy frameworks:

- SOC 2 Type II certification
- ISO 27001 certification
- GDPR compliance for European data
- HIPAA compliance capabilities for healthcare data
- Industry-specific certifications as applicable

Data Location and Sovereignty:

Some commercial agreements allow specifying where data is processed and stored geograph-

© 2025 Merlin Search Technologies, Inc.

ically. This matters for organizations with data residency requirements or concerns about specific jurisdictions.

Subprocessor Disclosure:

Commercial agreements typically disclose what subprocessors (vendors the AI provider uses) have access to customer data and under what terms. This transparency allows assessment of the full security chain.

API vs. Web Interface

How you access AI services affects data handling:

API Access:

API (Application Programming Interface) connections allow your software to communicate directly with AI services without human use of web interfaces. APIs typically offer:

- **Programmatic control:** Your code controls exactly what data is sent and how
- **No conversation storage:** Unless you explicitly implement it, APIs don't store conversation history
- **Enterprise security:** API access often includes additional security options like private endpoints, enhanced authentication, and audit logging
- **Clear contractual terms:** API agreements make data handling explicit and enforceable

For production discovery workflows processing sensitive documents, API access through commercial agreements represents best practice.

Web Interfaces:

Web-based interfaces like ChatGPT's website or Claude's chat interface offer convenience but additional considerations:

- **Conversation storage:** Interfaces store conversations in their databases for user convenience
- **Account security:** Security depends partly on account protection (strong passwords, multi-factor authentication)
- **User settings:** Some interfaces offer settings controlling data retention or opting out of certain uses
- **Visibility:** Interfaces make AI accessible to users who might not understand security implications

For individual professional use on less sensitive tasks, web interfaces with commercial accounts can be appropriate. For production systems processing confidential client data at scale, API access provides better control and security.

Reading the Fine Print

Not all commercial AI services are equal. When evaluating providers, examine:

© 2025 Merlin Search Technologies, Inc.

1. **What exactly is prohibited:** Does the agreement prohibit all uses of your data, or just training? Are there exceptions for security monitoring, legal compliance, or service improvement?

2. **What happens to conversation history:** If using an interface that stores conversations, where is that data kept, how long, and who can access it?

3. **Deletion capabilities:** Can you delete data on demand? What happens when you delete conversations or close your account?

4. **Audit and logging:** What logs are maintained, for how long, and who can access them?

5. **Breach notification:** Does the agreement require notification if your data is breached?

6. **Liability provisions:** What recourse do you have if the provider breaches data handling terms?

7. **Third-party access:** Under what circumstances might the provider share your data with third parties or respond to the legal process?

Can LLMs Share Confidential Information?

This question troubles many legal professionals: even if commercial agreements prohibit data misuse, could the LLM itself somehow leak confidential information to other users? Could information you provide inadvertently appear in someone else's responses?

The technical answer is no, for reasons grounded in how LLMs actually work.

Why LLMs Can't Learn from Prompts

Recall from Chapters 1 and 2 that LLMs are static systems. Their neural network parameters are fixed after training completes. When you submit a prompt:

1. The LLM processes your input through its existing parameters
2. Generates a response based on those fixed parameters
3. Returns to exactly the same state afterward

Your prompt doesn't modify these parameters. The LLM cannot learn from your interaction. It has no mechanism to incorporate your information into its knowledge base.

This is fundamentally different from systems that learn continuously. Some AI systems do learn from user interactions, updating their models based on feedback. LLMs as currently deployed in commercial services don't work this way. They're inference-only systems—they apply their fixed knowledge to new inputs without modifying that knowledge.

Why Other Users Can't Access Your Data

The architecture prevents cross-user data exposure:

© 2025 Merlin Search Technologies, Inc.

Isolated context windows: Each request gets its own context window. User A's context window never contains User B's data. The LLM processes requests independently.

No persistent memory: After processing your request, the LLM retains nothing. When the next user's request arrives (possibly to the same LLM instance), it starts from a clean state with only that user's context window visible.

Randomized routing: Requests are typically routed to available compute resources pseudo-randomly. Even if you submit multiple requests in sequence, they might go to different LLM instances. There's no persistent "session" at the LLM level.

No query mechanism: LLMs have no ability to search previous requests or access other users' data. They can only process what appears in the current context window.

Theoretical Vulnerabilities

While the architecture prevents inadvertent leakage, some theoretical vulnerabilities warrant mention:

Training data reconstruction: Academic research has shown that with extensive effort, attackers might extract small amounts of training data from LLMs by carefully probing them. However, this affects training data (public text from the internet), not user prompts, which aren't part of training in commercial services.

Prompt injection attacks: Sophisticated attackers might craft prompts that attempt to manipulate the system into revealing information about its operation or configuration. However, these attacks don't expose other users' data because that data isn't accessible to the LLM.

Interface vulnerabilities: Security issues in the interface layer (web application bugs, database vulnerabilities) could potentially expose conversation data stored by the interface. This is why interface security matters, though it's separate from LLM security per se.

Provider access: The AI provider's employees or systems might technically access your data if they breach their own policies or if their systems are compromised. This is a provider trustworthiness question rather than an LLM architecture question.

Comparison to Other Cloud Services

To contextualize the risk, consider that sending data to commercial LLM services presents security characteristics similar to other cloud services legal professionals routinely use:

Email services: When you email client information using Office 365, Gmail, or other cloud email, that data sits on provider servers indefinitely. Email systems maintain persistent storage of your communications. Staff at email providers could technically access your email if they violated policies. Yet legal



More secure than Office 365?
Yes, when you think about it.

professionals use cloud email extensively, relying on contractual protections and provider trustworthiness.

1. **Document management systems:** Cloud-based document management platforms store client documents on provider servers long-term. This presents a larger attack surface than LLM requests that exist in temporary memory for seconds.

2. **Ediscovery platforms:** Discovery professionals routinely upload millions of potentially confidential documents to third-party ediscovery platforms. These platforms process, store, and allow access to documents for months or years during litigation.

3. **Video conferencing:** Confidential client meetings occur over Zoom, Teams, or other platforms that route audio and video through provider infrastructure.

4. **Cloud storage:** Documents stored in Dropbox, Box, Google Drive, or SharePoint reside on provider servers with provider employees potentially able to access them.

The legal profession has collectively determined that these cloud services, when used with appropriate commercial agreements and security measures, provide reasonable expectation of privacy sufficient to avoid confidentiality breaches or privilege waivers. The analysis for commercial AI services is similar.

Reasonable Expectation of Privacy

The legal standard typically governing whether using third-party services waives privilege or breaches confidentiality is "reasonable expectation of privacy." This framework, developed through case law and bar association guidance, assesses whether the circumstances surrounding communication or data storage suggest the parties intended to maintain confidentiality.

Legal Framework

Courts have consistently held that using third-party technology services doesn't waive privilege if users have reasonable expectation of privacy in their communications. This principle has been applied to:

Email: Using employer-provided or cloud-based email for privileged communications doesn't waive privilege despite third-party hosting, as long as users reasonably expect communications to remain private.

Cloud storage: Storing privileged documents in cloud services doesn't waive privilege when done through services providing appropriate security and confidentiality commitments.

Litigation support vendors: Sharing privileged documents with ediscovery vendors, forensic experts, or litigation support services doesn't waive privilege because these disclosures are made in furtherance of legal representation with reasonable confidentiality expectations.

Factors Supporting Reasonable Expectation

Several factors support finding reasonable expectation of privacy when using commercial AI

© 2025 Merlin Search Technologies, Inc.

services:

Contractual protections: Commercial agreements explicitly prohibiting data disclosure or use support reasonable privacy expectations. These contracts create enforceable obligations protecting confidentiality.
Industry-standard security: Using services that implement industry-standard encryption, access controls, and security measures indicates reasonable care to protect confidentiality.

Limited disclosure: Data shared with AI providers is disclosed only to the extent necessary for the service, not broadly published or made available to unrelated parties.

Purpose of disclosure: Using AI to analyze documents serves the purpose of legal representation, similar to engaging other technology providers or expert consultants.

Vendor reliability: Reputable AI providers with established track records, security certifications, and significant business incentives to protect customer data support reasonable privacy expectations.

ABA and State Bar Guidance

The ABA and various state bar associations have addressed technology use and confidentiality obligations:

ABA Formal Opinion 477R (Revised 2017): Addressed email and electronic communications, concluding that lawyers can generally use unencrypted email because they have reasonable expectation of privacy. The opinion emphasized that lawyers should understand the technology they use and take reasonable security measures.

ABA Formal Opinion 483 (2018): Discussed cloud storage, concluding that lawyers may use cloud storage with reasonable care to ensure confidentiality. Factors include understanding how providers will handle data, evaluating security measures, and staying informed about developments.

ABA Formal Opinion 512 (2024): Specifically addressed generative AI, emphasizing lawyers' duties to understand tools they use and take reasonable measures to protect confidential information. The opinion didn't treat AI as categorically different from other cloud technologies regarding confidentiality.

State bar opinions: Various state bars have issued guidance generally consistent with ABA positions, emphasizing reasonable care, understanding of technology, and appropriate vendor selection rather than prohibiting cloud technology use.

Applying Standards to AI Services

Using commercial AI services through appropriate agreements satisfies reasonable expectation of privacy for several reasons:

More ephemeral than alternatives: Data sent to LLMs exists in temporary memory for seconds, unlike email or documents stored indefinitely on cloud servers. The brief existence actu-

© 2025 Merlin Search Technologies, Inc.

ally reduces exposure compared to persistent storage.

Strong contractual protections: Commercial AI agreements typically include specific prohibitions on data use, often stronger than generic cloud service terms.

No persistent storage in LLM: The LLM itself stores nothing. Only interface applications maintain conversation history, and this is typically deletable by users.

Security measures: Enterprise AI providers implement security measures comparable to or exceeding those of other legal technology providers.

Professional acceptance is growing: As legal professionals gain sophistication about AI security, using commercial AI services with appropriate safeguards is becoming accepted practice, much as cloud email and storage became accepted.

The key is using AI services properly—commercial agreements, understanding data handling, appropriate vendor selection, and reasonable security measures. Using consumer AI services without commercial protections for confidential client information would not satisfy reasonable expectations of privacy. Using commercial services with proper agreements likely does.

Special Considerations for Regulated Data

Certain types of data trigger additional regulatory requirements beyond general confidentiality obligations:

Healthcare Data (HIPAA)

Protected health information (PHI) under HIPAA requires specific safeguards. If your legal work involves healthcare data, ensure:

- AI vendors offer HIPAA-compliant configurations
- Business Associate Agreements (BAAs) are in place
- PHI is appropriately de-identified before processing, or
- Full HIPAA safeguards are implemented

Financial Data

Financial services data may be subject to regulations like GLBA (Gramm-Leach-Bliley Act) or industry-specific requirements. Financial institutions using AI services should verify compliance with applicable regulations.

International Data

Data from European individuals triggers GDPR requirements. Data from other jurisdictions may trigger additional data protection laws. Consider:

- Where data is processed and stored
- Data transfer mechanisms for cross-border data movement

© 2025 Merlin Search Technologies, Inc.

- Individual rights under applicable privacy laws
- Documentation and consent requirements

Government Classified Information

Classified government information requires specific security clearances and approved systems. Standard commercial AI services are not appropriate for classified data. Government agencies must use approved systems meeting classification-level requirements.

Trade Secrets

While trade secrets don't trigger specific regulations like healthcare data, their value demands careful protection. Consider whether the benefits of AI analysis outweigh risks of exposure, even through secure commercial services.

Looking Ahead

This chapter examined security, privacy, and confidentiality considerations when using AI services. We explored how AI systems handle data, what protections commercial agreements provide, why the technical architecture prevents inadvertent disclosure, and how using AI services compares to other cloud technologies legal professionals routinely employ.

The conclusion many discovery professionals have reached is that commercial AI services, used appropriately, present manageable security risks comparable to other cloud technologies. The key lies in understanding the technology, selecting reputable providers with strong contractual protections, implementing reasonable security measures, and training users about appropriate use.

Security and confidentiality concerns, while legitimate, need not prevent AI adoption. They should inform how you adopt AI—driving thoughtful vendor selection, appropriate safeguards, and reasonable security practices.

With concerns about data security addressed, we can now turn to the broader ethical framework governing AI use in legal practice. Chapter 5 examines the comprehensive ethical obligations lawyers face when using AI, synthesizing guidance from the ABA and state bar associations into practical frameworks for responsible AI deployment. Security is one component of ethical AI use, but the ethical considerations extend further, encompassing competence, supervision, transparency, billing, and professional judgment.

© 2025 Merlin Search Technologies, Inc.

MERLIN

05

Ethical and Professional Responsibilities

© 2025 Merlin Search Technologies, Inc.

MERLIN

CHAPTER FIVE

Ethical and Professional Responsibilities

The integration of Generative AI into legal practice presents ethical questions that extend far beyond security and accuracy. These questions touch the core of what it means to be a competent legal professional in an era of rapidly advancing technology. Can lawyers delegate legal analysis to AI systems? What must they disclose to clients about AI use? How should they bill for AI-assisted work? What supervisory obligations arise when staff use these tools?

These questions aren't hypothetical exercises. They affect daily practice decisions, client relationships, professional reputations, and potentially bar licenses. The legal profession has begun developing frameworks to address them, though the guidance continues to evolve as both the technology and our understanding of its implications mature.

This chapter synthesizes ethical guidance from the ABA and state bar associations into practical frameworks for responsible AI use. We examine core obligations, explore how they apply to specific AI applications in discovery, and provide decision-making tools for navigating ethical uncertainties. The goal is not to provide definitive answers to every situation—the technology evolves too rapidly for that. Instead, we aim to establish principles and analytical approaches that remain relevant as capabilities advance.

Why Ethics Matter Differently in Legal AI

The legal profession operates under distinctive constraints that shape how we must approach technology adoption. Lawyers serve as fiduciaries to their clients, owing duties of loyalty, confidentiality, and competent representation that exceed ordinary commercial relationships. These heightened obligations mean lawyers cannot simply adopt useful technology the way other professionals might. Each technology decision must be evaluated through the lens of client interests and professional responsibilities.

Legal work often occurs in adversarial contexts where opposing parties scrutinize work product for weaknesses. If you use AI for document review, opposing counsel might challenge your methodology. If AI assists in brief writing, courts might question accuracy and attribution. These challenges don't mean AI use is inappropriate, but they demand defensible implementation and clear documentation.

ABA Formal Opinion 512: Comprehensive Analysis

On July 29, 2024, the ABA Standing Committee on Ethics and Professional Responsibility issued Formal Opinion 512, providing the profession's first comprehensive ethical guidance specifically addressing Generative AI use. This landmark opinion doesn't create new ethical rules but interprets existing Model Rules of Professional Conduct as they apply to AI technology.

Understanding Opinion 512 is essential for any legal professional using or considering AI tools. The opinion addresses six core areas: competence, confidentiality, verification, transparency,

supervision, and billing. We examine each in detail.

1. Duty of Competence

The Rule: Model Rule 1.1 requires lawyers to provide competent representation, defined as "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation."

Application to AI: The ABA interprets this duty to require that lawyers using AI tools develop a reasonable understanding of how those tools work, what they can and cannot do reliably, and what risks they present. Notably, the opinion emphasizes that lawyers need not become AI experts or understand technical implementation details. What's required is functional understanding sufficient to use tools effectively and recognize their limitations.

Practical Implications:

Lawyers using AI should understand what the specific tool does, what types of outputs it generates, known limitations and failure modes, appropriate use cases, and how to verify outputs for accuracy. The competence duty extends to staying informed about technological developments through CLE programs, professional publications, consulting with technology experts when implementing new tools, and testing tools on representative tasks before production deployment.

Competence requirements vary by practice area. Discovery professionals need deeper understanding of AI document analysis capabilities than general practitioners. Competent lawyers recognize when tasks exceed their technological understanding and seek assistance—consulting technology specialists, working with vendors, or declining to use AI for applications beyond current comprehension.

2. Confidentiality and Data Protection

The Rule: Model Rule 1.6 requires lawyers to maintain client confidentiality, prohibiting disclosure of information related to representation without client consent except in limited circumstances.

Application to AI: Opinion 512 concludes that lawyers may use AI tools that process confidential information, but must take reasonable steps to protect that information from unauthorized disclosure.

Practical Implications:

Before sending confidential information to AI services, lawyers must evaluate data handling practices. This evaluation should address what the vendor does with customer data, whether data is used for training or improvement, how long data is retained, what security measures protect data, whether the vendor complies with relevant regulations, and what happens if the vendor experiences a breach.

Using consumer AI services for confidential information likely violates Rule 1.6 because these services may use data for training. Commercial services with appropriate agreements gener-

© 2025 Merlin Search Technologies, Inc.

ally don't raise the same concerns. Lawyers must implement reasonable security measures when using AI, which might include using services with strong encryption, implementing access controls, enabling multi-factor authentication, using private deployments for highly sensitive matters, and training staff on security practices.

What's "reasonable" depends on sensitivity of information, nature of the tool, and available security options. Highly sensitive matters (national security, trade secrets, high-stakes litigation) might warrant more stringent measures than routine matters.

3. Independent Verification and Due Diligence

The Rule: Model Rule 1.1's competence requirement implies that lawyers must ensure work product meets professional standards regardless of what tools assisted in creation.

Application to AI: Opinion 512 emphasizes that lawyers cannot blindly rely on AI outputs. They must independently verify AI-generated content before using it in representation. The opinion specifically references cases where lawyers were sanctioned for citing fake cases generated by ChatGPT, using these incidents to illustrate verification imperatives.

Practical Implications:

The degree of verification needed depends on how AI outputs are used and what risks errors present. Court filings require thorough verification—case citations must be checked, factual claims must be verified against sources, legal arguments must be evaluated for soundness. Client communications should be verified to the degree appropriate for the subject matter. Internal work product might require less immediate verification, but anything used for decision-making must be verified before reliance. Document review AI classifications require validation through sampling and quality control processes.

Verification should address factual accuracy, citations (verify that cited authorities actually exist and say what's claimed), logical soundness (evaluate whether reasoning is sound), completeness (consider whether AI missed important issues), and context appropriateness (assess whether outputs properly account for case-specific circumstances).

Verification methods include source checking for any citations or references, cross-referencing AI outputs against other sources, having subject matter experts review AI-generated content in their specialties, sampling large volumes of AI outputs to assess reliability, and considering how opposing counsel might challenge AI-assisted work. Document verification efforts, particularly for work that might be scrutinized, as notes provide evidence of due diligence.

4. Transparency and Client Communication

The Rule: Model Rules 1.4 (Communication) and 8.4 (Misconduct) create obligations to communicate with clients and to be truthful in professional dealings.

Application to AI: Opinion 512 suggests lawyers should consider disclosing AI use to clients, particularly when it materially affects representation or clients might reasonably expect to know. The opinion stops short of mandating disclosure in all circumstances, recognizing that routine tool use might not require disclosure any more than using word processors or email

© 2025 Merlin Search Technologies, Inc.

requires specific client notification.

Practical Implications:

Disclosure is advisable when using novel or experimental tools not yet standard in practice, when AI substantially influences analysis or strategy, when clients have voiced concerns about AI or technology, for highly sensitive matters, or when AI use creates risks clients might not anticipate. Some lawyers include AI use disclosure in engagement letters or matter-specific communications.

Communication should explain in accessible terms what AI tools do and how they'll be used, describe benefits like efficiency or cost savings, acknowledge limitations and safeguards that address them, explain how confidential information will be protected, and clarify that humans remain responsible and will verify outputs.

Disclosure may not be necessary when AI use is comparable to other standard tools (like legal research databases) or plays a minor role in work product. As AI use becomes standard practice, disclosure requirements likely diminish.

Regarding disclosure to courts and opposing counsel, there's no general disclosure duty to announce AI use. However, the duty of candor to tribunals requires that statements to courts be accurate regardless of what tools produced them. In discovery, if opposing counsel asks about methodologies or tools used, truthful responses are required. Some courts have adopted local rules requiring disclosure of AI use in filings, and lawyers must comply with applicable rules.

5. Supervisory Responsibilities

The Rule: Model Rules 5.1 (supervision of lawyers) and 5.3 (supervision of nonlawyer assistants) require lawyers with managerial authority to ensure that subordinates' conduct complies with professional obligations.

Application to AI: Opinion 512 extends supervisory obligations to AI tool use. Partners and legal managers must ensure that lawyers and staff using AI do so competently, ethically, and consistently with the firm's obligations to clients.

Practical Implications:

Law firms and legal departments should develop clear policies governing AI use. These policies might address approved tools and what AI services may be used under what circumstances, prohibited uses (e.g., consumer AI services for confidential matters), vendor approval processes, data handling requirements, verification standards, documentation requirements, and escalation procedures.

Supervisory lawyers must ensure subordinates receive adequate training on ethical obligations, competent use of specific tools, security practices for protecting confidential information, verification techniques, and how to report concerns or errors.

Supervision isn't one-time training but ongoing oversight through quality review (sampling work product to ensure AI is used appropriately), compliance monitoring (checking that policies are followed), staying informed about changes to tools in use, addressing issues when

© 2025 Merlin Search Technologies, Inc.

they arise, and refining policies based on experience.

6. Billing Practices

The Rule: Model Rules 1.5 (Fees) and 8.4 (Misconduct) prohibit charging unreasonable fees or billing fraudulently.

Application to AI: Opinion 512 addresses how AI use affects billing, particularly for hourly billing arrangements. The opinion concludes that lawyers cannot bill clients for time spent learning to use AI tools, and that efficiency gains from AI should generally benefit clients through reduced fees rather than increased profits through maintained billing at pre-AI rates for faster work.

Practical Implications:

Lawyers cannot bill for time spent learning to use AI tools (analogous to not billing clients for learning law), troubleshooting AI tool problems that didn't arise from case-specific circumstances, or experimenting with new AI approaches without authorization.

If AI allows completing tasks faster, billing should generally reflect actual time invested, not pre-AI time estimates. For example, if document review that previously took 40 hours now takes 10 hours with AI assistance, billing 40 hours would be inappropriate. For alternative fee arrangements (fixed fees, success fees), AI efficiency gains might benefit the firm rather than requiring fee reductions, depending on the agreement.

Consider addressing AI use in fee agreements, particularly for hourly billing. This provides transparency about how fees are calculated. Clients expecting traditional hourly billing may need education about how AI affects time investment and billing.

Additional Ethical Considerations

Bias and Fairness

LLMs learn from training data that reflects biases present in society and historical texts. This creates risks of perpetuating or amplifying bias in legal work. Training data may overrepresent certain demographics or perspectives, legal texts reflect historical practices that may have been discriminatory, and language patterns can encode subtle biases about race, gender, socioeconomic status, and other characteristics.

In legal AI applications, bias might manifest as biased document classification that systematically miscategorizes documents related to certain groups, predictive analysis that reflects historical biases in judicial decisions, AI-drafted content that inadvertently uses biased language or framing, or pattern recognition that flags behavior more common in certain demographic groups.

Lawyers using AI should understand potential bias risks in their specific applications. Before deploying AI at scale, test for biased outcomes across demographic categories. Ongoing monitoring can detect bias that wasn't apparent in initial testing. When bias is detected, take steps to mitigate it by adjusting prompts, using different tools, or increasing human oversight. Document bias analysis and mitigation efforts, particularly for applications in sensitive contexts.

© 2025 Merlin Search Technologies, Inc.

Criminal justice applications deserve special scrutiny. AI used in matters affecting liberty, employment discrimination cases, housing discrimination, or other civil rights contexts requires heightened attention to bias risks.

Professional Judgment and Independence

Model Rule 2.1 requires lawyers to exercise independent professional judgment and render candid advice. AI tools, no matter how sophisticated, cannot replace this judgment.

Treat AI outputs as information to inform judgment, not as final answers. Approach AI recommendations with the same critical thinking applied to any other source. Consider case-specific factors AI might not fully account for. Ensure decisions prioritize client interests, not merely what AI suggests is efficient or likely.

Don't delegate decision-making to AI even when its suggestions seem reasonable. Vendor marketing shouldn't drive adoption of AI tools not actually beneficial for clients. Pressure to use AI for efficiency can't override professional judgment about when AI is inappropriate.

Certain decisions particularly require human judgment: settlement recommendations, litigation strategy choices, risk assessments for clients, advice on transaction structures, and ethical decisions within representation. AI can inform these decisions but cannot make them. The lawyer's judgment, informed by experience, knowledge of the client, understanding of context, and professional responsibility, must remain central.

State Bar Guidance

While Opinion 512 provides the ABA's comprehensive framework, individual state bars govern attorney conduct within their jurisdictions. Several states have issued their own guidance, including California, Florida, New York, and the District of Columbia. The guidance from these states largely aligns with Opinion 512, emphasizing that existing ethical rules apply to AI use and require competence, confidentiality protection, verification, appropriate supervision, and reasonable billing.

States sometimes emphasize issues particularly relevant to their jurisdictions—privacy laws, specific court rules, or local practice norms—but the core principles remain consistent: AI is permitted but must be used competently, ethically, and with appropriate safeguards. Practitioners should check their own jurisdiction's guidance for any specific requirements or interpretations that apply to their practice.

Looking Ahead

This chapter examined the ethical framework governing AI use in legal practice, synthesizing guidance from the ABA and state bar associations into practical decision-making tools. The core principles—competence, confidentiality, verification, transparency, supervision, and reasonable billing—provide foundation for responsible AI adoption. These principles prepare us for Part Two, where we explore advanced AI capabilities including agentic systems that can plan multi-step workflows, use tools strategically, and iterate toward goals autonomously.

© 2025 Merlin Search Technologies, Inc.

MERLIN

PART TWO

AGENTIC AI AND ADVANCED ARCHITECTURES

© 2025 Merlin Search Technologies, Inc.

PART TWO

Agentic AI and Advanced Architectures

Part One equipped you with foundational knowledge about Large Language Models—what they are, how they work, their limitations, and the ethical framework governing their use. Now we turn to two major evolutions in legal AI that move beyond simple prompt-response interactions: agentic systems that can plan and execute complex multi-step workflows autonomously, and reasoning models that can think through difficult problems methodically before responding.

These advances represent a fundamental shift in how AI assists legal work. Agentic systems don't just respond to questions—they pursue goals, breaking assignments into steps, executing searches, analyzing results, identifying gaps, and iterating until they produce comprehensive work products. Reasoning models don't just generate immediate responses—they engage in extended deliberation, working through logical chains, considering multiple perspectives, and evaluating their own conclusions before answering.

For discovery professionals, these capabilities address critical needs. Real discovery investigations don't follow linear paths—you search for information about Topic A, which reveals Custodian B is important, leading you to their documents, which mention Event C you hadn't known about. Human investigators naturally follow these threads. Simple LLM queries can't, but agentic systems can. Similarly, privilege determinations and credibility evaluations require working through multiple considerations systematically. Quick responses often miss important nuances. Reasoning models provide the deliberate analysis these tasks require.

What You'll Learn

Part Two explores how agentic systems use goal-directed behavior, multi-step planning, tool use, and iterative refinement to conduct investigations rather than just answer questions. You'll understand how reasoning models employ extended chain-of-thought deliberation to work through complex legal analysis, contradictory evidence, and strategic decisions. We examine when to deploy each capability, how to maintain quality control and oversight, and what ethical obligations intensify as AI systems become more autonomous and sophisticated.

Why These Capabilities Matter

Discovery work inherently involves both complex multi-step processes and difficult analytical problems. Many discovery questions can't be answered by reading single documents—they require synthesizing information across dozens or hundreds of sources, identifying patterns, recognizing contradictions, and building coherent narratives from fragments. Agents can orchestrate this synthesis in ways that simple prompts cannot. And as investigations progress, understanding deepens and priorities shift. Agentic systems can adapt to evolving contexts, refining their approaches based on what they've learned, while reasoning models can reconsider earlier conclusions in light of new information.

© 2025 Merlin Search Technologies, Inc.

These advanced capabilities promise dramatic transformation. Tasks requiring days of human analysis might be accomplished in hours. Investigations demanding extensive manual synthesis could be conducted systematically at scale. But these capabilities also create challenges. Autonomous systems operating with less granular human oversight require different quality assurance approaches. Verifying that an agent properly conducted a multi-step investigation or that a reasoning model correctly evaluated a complex legal question proves more complex than verifying a simple document summary. The ethical questions about professional judgment and human oversight become more acute when systems operate autonomously rather than simply responding to prompts.

How to Approach Part Two

Agentic AI and reasoning models represent the cutting edge of current legal technology. Not every discovery professional needs to implement these systems immediately. But understanding these capabilities prepares you to evaluate vendor claims, plan technology roadmaps, and stay current as the technology evolves. The trajectory of legal AI development points clearly toward both increased autonomy and more sophisticated reasoning.

As you read Part Two, continue thinking about your practice and workflows. Where do you encounter complex multi-step tasks that could benefit from autonomous AI assistance? What analytical problems require careful reasoning through multiple considerations? What investigations require following threads through multiple sources? These are the applications where advanced AI capabilities shine—and where Part Two's insights will prove most valuable.

Moving Forward

Part Two explores territory that many legal professionals haven't yet encountered directly. These advanced AI capabilities remain newer and less widely deployed than simpler LLM applications. But they represent where legal AI is moving, and understanding them positions you to lead rather than follow as capabilities mature and become more accessible. The technical foundation established here prepares you for the practical applications demonstrated in Part Three, where we show these advanced systems transforming actual discovery workflows.

Let's begin by examining what makes AI systems "agentic" and how autonomous assistance differs from simple prompt-response interactions.

© 2025 Merlin Search Technologies, Inc.

MERLIN

06

Understanding Agentic AI Systems

© 2025 Merlin Search Technologies, Inc.

MERLIN

CHAPTER SIX
Understanding Agentic AI Systems

In Chapter 1, we described Large Language Models as powerful prediction engines that generate text by anticipating what word should come next in a sequence. In Chapters 2 through 5, we explored how these systems work, their limitations, and the frameworks for using them responsibly. Throughout Part One, we examined LLMs as responsive systems—you provide a prompt, the LLM generates a response, and the interaction concludes.

Agentic AI systems operate fundamentally differently. Rather than simply responding to prompts, they pursue goals through autonomous planning and execution. Instead of a single request-response cycle, agentic systems orchestrate dozens or hundreds of LLM calls, each building on previous results, adjusting strategies when approaches fail, and iterating toward objectives without constant human direction.

This chapter explores what makes AI systems "agentic," examining the components that enable autonomous behavior, the architectural patterns that organize these components, and the practical implications for discovery professionals considering these advanced capabilities. Chapter 7 then examines reasoning models—a complementary advance that enables extended deliberation on complex analytical problems. Together, these capabilities represent the next evolution in AI-assisted discovery work.

Defining Agentic AI

The term "agentic" derives from "agency"—the capacity to act independently and make autonomous decisions toward goals. Agentic systems differ from traditional LLM applications in several key ways.

Goal-oriented operation and planning: Traditional LLM interactions are task-oriented. You specify a task ("Summarize this document"), and the system completes it. Agentic systems are goal-oriented. You specify an objective ("Identify all relevant evidence"), and the system determines what tasks are needed to achieve that objective.

This distinction matters because goal achievement requires planning and adaptation, not just following instructions. When you assign an agent a goal, it first analyzes what's needed, breaks the objective into manageable subtasks, and sequences those tasks logically.

Tool use capability: While LLMs generate text, agentic systems can invoke external tools—search functions, data extraction utilities, calculation capabilities, file operations, or specialized analysis tools. Critically, agents decide which tools to use and when, based on task requirements.

Self-improvement through iteration: Agentic systems evaluate their own outputs and refine them through multiple passes. After completing initial analysis, an agent might identify gaps in its work, recognize that conclusions lack sufficient support, or determine that alternative

© 2025 Merlin Search Technologies, Inc.

approaches might yield better results. The agent can then conduct additional searches, re-analyze information, or approach problems differently.

Related to this iterative capability is error detection. When agents make mistakes—searching with terms that return irrelevant results or drawing incorrect inferences—they can often detect these errors and take corrective action.

State management: While individual LLMs are stateless (Chapter 2), agentic frameworks maintain state across multiple operations. The agent "remembers" what it's already done, what it learned from previous steps, and how current work connects to the broader goal. This allows agents to conduct coherent investigations despite being built on stateless LLMs.

What Agentic Systems Are Not

Understanding agentic AI also requires clarity about what it isn't:

Not autonomous AGI: Agentic AI systems are not artificial general intelligence (AGI) capable of human-like reasoning across all domains. They're specialized tools designed for specific types of tasks, operating within defined boundaries.

Not learning systems: Current agentic systems don't learn from experience in the way humans do. Each investigation starts fresh. The agent doesn't remember previous matters or accumulate expertise over time. (Some architectures maintain learnings within single sessions, but this differs from persistent learning across sessions.)

Not independent decision-makers: Despite autonomy within workflows, agents operate under human direction. Humans set goals, approve deployment, and remain responsible for outcomes. Agents are sophisticated tools, not independent practitioners.

Not replacements for human judgment: Agents can execute complex investigations, but professional judgment about strategy, significance, and next steps remains human responsibility. We explore this extensively in Chapter 9's ethical discussion.

Not infallible: Agentic systems make mistakes. They can pursue unproductive directions, miss important information, or draw incorrect conclusions. Autonomy doesn't equal accuracy. All the limitations from Part One—hallucinations, context window constraints, training cutoffs—still apply.

Why These Capabilities Matter for Discovery

Understanding what agentic systems are—and are not—prepares us to examine why these capabilities matter for discovery work. Real investigations don't follow linear paths. You search for Topic A, which reveals Custodian B, whose documents mention Event C. Each discovery shapes the next inquiry. Simple LLM interactions can't pursue these emergent leads autonomously. Agentic systems can.

Many discovery questions require synthesizing information scattered across dozens of documents. "What did the company know about the defect, and when?" can't be answered by any single email, report, or memorandum. Each source provides one piece; the answer emerges

© 2025 Merlin Search Technologies, Inc.

through systematic analysis across the entire set. Agentic systems orchestrate this process autonomously—following threads, identifying patterns, recognizing contradictions, and building coherent understanding.

Consider the goal “Identify all documents related to the patent dispute.” A simple LLM query might search for “patent dispute” and return results. An agentic system might plan:

1. Search for documents using relevant patent terms
2. Analyze results to identify specific patents, parties, and time periods
3. Conduct targeted searches for the specific patents and custodians identified
4. Build a chronology of key events and identify information gaps
5. Conduct additional searches to address those gaps
6. Synthesize findings into a comprehensive report

The agent creates this plan autonomously, adjusting as execution reveals new information or approaches prove ineffective.

These systems can process massive document volumes while maintaining sophisticated analysis and incorporating quality control. Rather than producing search results for human review, they conduct investigations and deliver analyzed intelligence. This doesn’t replace professional judgment about strategy and significance, but it expands what’s possible in terms of comprehensiveness, speed, and consistency.

Components of Agentic Systems

Agentic AI systems comprise several technical components working together. Understanding these components at a high level helps you evaluate agentic tools and recognize their capabilities and limitations. An orchestration layer coordinates planning, execution, memory, and evaluation components to enable goal-directed behavior.

Planning: The planning component receives goals and creates action plans by prompting an LLM to analyze the goal and generate steps. The system considers available tools, breaks the goal into manageable tasks, and sequences those tasks logically.

Different systems use different planning approaches. Some generate complete plans upfront and execute all steps in sequence—effective for well-defined problems. Others create high-level plans initially and fill in details as work proceeds, adapting to discoveries. Still others generate each next step based on current circumstances without extensive forward planning. For discovery applications, approaches that combine strategic planning with flexibility to adapt work best.

Execution and decision-making: The execution component carries out plan steps by invoking tools or prompting LLMs to perform specific tasks. When a step requires using a tool—searching documents, extracting data, or performing calculations—the system identifies the appropriate tool, formats the input, invokes the tool, and processes the output.

During execution, agents must make numerous tactical decisions: Which tool to use? What search terms are likely effective? Do results answer the current question? Should the plan be modified based on new information? The system makes these decisions by prompting LLMs with the decision context and criteria, leveraging their reasoning capabilities to select appro-

© 2025 Merlin Search Technologies, Inc.

priate next steps.

After each action, the execution component updates the system's state—marking completed steps, recording outputs, updating understanding based on new information, and determining what comes next. This state management allows the agent to maintain coherent progress despite being built on stateless LLMs.

Memory: Memory components maintain information across the multiple LLM calls that comprise an agentic workflow. This working memory holds the goal, the current plan, what's been accomplished, results from previous steps, and items requiring further investigation.

Since context windows are limited (Chapter 2), memory systems must manage what information gets included in prompts. They summarize extensive previous work to fit available space, prioritize the most relevant information for the current task, and retrieve specific details when needed. This allows agentic systems to conduct extensive investigations despite context window constraints.

Evaluation and verification: As we emphasized earlier, verification is critical for any AI outputs. Agentic systems can incorporate verification into their workflows by prompting LLMs to critically assess work product. Agents can evaluate whether search results actually relate to search terms, whether conclusions follow from evidence, whether there are gaps or inconsistencies, and whether outputs address the original goal.

More sophisticated verification involves cross-checking across sources—confirming that multiple independent sources support conclusions, identifying contradictions, verifying that citations reference actual documents, and ensuring extracted data remains consistent across mentions. These verification mechanisms help identify outputs requiring additional human review.

These components work together under an orchestration layer that manages overall workflow—initiating planning, sequencing execution, handling errors, determining when goals are achieved, and compiling final outputs. When problems arise, the orchestration layer determines appropriate responses: retry the operation, try an alternative approach, or escalate to human oversight. This coordination enables agentic systems to pursue complex investigations while adapting to challenges and maintaining progress toward goals.

Agentic Architectures

These components can be organized in various architectural patterns, each with different strengths and appropriate use cases.

Single-Agent Architecture

The simplest agentic architecture involves one agent with all components integrated. A single agent receives goals, plans approaches, executes steps, maintains state, and produces outputs. This design offers important advantages: it's simpler to build and maintain, maintains consistency through a single perspective, operates efficiently without coordination overhead, and provides transparency since you're tracking one agent's reasoning rather than multiple interacting agents.

© 2025 Merlin Search Technologies, Inc.

However, single agents face inherent limitations. They can't be simultaneously optimized for diverse tasks, everything runs sequentially rather than in parallel, and very complex goals may overwhelm single-agent planning capabilities. For discovery applications, single-agent systems work well for document analysis within coherent scope, timeline construction from clear sources, targeted investigations with defined boundaries, and tasks requiring consistent analytical perspective. A single agent investigating communications between parties can plan search strategies, execute searches, analyze results for relevance, identify key items for detailed review, extract critical information, synthesize findings, identify gaps, conduct supplemental searches, and compile a comprehensive report—all while maintaining context and coherence throughout.

Multi-Agent Architecture

Multi-agent systems employ specialized agents, each optimized for particular tasks, working together on complex goals. Rather than one agent handling everything, multiple agents collaborate—each focusing on what it does best. For discovery applications, this specialization proves valuable. A search agent might excel at finding relevant documents across large collections, while an analysis agent focuses on extracting key information. A synthesis agent then combines findings into coherent narratives, while a quality assurance agent verifies accuracy and completeness.

The key benefit is that each agent can be optimized for its specific role, leading to higher quality results than a single generalist agent could achieve. However, coordinating multiple agents requires sophisticated orchestration to ensure coherent output and manage workflow complexity.

Hierarchical Agent Architecture

Hierarchical systems organize agents in tiers, with high-level strategic agents directing lower-level tactical agents. Strategic agents determine overall approach and break goals into major sub-goals. Tactical agents plan specific approaches to achieve those sub-goals. Execution agents perform concrete tasks like searching documents or extracting data.

This hierarchical approach scales well to very complex investigations—such as enterprise-scale discovery spanning multiple matters, or complex investigations requiring both strategic oversight and detailed execution across diverse document types and time periods. The structure allows for parallel processing while maintaining strategic coherence.

Control Mechanisms and Guardrails

Autonomous systems require controls ensuring they pursue goals appropriately without running indefinitely or producing unacceptable outputs. These guardrails protect against runaway execution while maintaining the benefits of autonomous operation.

Key control mechanisms include iteration limits to prevent endless refinement, scope constraints to keep investigations focused on relevant materials, and quality thresholds ensuring outputs meet professional standards. Human checkpoints provide strategic oversight at key milestones, while escalation triggers automatically involve human oversight when agents encounter uncertainty or contradictions.

© 2025 Merlin Search Technologies, Inc.

These controls ensure that agentic systems remain powerful tools under human direction rather than independent decision-makers. The goal is to harness autonomous capabilities while maintaining professional oversight and responsibility for outcomes.

Looking Ahead

This chapter introduced the core concepts of agentic AI—what makes systems "agentic," what components enable autonomous behavior, how those components can be architecturally organized, and what control mechanisms ensure appropriate operation.

Understanding these foundations prepares you for Chapter 7, where we examine reasoning models and their role in complex analytical problems. While agentic systems excel at multi-step workflows and autonomous investigation, reasoning models provide extended deliberation on difficult questions requiring systematic evaluation. Following that technical foundation, Part Three demonstrates both capabilities through detailed discovery applications—showing how the architectures and components we've discussed here transform actual legal investigations.

Chapter 8 will demonstrate that agentic capabilities aren't merely theoretical possibilities—they're practical tools already transforming how discovery professionals approach complex investigations. The architectures and components we've discussed here provide the foundation for workflows that dramatically expand what's possible in document analysis, investigation, and synthesis at scale.

© 2025 Merlin Search Technologies, Inc.

MERLIN

07

Reasoning Models and Deep Analysis

© 2025 Merlin Search Technologies, Inc.

CHAPTER SEVEN

Reasoning Models and Deep Analysis

In September 2024, a new category of AI model emerged that changed how we think about machine intelligence. OpenAI released the beta of its o1 model ("strawberry"), followed by similar capabilities from other providers, introducing what we now call reasoning models. These systems represent a fundamental shift in how AI approaches complex problems.

Unlike standard large language models that generate responses immediately, reasoning models take time to think through problems step by step. They work through complex logical chains, consider multiple perspectives, and evaluate their own conclusions before responding. For discovery professionals, this capability opens new possibilities for analyzing intricate legal questions, evaluating conflicting evidence, and conducting thorough investigations.

What Are Reasoning Models?

Standard LLMs produce responses in a single forward pass. You ask a question, and the model generates an answer token by token, completing the response in seconds. This approach works well for many tasks, but it has limitations when dealing with complex analytical problems that benefit from deliberate, step-by-step reasoning.

Reasoning models use a different approach. When you submit a query, these models engage in extended internal thinking before responding. They break problems into components, explore different analytical paths, identify potential errors in their reasoning, and refine their approach. This process, often called chain-of-thought reasoning, mirrors how human experts work through difficult problems.

The key distinction is deliberation. Standard models are fast but don't reconsider their initial analytical path. Reasoning models take more time but produce more thorough, logically consistent analysis. For a simple question like summarizing a single document, this extra processing offers little benefit. For complex tasks like analyzing contradictory testimony across dozens of depositions or evaluating multi-factor privilege determinations, the improvement can be substantial.

The Thinking Process

When you use a reasoning model, the system goes through several phases:

Problem decomposition: The model breaks your question into logical components. For a discovery task like "identify all documents that support the plaintiff's theory of conspiracy," it might consider what evidence would demonstrate conspiratorial conduct, what time period matters, which custodians would be involved, and what communication patterns would be relevant.

Analytical exploration: The model considers multiple approaches to the problem. It might

evaluate different legal frameworks, consider various interpretations of ambiguous facts, or explore alternative explanations for patterns in the evidence.

Self-correction: As it works through the problem, the model identifies potential errors or gaps in its reasoning. It might recognize that an initial conclusion conflicts with other evidence, or that it has overlooked an important consideration.

Synthesis: The model integrates its various analytical threads into a coherent response, weighing different factors and qualifying conclusions where appropriate.

This internal process can take from a few seconds to several minutes, depending on the complexity of the task. Some systems show you a summary of this thinking process, while others keep it hidden. Either way, the result is typically more thorough and logically consistent than what standard models produce.

When Reasoning Helps in Discovery

Not every discovery task benefits from extended reasoning. Simple document summaries, straightforward searches, and routine categorization work fine with standard models. Reasoning models excel at tasks that require:

Multi-step logical analysis: Problems that require working through several layers of reasoning benefit from the model's ability to maintain logical consistency across multiple steps. Evaluating whether a document is privileged, for instance, might require analyzing the nature of the communication, the relationship between parties, the subject matter, and whether any exceptions apply.

Contradictory evidence: When testimony conflicts or documents present inconsistent information, reasoning models can methodically work through the contradictions, identify which sources are most reliable, and explain the discrepancies.

Complex pattern recognition: Identifying subtle patterns across large document sets, such as unusual communication sequences that might indicate coordination or concealment, requires the kind of deliberate analysis that reasoning models provide.

Legal and factual synthesis: Tasks requiring integration of legal standards with complex factual scenarios benefit from reasoning models' ability to work through multiple considerations systematically.

Practical Applications in Investigation and Discovery

Let's examine how reasoning models enhance several common discovery tasks:

Privilege Analysis

Privilege review represents one of the most complex analytical tasks in discovery. A document might be privileged, not privileged, or privileged with exceptions. The analysis requires considering multiple factors: the nature of the communication, the parties involved, the subject matter, the purpose of the communication, and any applicable exceptions like the crime-fraud

exception or common interest doctrine.

A reasoning model can work through this analysis methodically. Given an email chain with multiple participants, some in-house counsel and some business personnel, discussing both legal strategy and business operations, a reasoning model might:

- Identify each communication segment and determine who sent it to whom
- Analyze the subject matter of each segment separately
- Determine whether counsel was providing legal advice or business guidance
- Consider whether business participants on the email waive privilege
- Evaluate whether any exceptions apply
- Reach a reasoned conclusion with appropriate qualifications

This kind of multi-factor analysis is exactly what reasoning models do well. They can maintain logical consistency across the various considerations and avoid the kind of oversimplification that leads to errors in privilege review.

Analyzing Contradictory Testimony

Consider a case with testimony from fifteen witnesses about a series of meetings that occurred over six months. Some witnesses remember certain details clearly while others are vague. Testimony conflicts about who said what, when decisions were made, and what information various participants had at different times.

A reasoning model can systematically work through these contradictions. It might start by creating a timeline of undisputed facts, then layer in each witness's testimony about disputed events, identify where accounts diverge, consider the reliability of different witnesses based on their role and proximity to events, and synthesize a coherent narrative that accounts for the inconsistencies.

The model's ability to hold multiple competing accounts in working memory, evaluate them against independent evidence, and reason about credibility makes it particularly valuable for this kind of analysis. Rather than simply noting contradictions, it can work through possible explanations and evaluate which accounts are best supported by the full record.

Timeline Construction with Logical Inference

Building comprehensive chronologies from document sets often requires logical inference. A document might reference "the meeting last Tuesday" without specifying which Tuesday. Another might mention "the proposal we discussed" without naming the proposal. Constructing an accurate timeline requires piecing together these references through careful reasoning.

Reasoning models can work through these inference chains. They can identify references to the same event across different documents, use context clues to date undated materials, recognize when documents are responding to earlier communications, and flag timeline gaps that suggest missing documents. This kind of detective work, which requires holding multiple pieces of information in mind while working through logical connections, plays to reasoning models' strengths.

© 2025 Merlin Search Technologies, Inc.

Document Authenticity Assessment

Sometimes you need to evaluate whether a document is what it purports to be. An email's metadata might not match its apparent content. Document properties might be inconsistent with the stated creation date. References in the document might conflict with other evidence about when it was created.

A reasoning model can systematically evaluate authenticity concerns. It might examine metadata, compare document formatting and language with other documents from the same alleged time period, check whether referenced events had actually occurred by the document's stated date, identify anachronisms, and evaluate the overall pattern of evidence about when and how the document was created. This kind of forensic analysis requires exactly the sort of methodical, multi-factor reasoning that these models provide.

Evaluating Opposing Arguments

Reviewing opposing counsel's briefs, productions, or privilege logs often requires identifying logical gaps, unsupported assertions, or inconsistencies. A reasoning model can work through an opponent's argument step by step, test each premise against the evidence, identify logical leaps, spot selective use of facts, and craft responses that address the argument's actual weaknesses rather than straw men.

This application is particularly valuable when dealing with sophisticated opposition. The model's ability to engage with complex legal reasoning, maintain logical rigor, and identify subtle flaws in argumentation helps you respond more effectively than you might with quick-glance review.

Investigation Planning

At the start of an investigation, you often face choices about where to focus limited resources. Which custodians are likely to have relevant information? What time periods matter most? What search terms will be most productive? What patterns should you look for?

A reasoning model can help think through investigation strategy. Given the initial allegations and available information, it can reason about who would likely be involved, what kinds of evidence would be most probative, where you might find that evidence, and what sequence of investigative steps would be most efficient. This kind of strategic reasoning, which requires understanding the legal issues, anticipating what evidence might exist, and planning a logical investigation sequence, benefits from the model's ability to work through multiple considerations systematically.

Advanced Reasoning: Multi-Factor Relevance Analysis

Consider a trade secrets case where the plaintiff alleges the defendant hired away key employees and used them to develop competing products. Relevance depends on several factors: when the employees were hired, what knowledge they had, what they worked on after joining the defendant, whether they accessed old files, what the competitive products actually contain, and the timing of product development.

© 2025 Merlin Search Technologies, Inc.

A standard LLM given a document about hiring practices might flag it as relevant because it mentions employee transfers. But a reasoning model can work through the full analysis. It might determine that while the document discusses employee transfers, it relates to a different department that had nothing to do with the technology at issue, involves transfers that occurred after the relevant time period, and therefore is not actually relevant despite superficial similarity to the case themes.

This kind of contextual reasoning, which requires understanding the legal theory, knowing what facts would actually matter to that theory, and distinguishing between superficially similar situations, represents exactly where reasoning models add value. They can avoid false positives that result from simplistic pattern matching and instead evaluate relevance based on actual legal and factual significance.

Cost and Speed Considerations

Reasoning models cost more and run slower than standard models. A task that takes a standard model five seconds might take a reasoning model thirty seconds or even minutes. Cost per query can be five to ten times higher. These tradeoffs mean you need to be strategic about when to use reasoning capabilities.

Use reasoning models for high-value analytical tasks where accuracy and thoroughness justify the extra cost and time. Privilege determinations on sensitive documents, analysis of key testimony, evaluation of critical evidence, and strategic planning benefit from the enhanced reasoning. Routine tasks like document categorization, simple summaries, or high-volume review are better suited to standard models.

Think of reasoning models as the discovery equivalent of bringing in senior counsel for the complex issues while letting junior attorneys handle routine matters. You would not have a partner review every document in a production, but you would want partner-level attention on the difficult privilege calls or the key contradictory evidence. Similarly, you would not run reasoning models on every task, but you would use them where careful analysis matters most.

Combining Reasoning Models with Other AI Capabilities

Reasoning models work well in combination with other AI capabilities. You might use standard models for initial document review to identify potentially relevant materials, then apply reasoning models to analyze the most important documents in depth. Or you might use agentic systems to gather evidence from multiple sources, then employ a reasoning model to synthesize that evidence into a coherent legal analysis.

One effective pattern is the two-stage approach: use fast, inexpensive models to filter and organize large volumes of information, then use reasoning models for the complex analytical tasks that require careful thought. This combines efficiency for routine tasks with thoroughness for difficult problems.

Another useful combination is having reasoning models verify the output of other AI systems. An agent might analyze thousands of documents and produce a report, then a reasoning model might review that report for logical consistency, check whether conclusions are properly supported, and identify any analytical gaps. This provides a form of automated quality control.

© 2025 Merlin Search Technologies, Inc.

Limitations and Cautions

Reasoning models are not perfect. They can still hallucinate, though typically less than standard models. Their extended thinking process provides more opportunities to catch and correct errors, but it does not eliminate them. The same verification practices that apply to standard model output apply equally to reasoning models.

The thinking process, while more thorough than standard models, still differs from human reasoning in important ways. The models lack true understanding and cannot apply common sense in the way humans do. They can follow logical chains impressively well, but they can also pursue reasoning paths that seem logical but rest on faulty premises. Human oversight remains essential.

Reasoning models work best with clear, well-defined tasks. Vague instructions or ambiguous questions can lead them down unproductive analytical paths. When using reasoning models, invest time in crafting precise prompts that specify what analysis you want, what factors to consider, and what kind of conclusion you need.

The extended thinking time can sometimes be unproductive. The model might explore analytical paths that turn out to be irrelevant, or spend time on considerations that do not ultimately matter. This is part of the deliberation process, but it means that longer thinking time does not always equal better results. For simple questions, the extra time provides no benefit.

Best Practices for Using Reasoning Models

Several practices will help you get the most from reasoning models:

Be specific about what reasoning you want: Rather than asking a reasoning model to “analyze this document,” explain what kind of analysis matters. “Evaluate whether this email is protected by attorney-client privilege, considering the parties involved, the subject matter, and whether the primary purpose was to provide legal advice” gives the model clear analytical direction.

Provide relevant context: The model reasons better when it has the necessary background. Include relevant legal standards, case-specific facts, and any constraints on the analysis. If you are asking about privilege, provide your jurisdiction’s privilege standards rather than making the model guess.

Use for genuinely complex tasks: Simple yes/no questions or straightforward document summaries do not benefit from extended reasoning. Save reasoning models for tasks that actually require working through multiple considerations or complex logical chains.

Request explanations of reasoning: Ask the model to show its work. Understanding how it reached a conclusion helps you verify the analysis and identify any logical errors. Many reasoning models can provide summaries of their thinking process.

Verify conclusions: Even carefully reasoned analysis can be wrong. Check citations to source documents, verify that legal standards are applied correctly, and test conclusions against your own understanding of the facts and law.

© 2025 Merlin Search Technologies, Inc.

Iterate on difficult problems: If a reasoning model's first analysis seems incomplete or misses important considerations, provide feedback and ask it to reconsider. The model can incorporate your guidance into a revised analysis.

The Future of Reasoning Models in Discovery

Reasoning capabilities are likely to improve rapidly. Current models already show impressive analytical abilities, but they will become more sophisticated at identifying relevant considerations, catching their own errors, and reasoning about complex legal and factual problems. We can expect faster reasoning with better cost-efficiency, making these capabilities practical for a broader range of discovery tasks.

Integration with specialized legal knowledge will make reasoning models more effective. Models trained on case law, discovery practices, and legal reasoning will better understand what matters in legal analysis and how to apply legal standards to facts. This specialized knowledge combined with general reasoning ability will create powerful tools for legal analysis.

We will see reasoning models integrated into complete discovery workflows. Rather than being separate tools you invoke for specific tasks, reasoning capabilities will become embedded in document review platforms, investigation systems, and case management tools. The software will recognize when a task requires careful analysis and automatically apply reasoning models where appropriate.

Perhaps most intriguingly, reasoning models will enable AI systems that can explain their decisions in ways that satisfy legal and professional standards. When a model makes a privilege determination or evaluates document relevance, it will be able to provide a clear, logical explanation of its reasoning that lawyers can review, critique, and defend if challenged. This transparency will make AI analysis more trustworthy and defensible.

Conclusion

Reasoning models represent a significant advance in AI capabilities for discovery and investigation. Their ability to work through complex problems methodically, maintain logical consistency across multi-step analysis, and evaluate their own conclusions makes them valuable tools for the difficult analytical tasks that discovery professionals face daily.

These models are not replacements for human judgment. They are cognitive tools that extend our analytical capabilities, much as calculators extended our mathematical capabilities. Used thoughtfully, they help us analyze evidence more thoroughly, identify patterns we might miss, and work through complex logical problems more systematically. Used carelessly, they can lead us astray with confident-sounding but flawed analysis.

The key is knowing when and how to deploy reasoning capabilities. For routine tasks, standard models work fine. For complex analytical problems where logical rigor matters, reasoning models provide genuine value. As these capabilities mature and become more accessible, discovery professionals who learn to use them effectively will have significant advantages in handling complex investigations and litigation.

© 2025 Merlin Search Technologies, Inc.

The introduction of reasoning models marks another step in AI's evolution from simple automation to genuine cognitive assistance. Discovery professionals who understand these capabilities and integrate them thoughtfully into their practice will be better equipped to handle the increasingly complex and data-intensive investigations that characterize modern litigation.

© 2025 Merlin Search Technologies, Inc.

MERLIN

08

Agentic Workflows in Discovery

© 2025 Merlin Search Technologies, Inc.

MERLIN

CHAPTER EIGHT

Agentic Workflows in Discovery

In Chapter 6, we explored the components and architectures that make AI systems "agentic"—capable of planning, executing multi-step workflows, and iterating toward goals autonomously. Now we examine how these abstract capabilities translate into practical discovery applications that transform how legal professionals conduct investigations and analyze evidence.

The workflows we explore in this chapter go beyond simple prompt-response interactions. They represent sophisticated investigative processes where AI systems autonomously pursue complex objectives, make strategic decisions, adapt to emerging information, and produce comprehensive work product that previously required days or weeks of human effort.

The workflows we explore represent a spectrum of capabilities. Some, like autonomous document analysis and timeline construction, are beginning to appear in production systems. Others, particularly sophisticated multi-agent coordination and real-time deposition support, represent near-term possibilities that current technology could support but that haven't yet been fully realized in commercial products. Understanding this full range prepares you to evaluate emerging vendor offerings and envision how discovery work will evolve.

Document Analysis Agents

Document analysis represents one of the most promising applications of agentic AI in discovery, with early implementations beginning to appear in commercial systems.. Traditional document review follows linear patterns: search for documents, review results, refine searches, repeat. Agentic document analysis agents transform this into autonomous investigations that search strategically, evaluate results critically, and iterate until they achieve comprehensive understanding.

The Traditional Document Analysis Challenge

Discovery professionals routinely face scenarios like this:

A partner asks: "We need to understand what the board knew about the product defect and when they knew it. Review the document production and give me a comprehensive report."

Traditional workflow requires:

- Developing keyword searches for board-related documents
- Reviewing search results to identify relevant materials
- Reading board minutes, emails, and reports
- Building timeline of board knowledge
- Identifying gaps requiring additional searches
- Conducting supplemental searches

© 2025 Merlin Search Technologies, Inc.

- Synthesizing findings across sources
- Writing comprehensive report

This process might consume several days of attorney or senior paralegal time, involving iterative searching, extensive reading, and careful synthesis. Each iteration requires human decision-making about next steps.

How Document Analysis Agents Work

An agentic document analysis system would approach this same task autonomously. While the level of sophistication varies across current implementations, the workflow demonstrates the technology's potential:

Phase 1: Goal Understanding and Initial Planning

The agent receives the goal: "Understand what the board knew about the product defect and when they knew it."

The agent analyzes this goal, identifying key requirements:

- Subject matter: product defect information
- Key actors: board members
- Temporal component: when knowledge was acquired
- Deliverable: comprehensive report with timeline

The agent develops an initial strategic plan:

1. Identify who served on the board during relevant period
2. Search for board-related documents (minutes, presentations, reports)
3. Search for documents about the product defect
4. Identify intersection of board and defect documents
5. Analyze these documents for evidence of board knowledge
6. Build timeline of when information reached board
7. Identify key documents requiring detailed analysis
8. Search for related communications explaining or following up on board discussions
9. Assess completeness and identify gaps
10. Conduct targeted searches to address gaps
11. Synthesize findings into comprehensive narrative
12. Generate timeline visualization
13. Compile report with citations to source documents

Phase 2: Information Gathering and Analysis

The agent begins executing its plan, but unlike a human following a checklist, it adapts continuously based on what it discovers.

Initial board identification: The agent searches for documents mentioning "board of directors," "board meeting," "board members." It analyzes results to extract names of individuals

© 2025 Merlin Search Technologies, Inc.

serving on the board, compiling a list of key actors.

Strategic search execution: Rather than executing a single search, the agent orchestrates multiple search strategies simultaneously or sequentially:

- Semantic search for documents conceptually about product defects
- Boolean searches for specific technical terms about the defect
- Custodian-based searches focusing on board members identified
- Date-restricted searches around periods when defect was reported
- Email thread searches to capture board communications

The agent doesn't just execute these searches—it evaluates results. After each search, it assesses:

- Are results relevant to the goal?
- What new information do results reveal?
- What patterns emerge?
- What's still missing?
- What should be searched next?

Adaptive investigation: As the agent analyzes initial results, it discovers information that modifies its approach. Perhaps it finds:

- A critical email from an engineer to a board member six months earlier than expected
- References to a committee studying the issue (requiring committee searches)
- Technical reports presented to the board (requiring searches for those specific reports)
- Code names or project names used internally (requiring searches with those terms)

The agent dynamically adjusts its plan to pursue these leads:

Updated Plan:

- Original items 1-5: Completed
- NEW: Search for emails from Engineer X to board members 6+ months earlier
- NEW: Identify and search Safety Committee documents
- NEW: Search for "Project Phoenix" (code name discovered)
- NEW: Search for technical reports cited in board presentations
- Continue with original items 6-13, modified based on new findings

Phase 3: Deep Analysis and Synthesis

As relevant documents accumulate, the agent performs sophisticated analysis:
Document clustering: The agent groups related documents—email threads, meeting minutes with corresponding presentations, follow-up communications, etc. This clustering reveals relationships humans might miss across large document populations.

Timeline construction: The agent extracts temporal information from documents, building a chronology of when information about the defect reached various board members, when board meetings discussed it, and when decisions were made. This timeline identifies critical periods and potential gaps.

© 2025 Merlin Search Technologies, Inc.

Knowledge attribution: For each piece of defect information, the agent tracks:

- When did it first appear in documents?
- Which board members received it?
- In what context (board meeting, email, report)?
- What actions followed?
- Were there later contradictory statements?

Gap identification: The agent identifies information gaps that additional searching might fill:

- Board meetings without accessible minutes
- Time periods with no relevant communications
- Board members not appearing in documents when they should
- Technical details mentioned but not explained in available documents
- References to documents not found in production

Phase 4: Targeted Gap-Filling

Based on gap identification, the agent conducts targeted follow-up investigations:

Precision searches: Rather than broad searches, the agent executes highly specific searches designed to find particular missing information:

- “Board meeting” AND “March 15” (to find minutes from specific meeting)
- Custodian:”Board Member X” AND date range (to find that member’s communications during silent period)
- Attachment searches for technical reports referenced but not found

Relationship exploration: The agent explores relationships between documents and individuals:

- Who corresponded with whom about the defect?
- Which external parties provided information to board members?
- What documents were attachments to communications analyzed?

Iterative refinement: If targeted searches don’t fill gaps, the agent tries alternative approaches:

- Broader date ranges
- Related terms or synonyms
- Communications mentioning the missing documents
- Correspondence with administrative staff who might have relevant information

Phase 5: Synthesis and Reporting

After thorough investigation, the agent synthesizes findings:

Narrative construction: The agent builds a coherent narrative answering the original ques-

© 2025 Merlin Search Technologies, Inc.

tion:

- What defect information existed and when it was created
- How that information reached board members
- What board discussions occurred
- What decisions or actions board took
- Evolution of board knowledge over time
- Any contradictions or inconsistencies in the record

Evidence support: Every substantive claim in the narrative includes citations to specific documents with IDs allowing immediate verification:

"On January 15, the board received a preliminary report indicating potential defects in the product's safety mechanism (Doc ID 894523). The report noted that testing had revealed failure rates of 2.3% under stress conditions (Doc ID 894523, p. 7)."

Timeline generation: The agent creates a detailed timeline showing key events, knowledge milestones, and decision points, each linked to supporting documents.

Completeness assessment: The agent evaluates its own work:

- Were all aspects of the goal addressed?
- Are there remaining gaps or uncertainties?
- What information proved impossible to find?
- What alternative interpretations exist?

Final report compilation: The agent produces a comprehensive report including:

- Executive summary answering the original question
- Detailed narrative with full citations
- Timeline visualization
- List of key documents
- Methodology description
- Identified gaps or limitations
- Recommendations for further investigation if needed

Advantages Over Traditional Approaches

As these systems mature, agentic workflows will provide several advantages:

1. **Comprehensiveness:** The agent systematically pursues the question from multiple angles, less likely to miss relevant documents than human reviewers who might not think of all search strategies.

2. **Consistency:** The agent applies consistent analytical criteria across all documents, avoiding fatigue or attention lapses that affect human review.

3. **Speed:** What might require days of human work completes in hours or less, with the agent working continuously without breaks.

© 2025 Merlin Search Technologies, Inc.

4. **Documentation:** The agent's process creates a complete audit trail of search strategies, decision points, and analytical steps, supporting defensibility.

5. **Adaptability:** The agent dynamically adjusts its approach based on discoveries, pursuing leads that emerge during investigation.

6. **Scale:** The agent can process document volumes that would overwhelm human reviewers, maintaining quality across large populations.

Investigation Agents: Building Timelines and Mapping Relationships

Beyond analyzing documents to answer specific questions, agentic systems are being developed to conduct broader investigations that construct timelines, map relationships between entities, and identify patterns across large document populations. Some of these capabilities exist in production systems today, while others represent near-term development targets.

Timeline Construction Agents

Chronology proves critical in many investigations. What happened when? Who knew what at each point? How did events unfold? Constructing comprehensive timelines from scattered documents traditionally requires extensive manual work. Timeline construction agents automate this process while maintaining accuracy.

How timeline agents work:

The agent searches across the document population to identify relevant materials, then extracts temporal information from each document—not just document dates, but dates mentioned in content, temporal relationships between events, and even estimated timeframes when precise dates aren't stated.

From this temporal data, the agent identifies discrete events worthy of timeline inclusion: meetings, decisions, regulatory filings, personnel changes, and other significant occurrences. For each event, it associates relevant entities (people, organizations, products) and supporting documents. The agent then maps relationships between events, identifying causal connections, sequential dependencies, and contradictions that require reconciliation.

Critically, the agent identifies temporal gaps that suggest missing information—long periods without documented activity, events that should have intermediate steps, or referenced communications without documentation. These gaps guide targeted follow-up searches.

The agent produces timeline outputs in multiple formats: chronological narratives with document citations, visual timelines, sortable event tables, and network diagrams showing relationships between events.

Example application:

Consider an investigation into events leading to a product recall. The agent searches broadly, identifies thousands of relevant documents, and extracts dates and events from each. It iden-

© 2025 Merlin Search Technologies, Inc.

tifies key events spanning product design decisions through customer complaints, internal investigations, executive discussions, and regulatory communications.
The agent maps how customer complaints triggered an internal investigati
on, which revealed a manufacturing issue traceable to a specific process change decided by particular executives. It identifies critical gaps—a two-month period between first complaints and internal investigation, missing documentation of executive discussions before the recall decision, and absent quality control reports during the critical period. Based on these gaps, it conducts targeted searches to find missing information.

The result: a comprehensive timeline with 200+ events, each linked to supporting documents, along with visual representations highlighting key decision points and flags for events requiring additional investigation.

Value for discovery:

Timeline agents—in current and emerging forms—process document volumes humans cannot manually review, identify temporal patterns across scattered materials, and find gaps suggesting missing or withheld documents. They support both case strategy and depositions while creating demonstrative aids for hearings or trial. The automation doesn't replace professional judgment—it provides comprehensive chronological frameworks that legal professionals can analyze, challenge, and refine.

Relationship Mapping Agents

Understanding who communicated with whom about what, which organizations had business relationships, and how entities connect reveals patterns critical to investigations. Relationship mapping agents extract and visualize these connections.

How relationship mapping agents work:

The agent extracts entities from document populations—people, organizations, products, locations, and concepts—then identifies relationships between them. These relationships span multiple dimensions: communications patterns, organizational affiliations, business partnerships, knowledge sharing, spatial connections, and conceptual links around shared topics.

For each relationship, the agent characterizes its nature: strength based on interaction frequency, directionality (one-way versus reciprocal), temporal patterns showing when relationships existed, context distinguishing business from personal connections, topics discussed, and even sentiment or tone. The agent builds network representations where nodes represent entities and edges represent relationships, with edge weights reflecting relationship strength and temporal layers showing evolution over time.

The agent identifies significant patterns—central actors who are highly connected, clusters of entities with dense internal connections, and bridges connecting otherwise separate groups. It maps information flow paths and flags unusual relationships or unexpected connections. Critically, it performs anomaly detection, identifying communications outside expected patterns, connections between entities that shouldn't interact, sudden relationship intensity changes, individuals apparently avoiding documentation, and information silos.

© 2025 Merlin Search Technologies, Inc.

Example application:

Consider mapping relationships among executives, board members, and consultants during merger negotiations. The agent extracts 147 individuals and 23 organizations from the document production, then analyzes communications to identify relationship patterns. It discovers that the CFO had frequent communications with a competing bidder, an external consultant appeared in many board discussions despite little documentation of retention, separate communication streams existed between board committees and management, key strategic decisions were made via personal rather than corporate email accounts, and one board member never appeared in any merger-related documents.

The agent maps the communication network, revealing primary participants and interaction frequencies, information flow from legal counsel through the CEO to the board, side channels bypassing formal processes, and external parties receiving internal information. It flags the CFO's competitor communications, the undocumented consultant role, the absent board member, and the use of personal accounts for corporate decisions—each warranting investigation. The deliverable includes visual network diagrams and targeted search recommendations for follow-up.

Value for discovery:

Relationship mapping agents reveal hidden connections among parties, identify individuals with knowledge relevant to depositions, expose information flows and potential leaks, and support conspiracy or fraud investigations. They process communication volumes that would overwhelm manual analysis, identifying patterns invisible in small samples. The resulting network visualizations guide strategic investigation priorities and create compelling demonstrative evidence for hearings or trial.

Real-Time Deposition Support

One of the most ambitious agentic applications under development is real-time support during depositions, where attorneys need immediate access to testimony, documents, and case information. While elements of this capability exist today—such as transcript search and document retrieval—the integrated system described here represents a near-term development target rather than a current product.

The Deposition Challenge

During depositions, attorneys must listen carefully to testimony, formulate follow-up questions, identify contradictions with prior statements, recall relevant documents, adjust examination strategy dynamically, and take detailed notes. This cognitive load makes it difficult to simultaneously search case materials for supporting documents or prior testimony. Associates often sit beside examining attorneys to search documents, but even skilled associates can't search fast enough to keep pace with dynamic examinations.

Agentic Deposition Assistant

An agentic deposition assistant would provide immediate support throughout the deposition

process. Before the deposition, the agent analyzes all prior testimony from the witness, reviews associated documents, identifies key topics likely to arise, and builds an anticipatory search index. During the deposition, it processes the transcript in real-time via court reporter live feed or audio-to-text, identifies key claims or statements, automatically searches for related documents, flags contradictions with prior testimony, and surfaces relevant documents proactively.

The dynamic assistance works like this: When a witness claims "I never saw that report before March," the agent immediately searches for the report, finds an email from January showing the witness received it, alerts the attorney to the contradiction, and displays the email with relevant portions highlighted. When a witness mentions "We discussed that in the board meeting," the agent retrieves board meeting minutes, identifies the relevant discussion, summarizes what the minutes say, and flags discrepancies between testimony and minutes. When an attorney asks about "communications with Acme Corp," the agent pulls all Acme Corp communications, summarizes key themes, identifies documents not yet covered in questioning, and suggests follow-up questions.

After the deposition, the agent generates a detailed summary of testimony, identifies all contradictions with documents or prior testimony, creates an issue-spotting memo highlighting significant testimony, updates the case knowledge base with new information learned, and generates recommended document requests based on testimony gaps.

Implementation Requirements

Effective deposition agents will require low-latency operation with responses arriving within seconds, high accuracy to avoid disrupting examination flow with false contradictions or irrelevant documents, unobtrusive interfaces delivering information via tablet or laptop without distracting the attorney from the witness, attorney control allowing explicit queries or autonomous operation, and confidence signaling so attorneys can judge the reliability of findings.
Multi-Agent Investigation Systems

For complex investigations requiring diverse analytical approaches, multi-agent systems represent an emerging architecture where specialized agents collaborate under coordination frameworks. While research prototypes demonstrate this approach, commercial implementations are still developing.

Rather than one generalist agent handling everything, a sophisticated multi-agent discovery system deploys specialized agents for search strategy, document classification, content analysis, timeline building, relationship mapping, synthesis, and quality assurance. Each agent excels at its specific role, leading to higher quality results than a single agent could achieve. Multiple agents work simultaneously, dramatically reducing completion time while maintaining quality through verification processes.

The multi-agent example that follows illustrates how current technology could be orchestrated to handle complex discovery tasks. Consider this a roadmap for where the technology is heading rather than a description of widely available products.

© 2025 Merlin Search Technologies, Inc.

Example: Multi-Agent Product Liability Investigation

This example demonstrates how specialized agents could collaborate in a mature implementation to comprehensively analyze documents related to product defect allegations.

The investigation begins with a Search Strategist Agent developing search strategies covering product names, defect descriptions, customer complaints, engineering documents, quality control, and executive communications—yielding 50,000 potentially relevant documents. A Document Classifier Agent categorizes these into customer communications, internal emails, engineering documents, quality control reports, executive materials, and marginally relevant documents, routing high-priority categories forward while flagging potentially privileged documents for human review.

In the detailed analysis phase, a Content Analyzer Agent reads the high-priority documents, extracting key information about defect awareness, engineering responses, and executive decisions while generating summaries and identifying documents requiring deep analysis. Simultaneously, a Timeline Builder Agent processes documents to construct a chronology showing first customer complaints 18 months before recall, engineering investigation initiation, defect confirmation in testing, executive presentations, regulatory notification, and recall announcement—identifying significant timeline gaps requiring investigation. A Relationship Mapper Agent analyzes communications to map information flow from customer service through engineering to executives, identifying which executives received what information when, flagging concerning patterns like executives avoiding documentation.

Based on gaps and anomalies identified by the Timeline and Relationship agents, the Search Strategist Agent conducts a second pass with targeted searches for missing engineering reports, executive communications during suspicious periods, and external consultant correspondence. The Content Analyzer performs detailed analysis of newly discovered documents, cross-referencing with existing understanding and identifying contradictions requiring resolution.

A Synthesis Agent combines findings from all specialized agents into a comprehensive narrative addressing what defect existed and when it was discovered, how information flowed through the organization, what decisions were made and by whom, whether responses were adequate and timely, and evidence of knowledge or intent. It identifies key documents and witnesses for litigation and develops theories of liability and defense. A Quality Assurance Agent reviews this narrative against source documents, verifies that citations support claims, checks timeline accuracy, identifies unsupported conclusions, and recommends additional investigation where needed. A Coordination Agent compiles final deliverables: executive summary, detailed investigative report with citations, timeline with all events, relationship network diagrams, key document list, witness list with knowledge summaries, and recommendations for depositions and discovery requests.

Finally, a senior attorney reviews the multi-agent investigation results, verifies key findings against source documents, identifies areas needing additional investigation, and provides feedback. Based on this feedback, the Coordination Agent directs specialized agents to conduct follow-up investigation, repeating the process until the attorney determines the investigation is sufficiently complete.

© 2025 Merlin Search Technologies, Inc.

Advantages of Multi-Agent Systems

When fully realized, multi-agent architectures would provide significant benefits. Specialization allows each agent optimized for its specific task to perform better than generalist approaches. Parallelization enables multiple agents to work simultaneously, dramatically reducing completion time. Quality assurance through dedicated QA agents catches errors before human review. The architecture offers scalability (additional agents can be added without redesigning the system), modularity (individual agents can be updated or replaced independently), and robustness (if one agent fails, others continue working and the system degrades gracefully rather than failing completely).

Looking Ahead

This chapter demonstrated how agentic AI capabilities translate into practical discovery workflows. The examples span a range of development stages: document analysis agents with basic autonomous capabilities are entering the market now, timeline construction and relationship mapping tools exist in limited forms, while sophisticated multi-agent investigation systems remain primarily in development. Together, they illustrate how discovery work will transform as these technologies mature.

As these capabilities develop, they won't replace human professionals—they'll augment human expertise, handling volume and routine analysis while legal professionals focus on strategy, judgment, and critical decision-making. The emerging result is discovery work that's faster, more comprehensive, and more consistent while remaining under professional oversight.

Chapter 9 addresses the critical question of how to implement and control these systems effectively. Agentic AI's power brings risks—autonomous systems can pursue unproductive directions, consume excessive resources, or produce outputs requiring extensive verification. The next chapter provides frameworks for deploying agentic systems with appropriate controls, ensuring they serve professional needs without creating unacceptable risks.

We'll examine when agentic approaches are appropriate versus when simpler methods suffice, how to set boundaries and guardrails, how to monitor agent behavior, and how to maintain human oversight while preserving efficiency benefits. These implementation considerations determine whether agentic AI becomes a valuable tool or an expensive liability in your practice.

© 2025 Merlin Search Technologies, Inc.

MERLIN

09

Ethical Considerations for Agentic Systems

© 2025 Merlin Search Technologies, Inc.

CHAPTER NINE

Ethical Considerations for Agentic Systems

Chapter 5 established the general ethical framework for using generative AI in legal practice—the duty of competence, confidentiality obligations, verification requirements, transparency expectations, and billing practices. Those principles apply to all AI use in legal work. But agentic systems raise additional ethical questions that warrant specific attention.

Agentic systems execute multi-step investigations between checkpoints. This improves efficiency but requires clear boundaries, auditable operations, and human control over strategy and conclusions. This chapter examines the ethical considerations unique to agentic AI—how increased autonomy affects professional responsibilities, what transparency obligations arise when systems operate autonomously, who bears responsibility when agents make mistakes, how to ensure quality and reliability standards are met, what limits should be placed on agent autonomy, and how to maintain the professional judgment that defines legal practice.

Autonomy and Professional Responsibility

The defining characteristic of agentic systems is their autonomy—the ability to pursue goals through self-directed actions rather than executing prescribed steps. This autonomy creates tension with professional responsibility rules premised on lawyers maintaining direct control over legal work.

The Control Question

Traditional professional responsibility rules assume lawyers directly control or closely supervise all work performed on behalf of clients. When a lawyer delegates work to an associate or paralegal, the lawyer maintains oversight through assignment instructions, interim reviews, and final approval. The supervising lawyer can intervene at any point, redirect the work, or reject outputs that don't meet standards.

Agentic systems operate differently. Once launched toward a goal, they make numerous operational decisions autonomously—which documents to retrieve, what search strategies to pursue, how to analyze findings, when to iterate versus moving forward. While humans establish goals and boundaries at the outset and review outputs at checkpoints, the agent operates independently between these touchpoints.

This raises the fundamental question: Does autonomous agent operation constitute improper delegation of professional responsibility? The answer depends on understanding what agents actually control versus what remains under human direction.

What Lawyers Control

Even with agentic systems, lawyers maintain control over all substantive and strategic deci-

sions. Lawyers decide what investigations to pursue, what goals to establish, what boundaries to set, what resources to authorize, when to intervene based on checkpoint reviews, what conclusions are supported by evidence, how to present findings to clients or courts, and what actions to recommend. Agents operate within the framework established by these human decisions.

The autonomy agents exercise is operational, not substantive. Agents decide how to search efficiently within defined scope and which documents to prioritize but don't determine investigation strategy or what conclusions follow from findings.

This distinction matters ethically. Professional responsibility rules require lawyers to exercise independent professional judgment on substantive matters affecting client interests. Those rules don't require lawyers to personally perform every operational task. The question is whether the lawyer maintains appropriate oversight and makes all substantive decisions.

Appropriate Delegation Boundaries

Agentic systems remain within appropriate delegation boundaries when lawyers establish clear investigation goals reflecting professional judgment about what information clients need, set boundaries ensuring agents operate within authorized scope, implement checkpoints providing opportunities to review progress and redirect if needed, verify agent outputs meet quality standards before use, and make all substantive decisions about conclusions and recommendations.

Agentic systems cross into inappropriate delegation when agents make strategic decisions about investigation scope or direction, determine what conclusions are supported by evidence without human verification, communicate findings or recommendations directly to clients without lawyer review, or make decisions involving professional judgment or ethical considerations.

The ethical line is that agents can autonomously execute work within parameters established by lawyers, but all decisions requiring professional judgment remain with human lawyers who bear ultimate responsibility.

Transparency and Disclosure Obligations

Transparency enables informed client consent, supports verification and accountability, facilitates oversight, and maintains trust in the legal system. Agentic systems raise particular challenges because their autonomous operation makes it harder to explain exactly what was done and why.

Transparency to Clients

Lawyers have ethical obligations to keep clients informed about significant developments in their matters and to explain matters to the extent necessary to permit clients to make informed decisions. These obligations apply when using agentic AI to perform work for clients. The threshold question is whether AI use constitutes a "significant development" requiring client notification. The answer depends on the context. When agentic AI is used for routine work where the method employed doesn't affect strategy or outcomes in ways clients should know

© 2025 Merlin Search Technologies, Inc.

about, notification may be unnecessary. When AI use involves novel methods, affects cost or timing significantly, or raises questions clients might reasonably want input on, disclosure becomes appropriate.

Best practice points toward proactive transparency. Many firms now include language in engagement letters indicating that AI tools may be used in providing legal services, subject to appropriate safeguards and oversight. This general disclosure establishes client awareness and consent without requiring notification for each specific instance.

When agentic systems produce findings or recommendations that will significantly affect client decisions, lawyers should explain the methodology. This doesn't require technical details about LLM architecture—it requires explaining in accessible terms what was done, how reliability was ensured, and what the findings mean.

Transparency to Courts and Opposing Counsel

Courts have inherent authority to inquire about methods used in litigation, and procedural rules may require disclosure in specific contexts like discovery disputes or expert testimony. When methodology becomes material—perhaps because opposing counsel questions completeness of review or because a discovery dispute raises questions about methods used—transparency becomes essential. Lawyers should not misrepresent AI-assisted work as entirely human-performed, nor should they obscure material information about methods when disclosure is relevant to assessing reliability or completeness.

Transparency About Limitations and Uncertainties

Most importantly, lawyers must be transparent about limitations and uncertainties in AI-assisted work. When agentic systems report that findings are based on limited information, when verification reveals potential gaps, or when results contain uncertainties, these limitations must be communicated appropriately. This transparency serves both ethical obligations and strategic interests. Clients deserve to know when conclusions rest on uncertain foundations. Courts deserve to understand limitations in evidence or arguments presented to them. Practically, this means preserving and communicating the caveats and qualifications that should appear in agent outputs. When agents report "based on available documents" or "three sources provide conflicting information," those qualifications should carry through to how findings are presented to clients and others.

Accountability and Error Attribution

When agentic systems make mistakes, questions arise about who bears responsibility and how accountability operates. These questions have both practical and ethical dimensions.

The Lawyer's Ultimate Responsibility

The foundational principle is clear: lawyers remain professionally responsible for all work product provided to clients, regardless of what tools or assistants contributed to that work. This applies to work performed by associates, paralegals, vendors, AI systems, or any other contributor. The lawyer's professional responsibility is non-delegable.

© 2025 Merlin Search Technologies, Inc.

When an agentic system produces inaccurate findings, misrepresents source documents, overlooks critical information, or makes other errors, the lawyer using that system bears professional responsibility. The lawyer cannot deflect responsibility by claiming the AI made the mistake. This principle is well established in other delegation contexts. When an associate makes a research error, the supervising partner remains responsible. The same applies to AI tools—they're tools the lawyer chose to use, and the lawyer remains responsible for outputs generated with those tools.

Comparative Fault and Mitigation

While ultimate responsibility rests with lawyers, practical questions of comparative fault and mitigation remain relevant. When a lawyer took all reasonable precautions but an AI system failed in unpredictable ways, disciplinary authorities may view the situation differently than when a lawyer deployed AI recklessly without safeguards.

Taking reasonable precautions doesn't eliminate lawyer responsibility but may mitigate consequences. A lawyer who implemented robust verification protocols, conducted appropriate sampling, and had reasonable systems for catching errors has exercised better judgment than a lawyer who blindly relied on AI outputs without verification—even if errors occurred in both cases. This creates practical incentives for robust quality assurance systems.

Learning from Errors

Accountability includes learning from mistakes to prevent recurrence. When agentic systems produce errors, organizations should investigate root causes, assess whether verification protocols should be enhanced, and determine whether guardrails should be adjusted.
Documentation plays an important role here. Maintaining records of errors identified, investigations conducted, improvements implemented, and lessons learned creates evidence of responsible oversight and continuous improvement.

Quality and Reliability Standards

Agentic systems must meet quality and reliability standards appropriate for legal work. These standards go beyond avoiding outright errors to ensuring outputs reliably meet professional standards.

Defining Adequate Quality

The quality threshold for agentic outputs should match what would be expected from human-performed work in similar contexts. For high-stakes litigation work where accuracy is critical, agentic outputs should meet the same stringent standards expected from senior attorneys. For preliminary internal investigations where some error tolerance exists, standards might be more relaxed—but not so relaxed that outputs are unreliable or misleading.

This situational approach recognizes that legal work involves varying stakes and quality requirements. Not every task requires perfect accuracy—some preliminary work accepts higher error rates as the trade-off for speed and efficiency. But even preliminary work must meet minimum reliability thresholds. Outputs that are more likely wrong than right, or that contain

© 2025 Merlin Search Technologies, Inc.

so many errors that they mislead rather than inform, fail to meet professional standards regardless of their preliminary nature.

Practically, this means calibrating agentic systems to task requirements. High-stakes work requires more stringent guardrails, more intensive verification, more conservative confidence thresholds, and more frequent human checkpoints. Lower-stakes work can accept somewhat lighter oversight—but never to the point where quality falls below acceptable minimums.

Reliability Over Time

Quality standards include reliability—the expectation that systems produce consistent results over time rather than alternating between excellent and poor performance unpredictably. Unreliable systems create risks even if average performance is acceptable, because users can't distinguish when to trust outputs versus when to be skeptical.

This requires ongoing monitoring of system performance. Initial validation showing a system performs well doesn't ensure continued reliability. Systems may degrade over time as underlying models are updated, as document collections change character, or as edge cases emerge that weren't present during initial testing. Regular quality checks ensure continued reliability and identify when performance degrades requiring intervention.

When reliability problems emerge, organizations must respond appropriately. Continued use of systems known to produce unreliable outputs violates professional obligations. The response might involve enhanced verification protocols, adjusted guardrails, more intensive human oversight, or discontinuing use for particular applications until reliability issues are resolved.

Comparative Reliability

An important consideration is how agentic system reliability compares to alternatives. Agentic systems need not be perfect to be ethically usable—they need to be as reliable as or more reliable than the alternatives they replace, considering both accuracy and efficiency. If human review of thousands of documents produces certain error rates and patterns, agentic review should aim for comparable or better accuracy while offering efficiency advantages. The standard is whether agentic approaches meet professional quality expectations considering the full context of alternatives, stakes, and trade-offs.

Autonomy Limits and Human Judgment

Determining appropriate limits on agent autonomy requires balancing efficiency benefits against risks of excessive delegation. Some decisions inherently require human judgment and should never be delegated to autonomous systems.

Decisions Requiring Human Judgment

Certain categories of decisions must remain exclusively in human hands because they involve professional judgment, ethical considerations, or client interests that machines cannot appropriately weigh.

© 2025 Merlin Search Technologies, Inc.

Strategic decisions about case direction, investigation scope, resource allocation, and litigation approaches require human judgment about client interests, risk assessment, and professional experience. Agentic systems can inform these decisions by gathering and analyzing information, but the decisions themselves must be made by lawyers exercising professional judgment.

Ethical decisions involving privilege, conflicts, disclosure obligations, billing practices, or professional conduct require applying ethical rules and professional judgment to specific facts. These decisions implicate lawyer responsibilities that cannot be delegated to autonomous systems. When agents encounter potential privilege issues, possible conflicts, or uncertain ethical questions, they must escalate to humans rather than attempting to resolve these matters autonomously.

Client counseling about strategy, settlement, risk assessment, or important decisions requires the attorney-client relationship and human understanding of client interests, values, and objectives. Agents should not communicate directly with clients about strategic matters or provide advice without human lawyer review and approval.

Novel situations without clear precedent require human judgment to navigate uncertainty and ambiguity. When agents encounter unprecedented patterns, unusual facts, or situations that don't fit established frameworks, they should escalate to humans who can apply judgment, creativity, and experience to novel problems.

Operational Autonomy Within Boundaries

Within appropriate boundaries, operational autonomy creates value without raising ethical concerns. Agents can autonomously decide which search strategies to use within validated methods, which documents to prioritize among relevant candidates, what level of detail to apply to different documents, how to structure outputs for clarity, when to iterate on searches versus moving forward, and what tools to use for particular tasks.

These operational decisions implement the strategy lawyers established but don't determine what that strategy should be. They involve technical optimization rather than professional judgment. Allowing agents to make these decisions autonomously captures efficiency benefits while preserving human control over matters requiring professional judgment.

The key is ensuring boundaries are clear, preventing agents from drifting into decisions requiring human judgment, implementing escalation mechanisms for edge cases, and maintaining human oversight at strategic points.

The Right Level of Autonomy

Finding the right autonomy level requires experimentation and adjustment. Too little autonomy eliminates the efficiency benefits justifying agentic approaches—if humans must approve every operational decision, the system becomes too cumbersome to provide value. Too much autonomy creates risks of inappropriate delegation and loss of necessary control.

Organizations should start with relatively limited autonomy, gradually expanding as experience builds confidence and as verification demonstrates that agents respect boundaries and make

© 2025 Merlin Search Technologies, Inc.

appropriate operational decisions. This graduated approach allows learning what autonomy levels work well for different tasks and contexts while managing risk of excessive delegation. Regular review of what decisions agents are making autonomously, whether those decisions reflect appropriate operational choices, whether agents are properly escalating decisions requiring human judgment, and whether the balance between autonomy and control is appropriate for each application guides adjustments over time.

Competence and the Duty to Understand

Professional responsibility rules require lawyers to maintain competence in technologies they use. For agentic systems, this creates particular challenges because the systems are complex and their operation can be opaque.

What Lawyers Should Understand

Lawyers using agentic systems need not become AI experts or understand technical implementation details, but they should understand what agentic systems can do reliably and what they cannot, what types of errors are possible and how to detect them, what verification and oversight are necessary, how to interpret confidence levels and caveats in outputs, what the boundaries and limitations are for specific applications, and when to seek expert assistance or additional verification.

This understanding should be practical rather than theoretical. Lawyers should know enough to use systems appropriately, identify when problems may exist, implement adequate safeguards, and explain to clients or courts in accessible terms what was done and why it's reliable.

Building Competence

Building this competence requires training, experience, and ongoing learning. Organizations deploying agentic systems should provide comprehensive training covering how the systems work at a conceptual level, what applications are appropriate, what safeguards and verification are required, how to identify potential problems, and where to get help when issues arise. Competence develops through supervised use where experienced reviewers assess whether users are deploying systems appropriately, interpreting outputs correctly, implementing adequate verification, and recognizing limitations. Over time, users develop intuitions about when systems are working well versus when problems may exist. Ongoing learning is essential because capabilities and limitations evolve.

Competence includes recognizing when expertise beyond one's own is needed. Lawyers need not be AI experts, but they should know when to consult with technical experts, when to seek specialized verification, when to engage vendors with relevant expertise, and when applications are complex enough to require additional oversight. This parallels how lawyers handle other technical areas—knowing when specialized expertise is needed to assess reliability, investigate problems, or validate approaches.

Preserving Professional Independence

Professional responsibility rules require lawyers to exercise independent professional judg-

© 2025 Merlin Search Technologies, Inc.

ment on behalf of clients. This independence must be preserved when using agentic systems. A significant risk with sophisticated AI systems is over-reliance—deferring to AI outputs rather than exercising independent judgment. When systems produce polished, confident-sounding analyses, the temptation is to accept them without adequate critical evaluation.

The fundamental principle is that professional judgment must be the final arbiter of what conclusions are reached, what strategies are pursued, what recommendations are made, and what work product is delivered to clients. Agentic systems can inform these judgments by gathering and analyzing information, but the judgments themselves must be made by lawyers exercising independent professional judgment based on their evaluation of all relevant factors.

This means sometimes disagreeing with agent outputs. If an agent reaches conclusions that don't seem supported by evidence, professional judgment should override the agent's conclusions. If an agent misses information that professional judgment says is important, the human judgment should prevail. If an agent suggests approaches that professional judgment says are inappropriate, independence requires following professional judgment over AI suggestions.

Professional responsibility doesn't require that lawyer judgment always be correct—it requires that important judgments be made by lawyers exercising genuine professional judgment rather than abdicating decisions to automated systems.

Client Consent and Informed Decision-Making

Clients have rights to make informed decisions about their legal representation, including what methods are used to provide services. This raises questions about what clients should know about AI use and what consent is required.

The Scope of Required Consent

Not every tool or method requires explicit client consent. Lawyers routinely make numerous decisions about how to perform work—what legal research databases to use, what document review approaches to employ, what investigation methods to pursue—without seeking client approval for each decision. Clients hire lawyers to exercise professional judgment about such matters.

AI use generally falls within this professional discretion when it involves routine applications of technology to improve efficiency or quality, when methods don't affect strategy in ways clients should be aware of, when appropriate safeguards and oversight are in place, and when costs are consistent with what clients expect.

Explicit consent becomes appropriate when AI use is novel or experimental, when methods might affect costs or timing in ways clients should know about, when clients have expressed views about AI use, or when the particular application raises questions clients might reasonably want input on. In these situations, explaining what's being done and why, obtaining client agreement, and documenting the understanding protects both client interests and lawyer interests.

Ongoing Communication

© 2025 Merlin Search Technologies, Inc.

Beyond initial engagement, ongoing communication about AI use may be appropriate when methods affect significant strategic decisions, when findings generated through AI assistance will inform important client decisions, or when clients express interest in understanding methodology. This communication should focus on what was done, why it's reliable, what limitations exist, and what the findings mean for the client's situation.

Looking Ahead

This chapter examined the ethical considerations unique to agentic AI systems—how autonomy affects professional responsibility, what transparency obligations arise, how accountability operates when agents make mistakes, what quality standards apply, where autonomy limits should be set, what competence requires, how to preserve professional independence, and what clients should know and consent to.

These ethical considerations are fundamental to responsible agentic AI deployment. Technology that violates professional obligations or compromises client interests fails regardless of its capabilities. The frameworks discussed in this chapter provide guidance for deploying agentic systems in ways that enhance legal practice while maintaining professional and ethical standards.

But ethical deployment is only part of ensuring these powerful systems serve client interests and advance justice. Part Three shifts focus to practical applications—how discovery professionals can actually use agentic and non-agentic AI to transform document analysis, transcript review, investigation workflows, contract management, and other core legal processes. The chapters ahead demonstrate how the technical capabilities from Part Two and the ethical framework from this chapter combine to enable new approaches to fundamental legal work.

© 2025 Merlin Search Technologies, Inc.

MERLIN

PART THREE

PRACTICAL APPLICATIONS IN DISCOVERY AND INVESTIGATIONS

© 2025 Merlin Search Technologies, Inc.

PART THREE

Practical Applications in Discovery and Investigations

Parts One and Two equipped you with a comprehensive understanding of generative AI technologies—from foundational LLM concepts through advanced agentic architectures and reasoning models, from technical capabilities through ethical frameworks. Now we turn to the question that matters most for discovery professionals: How do you actually deploy these technologies to transform your work?

Part Three demonstrates practical applications across the discovery lifecycle. Rather than theoretical discussions of what AI might do someday, these chapters show what AI systems are doing now in real discovery workflows. The examples presented here are based on actual implementations, though details have been generalized to protect confidentiality and distilled to highlight patterns applicable across diverse matters.

For discovery professionals, this is where abstract capabilities become concrete value. You'll see how workflows that traditionally require weeks of attorney time can be compressed to days, how synthesis that once demanded tedious manual work across dozens of depositions can now happen systematically, and how investigations transform when AI systems can autonomously pursue leads and identify patterns across massive document populations.

What You'll Learn

Part Three explores AI applications across core discovery workflows: document analysis and review (from initial assessment through final production), transcript and testimony analysis (extracting key testimony and identifying contradictions across depositions), timeline construction (gathering dates from thousands of documents and organizing complex sequences), relationship mapping (analyzing communication patterns and building network representations), contract analysis (reviewing agreements and extracting obligations at scale), and production review (assessing opposing parties' responses for completeness). Each application demonstrates how to deploy AI effectively while maintaining quality control and meeting professional standards.

Why These Applications Matter

Understanding that agentic systems can plan and execute multi-step workflows is valuable. Knowing specifically how to deploy such systems to analyze opposing counsel's production for completeness, to build comprehensive timelines from scattered communications, or to map relationship networks across thousands of documents—that's transformative.

Successful AI deployment doesn't mean eliminating human involvement. It means augmenting human capabilities to deliver better outcomes more efficiently: better early case assessment from rapid comprehensive analysis, more thorough investigations as AI systematically identifies patterns manual review might miss, enhanced quality control through systematic verification, improved cost-effectiveness by focusing expensive human expertise where it adds most

© 2025 Merlin Search Technologies, Inc.

value, and strategic advantages from comprehensive understanding of evidence enabling analyses that would be impractical with traditional methods.
These outcomes require understanding not just what AI can do but how to deploy it effectively within the constraints of legal practice—which is precisely what Part Three provides.

How to Approach Part Three

Part Three is structured to be immediately actionable. Each chapter examines a specific discovery workflow, showing both what AI can accomplish and how to implement it responsibly. You don't need to read chapters sequentially. If timeline construction is your immediate challenge, jump to Chapter 13. If you're wrestling with transcript analysis, start with Chapter 12. Each chapter stands alone while building on the technical foundations from Parts One and Two.

As you read, consider your own matters and workflows. Where do you encounter document volumes that make comprehensive human review impractical? What investigations require following threads through multiple sources? What synthesis challenges demand coordinating information across dozens of documents? These are the applications where AI capabilities shine—and where Part Three's insights will prove most valuable.

Pay particular attention to the verification protocols in each chapter. These aren't optional extras—they're essential components of responsible AI deployment in legal contexts. Every application requires human oversight, verification, and ultimate decision-making authority. The goal is augmentation, not automation.

Moving Forward

The chapters ahead show how the architectures and capabilities from Parts One and Two manifest in concrete discovery applications. You'll see detailed workflows, specific methodologies, actual results, and critical analysis of both strengths and limitations.

Part Three bridges the gap between capability and application, between understanding what agentic systems can do in theory and knowing how to deploy them effectively in practice. The AI-augmented future of discovery isn't coming—it's here. These chapters show you how to be part of it.

Let's begin with the most resource-intensive aspect of modern discovery: document analysis and review.

© 2025 Merlin Search Technologies, Inc.

MERLIN

10 Ethical Considerations for Agentic Systems

© 2025 Merlin Search Technologies, Inc.

CHAPTER TEN

Document Analysis and Review Workflows

Parts One and Two established the foundation—how generative AI and agentic systems work, their capabilities and limitations, implementation strategies, and ethical frameworks. Now we turn to practical applications. How do discovery professionals actually use these technologies to transform document analysis, improve investigation workflows, and deliver better outcomes for clients?

This chapter examines document analysis and review workflows—perhaps the most resource-intensive aspect of modern discovery. We'll explore how AI technologies, both agentic and non-agentic, can enhance efficiency and accuracy across the document review lifecycle, from initial assessment through production review. The examples and workflows presented here are based on real implementations, though details have been generalized to protect confidentiality.

The Document Analysis Challenge

Document review has always been the most expensive, time-consuming phase of discovery. A significant litigation matter might involve reviewing hundreds of thousands or even millions of documents. At traditional hourly rates, this review can cost millions of dollars and take months to complete. The scale creates not just cost problems but quality challenges—maintaining consistency and accuracy across large review teams working on massive document populations is difficult.

Traditional approaches to managing these challenges have limitations. Keyword search casts too wide a net, retrieving both relevant documents and massive volumes of false positives requiring manual review. Technology-assisted review using predictive coding improves efficiency but still requires substantial training sets and human review. Neither approach fundamentally changes the human-intensive nature of document analysis.

Generative AI offers different capabilities. Rather than just finding potentially relevant documents, AI can read, understand context, extract information, identify patterns, synthesize findings, and explain reasoning. These capabilities enable new workflows that reduce human effort while maintaining or improving quality.

Non-Agentic Document Analysis Workflows

We'll start with non-agentic applications—using LLMs in structured workflows where humans direct each step. These applications provide immediate value and require less infrastructure than agentic approaches.

Initial Document Assessment

When faced with a new document collection, legal teams need quick understanding of what they're dealing with. Traditional approaches involve sampling—reviewing a few hundred documents to get a sense of the population. This is slow and may miss important patterns.

LLM-powered assessment works differently. The system can rapidly process thousands of documents, generating summaries that reveal document types present, key topics and themes, important custodians and entities, relevant time periods, and potential issues or red flags. This assessment might take hours rather than days, providing comprehensive overview without extensive human review.

The workflow is straightforward. Select a representative sample of documents—perhaps 2,000-5,000 from different custodians, time periods, and document types. Process each through an LLM with prompts designed to extract key information. Aggregate the results to identify patterns across the population. Present findings to legal team for review and strategic planning. This approach provides early visibility into what the document population contains, enabling better planning for full review. It identifies custodians who appear central to key issues, reveals time periods with high relevant document concentration, flags potential privilege issues early, and surfaces unexpected topics deserving attention.

Document Categorization and Metadata Enrichment

Beyond initial assessment, LLMs excel at categorizing documents and extracting structured information. Traditional metadata is limited—file type, date, author, recipient. LLMs can add rich semantic metadata describing document content, purpose, and significance.

The workflow processes each document through prompts designed to extract specific information. For contract collections, this might include party names, effective dates, termination dates, financial terms, key obligations, and unusual provisions. For email collections, it might include primary topic, sentiment, action items, mentions of key issues, and communication patterns.

This enriched metadata serves multiple purposes. It enables more sophisticated searching—finding not just documents containing keywords but documents about specific topics or with particular characteristics. It supports better prioritization—identifying high-value documents deserving immediate attention. It facilitates quality control—flagging documents with characteristics suggesting they warrant special handling.

Implementation requires careful prompt design. Prompts must be specific enough to generate consistent, useful metadata but flexible enough to handle diverse document types. They should include examples illustrating what good categorization looks like. They should instruct the LLM to indicate uncertainty when documents don't fit cleanly into categories.

Document Summarization at Scale

Detailed review of every document remains necessary in many contexts, but LLM-generated summaries can improve efficiency even when human review is required. Reviewers can quickly scan summaries to identify documents requiring detailed attention, focusing human effort where it adds most value.

© 2025 Merlin Search Technologies, Inc.

Effective summarization requires balancing completeness with conciseness. Too brief and important details are lost. Too verbose and efficiency gains disappear. The optimal approach often involves tiered summarization—very brief summaries for rapid scanning, moderate-length summaries providing more context, and detailed summaries for complex documents.

Summaries should preserve key information like who sent/received the communication, what the primary topic is, what key facts or assertions are made, what action items or commitments exist, and what connection exists to case issues. They should indicate when documents contain potentially privileged content, when information conflicts with other sources, when facts are particularly significant, or when uncertainty exists about interpretation.

Quality control is essential. Summaries must accurately represent document content without hallucination or mischaracterization. This requires verification sampling where human reviewers compare summaries to source documents for accuracy. Error rates should be tracked and addressed through prompt refinement when needed.

Privilege Identification

Privilege review represents one of the most challenging document analysis tasks. Missing privileged documents risks waiver of protections. Over-designating documents as privileged wastes time and may draw challenge from opposing parties. The volume of documents requiring privilege assessment makes this time-consuming and expensive.

LLMs can assist privilege identification by flagging documents with characteristics suggesting potential privilege—communications with known counsel, legal advice-seeking language, confidential markings, attorney work product indicators, and litigation context. The system doesn't make final privilege determinations, which require human judgment, but it reduces the population requiring detailed human privilege review.

The workflow processes documents through privilege-screening prompts designed to identify privilege indicators. Documents flagged by the system receive priority human review by attorneys trained in privilege assessment. Those attorneys make final privilege determinations based on the full context that AI systems may miss—attorney-client relationship existence, communication purpose, whether litigation was anticipated, and whether communications were kept confidential.

This tiered approach focuses expensive attorney time on documents most likely to be privileged while allowing lower-cost processing for clearly non-privileged documents. It improves consistency by ensuring that documents with similar characteristics receive similar treatment. It reduces risk of privilege waiver by catching potentially privileged documents that might otherwise be missed in massive document populations.

Issue Identification and Tagging

Modern litigation often involves multiple issues, each requiring focused review. Traditional approaches involve training reviewers on issue definitions and relying on their judgment to identify relevant documents. This is inconsistent—different reviewers may reach different conclusions about issue relevance.

© 2025 Merlin Search Technologies, Inc.

LLMs can improve consistency by applying uniform criteria across all documents. Prompts define each issue specifically, provide examples of responsive documents, and instruct the system to explain its reasoning. Documents are processed through these issue-identification prompts, generating consistent tagging across the population.

The key is precise issue definition. Vague issues like “documents about the merger” produce inconsistent results. Specific issues like “documents discussing due diligence findings about environmental liabilities” enable more accurate identification. Issues should be mutually exclusive where possible to avoid confusion. Examples of clearly relevant, clearly irrelevant, and borderline documents help calibrate the system.

Human oversight remains important. Sample LLM issue determinations to verify accuracy. Review borderline cases where the system expressed uncertainty. Adjust issue definitions and prompts when patterns show the system misunderstands the issue. Track accuracy over time to ensure consistent performance.

Extracting Structured Data from Unstructured Documents

Many discovery tasks involve extracting structured information from unstructured documents—identifying all financial transactions mentioned in emails, extracting dates and events for timeline construction, capturing party obligations from contracts, or building databases of factual assertions.

LLMs excel at this extraction. Unlike traditional extraction methods requiring precise formatting or position, LLMs understand meaning and can extract information regardless of presentation. They can handle variations in terminology, parse complex sentences to extract relevant details, reconcile conflicting information from different sources, and format extracted data consistently.

The workflow defines what information to extract, provides extraction templates specifying fields and formats, processes documents through extraction prompts, and compiles results into structured databases. For timeline construction, this might extract dates, events, participants, and supporting citations. For contract analysis, it might extract party names, effective dates, termination provisions, financial terms, and key obligations.

Quality control verifies that extracted information accurately reflects source documents, that formatting is consistent across extractions, that uncertainties are flagged appropriately, and that citations enable verification. Sampling protocols similar to those for document review should be applied—heavier sampling for high-stakes extractions, lighter sampling for preliminary work.

Agentic Document Analysis Workflows

Agentic approaches enable more sophisticated workflows where systems pursue investigation goals autonomously within defined boundaries. These workflows require more infrastructure but deliver greater efficiency gains for complex tasks.

Comprehensive Issue Investigation

© 2025 Merlin Search Technologies, Inc.

Traditional issue investigation requires iterative human effort. An attorney searches for documents, reviews results, refines the search based on what was found, searches again, and continues until satisfied that relevant documents have been identified. This process is time-consuming and depends heavily on the attorney's search expertise and domain knowledge.

Agentic systems can pursue investigations autonomously. Given a well-defined investigation goal, the system can plan initial search strategy, execute searches and retrieve candidate documents, analyze retrieved documents for relevance, refine search strategy based on findings, iterate until diminishing returns suggest stopping, and synthesize findings into a coherent report with citations to source documents.

Consider an investigation into whether executives knew about product defects before recall decisions. An agentic system might start by searching for documents mentioning both the product and quality/defect terms. It analyzes results to identify which executives appear in relevant documents. It searches for communications involving those executives and the product. It identifies when discussions shifted from quality concerns to recall consideration. It searches for earlier communications suggesting knowledge of problems. It builds a timeline of escalating awareness. It generates a report synthesizing findings with citations enabling verification.

This autonomous investigation compresses work that might take attorneys days or weeks into hours. The system works systematically through the document population without fatigue or distraction. It maintains consistency in how documents are evaluated. It doesn't overlook relevant documents because of search strategy limitations.

Gap Analysis and Production Assessment

When reviewing opposing parties' document productions, attorneys need to assess completeness—identifying what should have been produced but appears to be missing. This requires understanding what documents would be expected given the producing party's business operations and communications patterns, comparing that expectation against what was actually produced, and identifying gaps suggesting incomplete production or targeted withholding.

Agentic systems can systematically pursue gap analysis. Given a production and understanding of what was requested, the system can analyze the production to identify custodians represented, time periods covered, document types present, communication patterns observed, and business activities reflected. It compares this against what would be expected based on the request scope, the producing party's operations, and typical business communications. It identifies gaps—custodians missing that should be present, time periods with suspiciously low document volumes, expected document types not found, and communication patterns showing breaks suggesting missing documents.

The system generates a report detailing identified gaps, explaining why each gap is concerning, suggesting possible explanations, and recommending follow-up investigation or production challenges. This systematic approach identifies completeness issues that might be missed in less structured review.

Timeline and Chronology Construction

Many investigations require building detailed timelines showing who knew what when, how

© 2025 Merlin Search Technologies, Inc.

events unfolded, and where gaps in the record exist. Manual timeline construction is tedious—reviewing thousands of documents, extracting date references, organizing events chronologically, cross-referencing related events, and identifying inconsistencies.

Agentic systems can automate much of this work. The system searches broadly for documents containing date references related to the issue. It extracts events and their dates from each document. It organizes events chronologically, grouping related events and identifying the sequence of developments. It flags inconsistencies where different sources provide conflicting dates or descriptions. It identifies gaps between major events where documents would be expected but aren't found. It generates timeline reports in formats suitable for human review and refinement.

The resulting timeline provides comprehensive foundation for understanding event sequences. Attorneys can review the timeline, verify key events against source documents, investigate gaps or inconsistencies, and incorporate the timeline into case strategy and presentations. The system handles the mechanical work of extraction and organization, freeing attorneys to focus on interpretation and strategic implications.

Relationship Mapping and Communication Analysis

Understanding relationships between entities—who communicated with whom, about what topics, and how frequently—provides crucial context for investigations. Manual relationship mapping requires reviewing thousands of communications, tracking patterns, and building network representations. This is time-consuming and prone to missing subtle patterns.

Agentic systems can systematically analyze communication patterns. The system identifies all entities appearing in the document collection—people, organizations, and other actors. It analyzes communications to extract relationship information—who sent messages to whom, what topics were discussed, how frequently they communicated, and what tone characterized communications. It builds relationship maps showing connections, communication frequency, topic focus, and relationship evolution over time.

These relationship maps reveal patterns that might otherwise be missed. They show who served as central hubs connecting different groups. They identify when relationships intensified or diminished. They reveal topic-specific communication networks—who discussed the merger, who discussed the product defect, who discussed the accounting treatment. This understanding informs investigation strategy and witness identification.

Multi-Dimensional Document Analysis

Some investigations require analyzing documents across multiple dimensions simultaneously. A contract review might assess legal compliance, financial implications, operational risks, and strategic alignment. A due diligence review might evaluate intellectual property, regulatory compliance, litigation exposure, and business operations. Conducting these multi-dimensional analyses manually requires multiple reviewers with different expertise, making coordination complex and expensive.

Multi-agent systems can parallelize these analyses. Specialized agents focus on particular dimensions—one analyzes legal compliance, another assesses financial implications, another

© 2025 Merlin Search Technologies, Inc.

evaluates operational risks, and another considers strategic alignment. Each agent analyzes the same documents from its specialized perspective. A coordinating agent synthesizes findings across dimensions, identifying documents with issues on multiple dimensions, reconciling conflicting assessments, and generating comprehensive reports.

This approach captures efficiency benefits from parallelization while maintaining specialized focus. Each specialized agent can be optimized for its particular analysis type. The coordinating agent ensures that insights are integrated rather than siloed. The result is more comprehensive analysis completed more efficiently than sequential human review could achieve.

Quality Assurance in AI-Assisted Document Review

Regardless of whether workflows are agentic or non-agentic, quality assurance is essential. AI systems must meet reliability standards appropriate for legal work.

Verification Sampling Protocols

The foundation of quality assurance is systematic verification through sampling. For high-stakes document review destined for production, verification should be rigorous—sampling 30-50% of AI-reviewed documents initially, reducing sampling as consistent accuracy is demonstrated. For lower-stakes preliminary review, lighter sampling of 10-20% may suffice. For new document types or issues where AI performance is unknown, begin with 100% human verification until accuracy is established.

Sampling should be stratified to ensure coverage across different document types, custodians, time periods, and relevance levels. Random sampling alone may miss concentrated errors. Stratified sampling ensures that errors affecting particular categories are detected.

Error Pattern Analysis

When verification identifies errors, analyze patterns to guide improvements. Are errors concentrated in particular document types, suggesting prompts need refinement for those types? Do errors cluster around specific custodians or time periods, suggesting those areas need enhanced scrutiny? Are certain types of analysis particularly error-prone, indicating where additional verification is needed?

Pattern analysis enables targeted responses rather than across-the-board measures that may be inefficient. If errors occur primarily with technical documents, enhance prompts and verification for technical content while maintaining standard approaches for other document types. If a particular custodian's communications consistently cause problems, flag that custodian for enhanced human review.

Continuous Monitoring

Quality assurance isn't just an initial validation exercise—it requires ongoing monitoring throughout document review. Track accuracy metrics over time to detect degradation. Monitor error rates by document type and analysis type. Assess whether AI and human reviewers show consistent judgment on relevance and issue identification. Investigate anomalies like

© 2025 Merlin Search Technologies, Inc.

sudden changes in error rates or unexpected patterns.

Continuous monitoring catches problems before they compound. If accuracy degrades, investigation can identify causes—perhaps document characteristics changed as review progressed into different time periods, or system updates affected performance. Early detection enables corrective action while most documents remain unreviewed.

Human-AI Disagreement Analysis

When human reviewers disagree with AI assessments, investigate why. Sometimes humans are right and AI made errors requiring correction. Sometimes AI identified relevant information humans missed, suggesting human error. Sometimes both perspectives have merit, indicating ambiguous documents deserving careful judgment.

This disagreement analysis reveals improvement opportunities. If AI consistently flags documents humans consider irrelevant, prompts may need refinement to better match human judgment. If humans consistently miss information AI identifies, training may need enhancement to help reviewers recognize relevant content. If disagreements cluster around genuinely ambiguous documents, escalation protocols can ensure those documents receive senior attorney review.

Cost-Effectiveness and ROI Considerations

AI-assisted document review must demonstrate cost-effectiveness to justify adoption. Understanding the economics helps make informed decisions about when and how to deploy these technologies.

Direct Cost Analysis

Direct costs include LLM API charges, infrastructure and platform costs, storage for documents and analysis results, and vendor fees for commercial AI review platforms. These costs are generally proportionate to document volume and analysis complexity. A simple relevance review might cost $0.10-0.50 per document in API charges. More complex multi-issue analysis might cost $1-3 per document. These costs compare favorably to traditional document review costs of $50-200 per document in attorney time.

However, direct AI costs don’t tell the complete story. Human oversight, verification sampling, error correction, and quality assurance add costs that must be factored into ROI calculations. A heavily verified AI review might cost $5-15 per document all-in, still substantial savings compared to traditional review but less dramatic than comparing API costs alone.

Efficiency Gains

Beyond direct cost savings, AI review offers efficiency benefits. Traditional review might take months for large document populations. AI-assisted review might compress this to weeks. This timeline compression has value beyond cost—it enables earlier case assessment, faster strategic decisions, and more flexible litigation posture.

© 2025 Merlin Search Technologies, Inc.

AI review also improves consistency. Human review teams inevitably show variation in judgment. AI systems apply consistent criteria across all documents, reducing variation and improving predictability of review outcomes. This consistency reduces disputes about review methodology and improves defensibility.

Strategic Value

Some benefits resist quantification but provide strategic value. Comprehensive early case assessment enables better settlement decisions. Identifying key documents quickly improves trial preparation. Understanding communication patterns reveals case dynamics informing strategy. These strategic benefits may dwarf cost savings for high-stakes matters.

The ROI calculation should consider total value delivered, not just cost reduction. A system costing $100,000 but enabling $2 million better settlement through improved early assessment shows strong ROI even without considering document review cost savings.

Integration with Existing Discovery Workflows

AI document analysis doesn't replace existing discovery infrastructure—it augments and integrates with it. Understanding integration points helps ensure smooth adoption.

Platform Integration

Most organizations have existing document review platforms housing document collections, supporting keyword search, enabling reviewer coding, tracking production, and managing workflows. AI analysis must integrate with these platforms rather than requiring migration to new systems.

Integration typically occurs through APIs enabling the AI system to retrieve documents from the review platform, process them through AI analysis, write results back to the platform as metadata or coding, and enable reviewers to see AI analysis alongside traditional search and review tools. This integration allows organizations to adopt AI capabilities while preserving existing investments in review platforms and workflows.

Workflow Coordination

AI analysis should fit naturally into established workflows rather than requiring wholesale process redesign. In traditional workflows, document populations flow through processing and deduplication, then keyword search and culling, then privilege review and quality control, and finally production. AI can enhance each stage—assisting with early case assessment after processing, improving search and culling efficiency, supporting privilege identification, and enabling production quality verification.

The key is identifying where AI adds most value for each matter and integrating AI support at those points while maintaining established workflows elsewhere. Not every matter benefits from AI deployment at every stage. Selective deployment where value is clear enables adoption without disruption.

© 2025 Merlin Search Technologies, Inc.

Skills and Training

Successful integration requires developing organizational capabilities. Reviewers need training on how to work with AI-generated analysis—understanding what it means when AI flags documents as relevant, how to interpret confidence levels, when to trust AI assessment versus conducting independent review, and how to provide feedback when AI analysis seems incorrect.

Case managers need skills in deploying AI appropriately—determining which matters benefit from AI support, configuring AI workflows for specific needs, monitoring AI performance during review, and troubleshooting when issues arise. These skills develop through experience and training, starting with pilot projects and expanding as confidence grows.

Looking Ahead

This chapter examined document analysis and review workflows using both non-agentic and agentic AI approaches. From initial assessment through production review, AI technologies offer opportunities to improve efficiency, enhance accuracy, and deliver strategic value. The workflows presented here demonstrate practical applications that organizations are already implementing successfully.

But document review represents only one aspect of discovery work. Chapter 11 turns to transcript analysis—depositions, hearings, witness interviews, and other recorded testimony. These materials present different challenges than documents, requiring different AI approaches while building on the principles established here. The techniques for analyzing, synthesizing, and extracting insights from testimony recordings show how AI capabilities extend across diverse discovery workflows.

© 2025 Merlin Search Technologies, Inc.

MERLIN

11

Transcript and Testimony Analysis

© 2025 Merlin Search Technologies, Inc.

MERLIN

CHAPTER ELEVEN

Transcript and Testimony Analysis

Depositions, hearings, trials, and witness interviews generate thousands of pages of testimony in complex litigation. Traditional approaches to managing this testimony rely on human-created summaries—associates or paralegals reading transcripts and distilling key points into memos or summary databases. This process is expensive, time-consuming, and variable in quality depending on who performs the work.

Generative AI transforms transcript analysis by reading and understanding testimony at scale, extracting key points systematically, identifying patterns across multiple witnesses, and enabling sophisticated queries about what testimony contains. Unlike document analysis where relevant information might be buried in thousands of files, transcript analysis requires processing long-form narrative content where context and nuance matter significantly.

This chapter examines how AI enhances transcript workflows—from basic summarization through sophisticated cross-examination preparation, from single-transcript analysis through synthesis across dozens of witnesses, from passive summaries through active investigation of testimony patterns. The approaches described here work with deposition transcripts, hearing and trial transcripts, witness interview notes, recorded statements, and other long-form testimonial content.

The Transcript Analysis Challenge

Testimony presents distinct challenges compared to documents. Transcripts are long—a single deposition might run 200-400 pages. They're conversational rather than structured—witnesses ramble, backtrack, clarify, and contradict themselves. They're dense with context—understanding a statement often requires knowing what was asked, what the witness said earlier, and what other witnesses said about the same topic.

Traditional transcript summaries address some challenges while creating others. A well-prepared summary distills key testimony efficiently, but creating that summary requires someone to read the entire transcript carefully. Quality varies with the summarizer's skill and attention. The summary reflects what the summarizer considered important, potentially missing information others might prioritize. And summaries become outdated as cases evolve—what seemed unimportant when the summary was created might become critical later.

These challenges multiply with volume. A case with thirty depositions generates thousands of transcript pages. Finding what Witness A said about Topic X requires either reading all of Witness A's testimony or hoping the summarizer flagged that topic. Identifying where Witness A and Witness B gave conflicting testimony requires either remarkable memory or painstaking cross-referencing. Understanding how testimony evolved across multiple deposition days demands synthesis that's difficult even for experienced attorneys.

AI doesn't eliminate the need for attorneys to understand testimony, but it fundamentally

changes how that understanding is built. Instead of relying entirely on linear reading or static summaries, attorneys can query transcripts dynamically, extract information systematically, and synthesize testimony at scale.

Understanding Transcript Structure

Before examining AI applications, it's important to understand how transcripts work technically and why this matters for AI analysis.

Legal transcripts follow consistent structures—page and line numbering, speaker identification, question-and-answer format. This structure is both an asset and challenge for AI analysis. The structured format enables precise citation—every statement can be referenced by page and line number. But the conversational nature means that understanding often requires reading beyond individual answers to grasp context.

Most modern transcripts exist as searchable text, either as native-born digital files or as OCR conversions from scanned images. This text serves as input for AI analysis. However, transcript quality varies. Some include timestamps, speaker labels, and formatting that helps AI understand structure. Others are plain text requiring AI to infer structure from content. Error rates from OCR or voice-to-text transcription can introduce inaccuracies that AI analysis must navigate.

For AI applications, transcripts typically are processed as single long documents rather than broken into arbitrary chunks. This preserves conversational flow and context. LLM context windows that might handle 200,000 tokens can process substantial testimony—roughly 150,000 words or 300-400 pages of transcript. This capacity means that many depositions fit entirely within single context windows, enabling comprehensive analysis without losing context across chunks.

When testimony exceeds context window capacity—multi-day depositions or trial transcripts running hundreds of pages—strategic chunking becomes necessary. The goal is preserving logical units (entire examination sequences, topical discussions, witness sessions) rather than arbitrary page breaks. Effective chunking maintains context while enabling systematic processing across lengthy testimony.

Non-Agentic Transcript Applications

We'll start with non-agentic applications where humans direct specific analytical tasks and AI executes them. These applications provide immediate value and require less infrastructure than agentic approaches.

Comprehensive Transcript Summarization

The most straightforward AI application is generating comprehensive summaries of individual transcripts. Rather than human summarizers reading and distilling testimony, AI can process entire depositions and produce structured summaries consistently and efficiently.

Effective summarization requires careful prompt design balancing detail with readability. Too brief and important nuances are lost. Too detailed and the summary becomes nearly as

© 2025 Merlin Search Technologies, Inc.

long as the original transcript. The optimal approach often employs tiered summarization—a high-level overview for quick orientation, section-by-section summaries providing moderate detail, and topic-focused summaries for key issues.

A well-structured transcript summary includes several components. The executive summary provides case context—who the witness is, their role in events, their relationship to parties, and the key topics covered. The chronological summary follows testimony flow, capturing major points in the order discussed. Topical summaries organize testimony by issue regardless of order discussed—what the witness said about Product Development, what they said about the Meeting, what they said about the Decision.

Citations are essential. Every substantive point in the summary should reference the specific transcript location (page and line numbers) supporting it. This enables rapid verification—attorneys can immediately check whether the summary accurately represents testimony without searching through hundreds of pages. Citations also support using summaries in motion practice or trial preparation where specific transcript references are required.

The summarization workflow is straightforward. The transcript is processed through prompts designed to generate the summary structure. For transcripts within context window capacity, this might be a single processing pass generating the complete summary. For longer transcripts, section-by-section processing followed by synthesis of section summaries produces the final output. The resulting summary undergoes verification sampling where attorneys compare summary statements to source testimony for accuracy.

Quality control focuses on three dimensions—accuracy (does the summary correctly represent what was said), completeness (are significant points included), and utility (is the summary actually helpful for case work). Accuracy checking samples summary statements and verifies them against transcripts. Completeness assessment compares summaries to human-generated summaries or to attorney knowledge of key issues to identify gaps. Utility evaluation solicits feedback from attorneys using summaries about whether they find them valuable.

Topic-Specific Extraction

Beyond comprehensive summaries, AI excels at extracting testimony about specific topics across entire transcripts. An attorney might need everything Witness A said about the Board Meeting, or all testimony about Product Testing, or any mention of Person X. Traditional approaches require either reading the entire transcript or relying on keyword search that misses relevant testimony phrased differently than expected.

AI can process transcripts with topic-focused prompts that identify relevant testimony regardless of specific wording. The prompt describes the topic of interest, provides context about why it matters, and instructs the system to extract any testimony touching on that topic even tangentially. The system processes the transcript and returns relevant excerpts with citations.

This approach works particularly well for topics that might be discussed using varied terminology. A prompt seeking testimony about “financial projections” will capture statements about “revenue forecasts,” “earnings estimates,” “budget assumptions,” and other related concepts that keyword search might miss. The AI understands semantic relationships between concepts rather than just matching words.

© 2025 Merlin Search Technologies, Inc.

The extraction workflow produces structured outputs organizing relevant testimony logically. For each topic, the output includes direct quotes from testimony (with page and line citations), context explaining the setting (what question elicited this response), and brief explanatory notes (why this testimony matters for the topic). This structure enables attorneys to quickly grasp what was said about topics of interest.

Verification focuses on both recall (did the system find all relevant testimony) and precision (is what it found actually relevant). Recall testing compares AI extraction against human review of the same transcript for the same topic. Precision testing evaluates whether flagged testimony actually relates to the topic. Both metrics inform prompt refinement to improve extraction quality.

Witness Comparison and Contradiction Analysis

One of the most valuable but tedious transcript tasks is comparing what different witnesses said about the same events or topics. Manual comparison requires reading multiple transcripts, finding relevant sections, and identifying where testimony aligns or conflicts. This process is time-consuming and error-prone.

AI can systematically compare testimony across witnesses. Given transcripts from multiple witnesses and a specific topic or event, the system can extract what each witness said, identify areas of agreement and disagreement, flag direct contradictions, note differences in detail or emphasis, and synthesize comparative analysis.

The comparison workflow starts by processing each transcript to extract testimony about the specified topic. The system then compares extracted testimony across witnesses, identifying statements that confirm each other, statements that contradict each other, statements that add details others omitted, and statements that differ in emphasis or interpretation. The output organizes these comparisons clearly, showing what each witness said alongside each other for easy review.

This analysis proves particularly valuable for impeachment preparation. If Witness A testifies that the meeting occurred in March but Witness B says April, that contradiction deserves attention. If Witness A describes a decision process as careful and deliberate while Witness B describes it as rushed and pressured, that difference in characterization matters. AI-assisted comparison surfaces these discrepancies systematically.

Quality assurance for comparison analysis requires verifying that identified agreements and contradictions are real rather than artifacts of misunderstanding testimony. Sampling protocols check significant contradictions against source transcripts to ensure the AI correctly interpreted what witnesses said. This verification is essential because contradictions drive important strategic decisions about witness credibility and case theory.

Question and Answer Indexing

Traditional transcript summaries organize by topic or chronology, but sometimes attorneys need different access patterns. Q&A indexing creates searchable databases where specific questions and answers can be retrieved individually along with surrounding context.

© 2025 Merlin Search Technologies, Inc.

This approach processes transcripts to identify each question-answer pair, extract both question and answer, generate brief descriptions of what each pair discusses, tag pairs with relevant topics or issues, and create citations enabling instant location in source transcripts. The resulting index enables searching by question type (e.g., all questions about documents), by topic (all Q&A about financial performance), or by witness response patterns (all questions the witness answered evasively).

The value becomes clear in deposition preparation and cross-examination. When preparing to depose Witness B, attorneys can search the Q&A index for all questions asked of Witness A about Topic X, seeing both what was asked and how Witness A responded. This reveals productive question formulations and suggests what to ask Witness B. During cross-examination, attorneys can quickly find prior testimony to impeach inconsistent testimony.

Implementation requires careful handling of conversational flow. Some testimony can't be isolated into clean question-answer pairs—witnesses give multi-part answers, attorneys ask follow-up questions, objections interrupt, and colloquy develops. The indexing system must handle these complexities while maintaining utility. Effective implementations group related exchanges into logical units rather than forcing artificial Q&A pair boundaries.

Timeline Construction from Testimony

Witnesses often testify about events over extended time periods, mentioning dates and temporal sequences scattered throughout depositions. Extracting these temporal references and organizing them chronologically creates valuable timelines showing when events occurred according to testimony.

AI can process transcripts to identify any statement mentioning dates, times, or temporal relationships (before, after, during). It extracts these references along with what the witness said happened at those times. It organizes extracted information chronologically. It identifies temporal contradictions or uncertainties. It generates timeline outputs showing event sequences according to witness testimony.

This timeline construction differs from document-based timelines because it reflects witness recollection rather than contemporaneous documentation. Comparing testimony-based timelines to document-based timelines often reveals discrepancies—witnesses misremembering dates, telescoping events closer together, or inverting sequences. These discrepancies themselves provide valuable impeachment material or suggest where memory failures affect testimony reliability.

The extraction process must handle temporal ambiguity carefully. Witnesses rarely provide precise dates for all events. They say "sometime in early 2023" or "a few weeks after the meeting" or "I think it was March, maybe April." The timeline system should preserve this uncertainty rather than forcing false precision. Outputs might show events on approximate timelines with confidence indicators reflecting how precisely the witness dated events.

Key Document Identification

Depositions often reference documents—emails shown to witnesses, contracts discussed, reports reviewed. Tracking which documents were discussed and what witnesses said about

© 2025 Merlin Search Technologies, Inc.

them creates valuable connections between testimony and documentary evidence.

AI can process transcripts to identify when documents are referenced (by exhibit number, description, or content), extract what the witness said about each document, capture whether the witness authenticated, explained, or disputed the document, and create indexed references connecting documents to testimony about them.

This document-testimony linking serves multiple purposes. It helps attorneys understand which documents witnesses recognize and what explanations they provide. It reveals which documents witnesses dispute or explain away, suggesting potential challenges. It enables finding testimony when you have the document (what did witnesses say about Email 247?) or finding documents when you have the testimony (the witness mentioned a report about safety concerns—which exhibit was that?).

Implementation challenges include handling situations where documents are referenced ambiguously (the witness mentions “that email” without exhibit numbers) or where transcript references are incomplete (exhibit markings noted but document descriptions unclear). Effective systems combine AI processing with human verification to ensure document-testimony links are accurate and complete.

Agentic Transcript Applications

Agentic systems enable more sophisticated testimony analysis where AI autonomously pursues investigation goals across transcript collections. These applications demonstrate how agentic capabilities transform testimony analysis from reactive (answer specific questions about transcripts) to proactive (investigate testimony systematically to uncover patterns and insights).

Cross-Examination Preparation Investigations

Traditional cross-examination preparation involves attorneys reading deposition transcripts, identifying inconsistencies and vulnerabilities, researching what other witnesses said about the same topics, and organizing material for effective cross. This process is labor-intensive and requires exceptional memory to track testimony patterns across multiple witnesses.

Agentic systems can pursue cross-examination preparation autonomously. Given a witness and upcoming testimony, the system can identify key topics the witness testified about, compare that testimony to what other witnesses said about those topics, search for inconsistencies between the witness’s testimony and documentary evidence, identify areas where the witness’s testimony seems implausible or unsupported, research what questions proved effective with similar witnesses, and compile comprehensive cross-examination preparation materials.

Consider preparing to cross-examine an executive who testified in deposition about decision-making around a product discontinuation. An agentic system might start by extracting everything the executive said about the decision. It would compare that testimony to what other executives said about the same decision, identifying contradictions or differences in emphasis. It would search documents contemporaneous with the decision for evidence confirming or undermining the executive’s account. It would analyze whether the executive’s testimony about the decision timeline aligns with email exchanges and meeting minutes. It would identify documents the executive should have known about but claimed not to recall. It would generate a

© 2025 Merlin Search Technologies, Inc.

structured report organizing all this material for the cross-examining attorney.

This autonomous investigation compresses work that might take an attorney days into hours. The system works systematically through the investigation without fatigue, identifies patterns a human might miss across large testimony collections, maintains consistency in how testimony is analyzed, and produces comprehensive results with citations enabling verification.

Testimony Gap Analysis

In complex litigation with many witnesses, gaps in testimony—topics that should have been covered but weren't—can be as important as what was said. Identifying these gaps requires understanding what testimony exists, what topics matter for the case, and what's missing. Agentic systems can systematically pursue gap analysis. Given a set of depositions and case issues, the system can catalog what topics each witness addressed, identify topics that weren't covered with particular witnesses, recognize patterns suggesting coordinated testimony or coached evasions, note areas where all witnesses claimed not to recall information, and generate reports highlighting testimony gaps deserving attention.

This gap analysis informs deposition strategy going forward—what to ask witnesses who haven't been deposed yet, what to pursue in supplemental depositions, what gaps to highlight in expert reports or motion practice. It also surfaces potential discovery deficiencies—if no witness can explain key decisions despite documentary evidence showing extensive deliberation, that suggests incomplete testimony deserving further investigation.

The investigation workflow operates autonomously within defined boundaries. The system doesn't decide what gaps matter—humans establish investigation goals and identify what topics are important. But within those parameters, the system systematically analyzes testimony coverage, identifies omissions, and reports findings for human evaluation.

Credibility and Consistency Assessment

Witness credibility often hinges on consistency—do they tell the same story repeatedly, or does their account shift? Do their descriptions align with other evidence, or do discrepancies suggest problems? Manual credibility assessment requires comparing testimony across multiple sessions, tracking evolving accounts, and maintaining awareness of corroborating or contradicting evidence.

Agentic systems can pursue credibility assessment systematically. Given testimony from a witness across multiple sessions (deposition, hearing, trial), the system can identify where accounts of the same events differ between sessions, analyze whether differences represent clarification or substantive change, compare testimony to documentary evidence for consistency, evaluate whether claimed lack of recollection seems selective or comprehensive, and assess whether testimony patterns suggest coaching or evasion.

The credibility assessment doesn't reach ultimate conclusions about truthfulness—that remains human judgment. But it systematically organizes information bearing on credibility, highlights inconsistencies deserving attention, identifies patterns suggesting strengths or vulnerabilities, and provides factual foundation for credibility arguments.

© 2025 Merlin Search Technologies, Inc.

For example, if a witness testified in deposition that they learned about Quality Issues from Report X in March, but later at hearing testified they learned about Quality Issues from Meeting Y in January, that's a significant shift. If the witness claimed not to recall specific emails in deposition but at hearing provided detailed recollections of those same emails, that pattern deserves scrutiny. Agentic systems surface these patterns systematically.

Multi-Witness Narrative Construction

Complex cases require understanding how multiple witnesses' accounts combine into coherent narrative of events. What's the overall story that emerges from all testimony? Where do witnesses agree, suggesting reliable facts? Where do they disagree, suggesting disputed issues?

Agentic systems can construct multi-witness narratives autonomously. Given depositions from all witnesses, the system can identify the key events and decisions that testimony addresses, synthesize what each witness said about those events, identify areas of agreement across witnesses, note disputed facts where witnesses conflict, construct chronological narrative of events based on testimony, and generate comprehensive reports with citations to supporting testimony.

This narrative construction requires sophisticated synthesis. The system must understand that different witnesses describing "the meeting" are referring to the same event even if they use different terminology. It must recognize when witness accounts complement each other (one witness describes the decision, another describes the implementation) versus when they conflict (one says the decision was made in March, another says May). It must weigh testimony differently based on witness position and knowledge (the CFO's testimony about financial analysis deserves more weight than a junior accountant's).

The resulting narrative provides framework for understanding the case. Attorneys can see what story emerges from all testimony, identify which facts seem established by multiple witnesses, recognize which issues remain disputed, and assess whether overall testimony supports or undermines case theory.

Expert Testimony Analysis

Expert witnesses present unique analytical challenges. Their testimony often contains technical content requiring careful analysis. Their opinions must be evaluated for consistency with disclosed reports. Their methodology deserves scrutiny. Comparing competing experts requires understanding technical differences.

Agentic systems can analyze expert testimony systematically. Given expert deposition transcripts and disclosed reports, the system can extract opinions expressed in testimony, compare testimony to disclosed reports for consistency, identify bases and reasoning for opinions, flag methodology issues or unsupported assumptions, compare competing experts' opinions on the same topics, and generate analysis highlighting strengths and vulnerabilities.

For example, analyzing a damages expert's testimony, the system might extract all assumptions underlying the damages calculation, compare those assumptions to what was disclosed in the expert report, identify where opposing counsel challenged assumptions in deposition,

© 2025 Merlin Search Technologies, Inc.

note where the expert acknowledged limitations or uncertainties, research whether the expert's methodology has been accepted in other cases, and compile material for Daubert challenges or cross-examination.

This analysis is particularly valuable because expert testimony often runs hundreds of pages covering complex technical topics. Human analysis risks missing important details or failing to recognize methodological issues. Systematic AI analysis ensures comprehensive coverage while organizing technical content accessibly for attorneys preparing to challenge or defend expert opinions.

Quality Assurance for Transcript Analysis

Transcript analysis requires rigorous quality assurance because testimony analysis directly affects case strategy, motion practice, and trial preparation. Errors or omissions can lead to missing impeachment opportunities, making unsupported arguments, or pursuing unproductive strategies.

Accuracy Verification

The foundation of transcript QA is verifying that AI analysis accurately represents what testimony contains. This requires systematic sampling where human reviewers compare AI outputs to source transcripts.

For summaries, verification samples summary statements and checks them against cited testimony. Does the summary accurately capture what the witness said? Are citations correct? Does the summary omit important context that changes meaning? Are quoted portions accurate? Error rates should be tracked by statement type—are certain kinds of testimony (technical content, temporal references, opinion statements) more error-prone?

For extractions and comparisons, verification checks both recall and precision. Did the system find all relevant testimony (recall)? Is flagged testimony actually relevant (precision)? For contradiction analysis, verification confirms that identified conflicts are real rather than artifacts of misunderstanding nuanced testimony.

Verification intensity should match stakes. High-stakes testimony analysis destined for motion practice or cross-examination deserves 100% verification of key points. Lower-stakes preliminary analysis might use lighter sampling. New testimony types or technical content warrant heavier verification until accuracy is established.

Citation Validation

Citations are essential for transcript analysis—they enable verification and support using analysis in legal work. Citation validation ensures that every substantive statement in analysis includes correct transcript references.

Validation checks that cited page and line numbers exist in the transcript, that cited text actually appears at specified locations, that citations accurately represent what testimony says (not cherry-picking out of context), and that citation density is sufficient (all substantive claims

© 2025 Merlin Search Technologies, Inc.

supported, not just selected statements).

Automated validation can check citation existence and text matching. Human review is needed for context assessment—does the citation provide fair representation of testimony or does broader context change meaning? This is particularly important for impeachment material where opposing counsel will scrutinize context carefully.

Completeness Assessment

Beyond accuracy, transcript analysis should be complete—covering significant testimony rather than missing important material. Completeness assessment compares AI analysis to human expectations about what should be included.

Approaches include comparing AI summaries to human-generated summaries to identify gaps, having attorneys review transcripts after reading AI analysis to note omissions, tracking whether case-relevant topics received appropriate attention, and soliciting feedback from users about whether analysis covered what they needed.

Completeness failures often reflect prompt deficiencies rather than fundamental AI limitations. If analysis consistently misses certain testimony types, prompts can be refined to emphasize those areas. Iterative improvement based on completeness feedback enhances analysis quality over time.

Consistency Checking

For multi-transcript analysis, consistency matters—does the system analyze similar testimony consistently across different witnesses? Inconsistent treatment might mean Witness A's testimony about Topic X receives detailed attention while Witness B's similar testimony is glossed over.

Consistency checking involves comparing analysis across transcripts, evaluating whether similar testimony receives similar treatment, assessing whether contradictions are identified symmetrically (if A contradicts B, is that noted in analysis of both transcripts?), and verifying that importance assessments align with case strategy rather than arbitrary variation.

Inconsistency often stems from processing transcripts separately without cross-reference. Multi-transcript agentic systems that analyze testimony collections holistically show better consistency than transcript-by-transcript processing. When transcripts must be processed separately, post-processing synthesis can improve consistency.

Integration with Discovery Workflows

Transcript analysis doesn't stand alone—it integrates with broader discovery workflows including case assessment, motion practice, trial preparation, and ongoing investigation strategy.
Case Assessment Integration

Early case assessment requires understanding both documents and testimony. Effective integration ensures that transcript analysis informs strategic decisions alongside document analysis.

© 2025 Merlin Search Technologies, Inc.

Integration approaches include using transcript timelines alongside document timelines to identify temporal discrepancies, connecting testimony about documents to document analysis to understand witness perspectives, incorporating witness credibility assessments into overall case evaluation, and synthesizing testimony-based factual narrative with document-based evidence into comprehensive case understanding.

This integration reveals relationships between evidence types. When witness testimony aligns with documentary evidence, that strengthens factual positions. When testimony and documents conflict, that suggests issues deserving investigation or targets for impeachment.

Deposition Strategy Development

Transcript analysis from completed depositions informs strategy for upcoming depositions. Understanding what previous witnesses said, what questions proved effective, what topics remain unexplored, and what contradictions exist guides preparation.

Integration includes using prior testimony analysis to identify what to ask new witnesses, developing questions based on what earlier witnesses said, preparing impeachment material based on contradiction analysis, and structuring deposition flow based on testimony gap analysis showing what needs coverage.

This integration enables systematic improvement across deposition sequences. Later depositions benefit from lessons learned in earlier ones. The investigation becomes increasingly focused as testimony analysis reveals which theories are supported and which deserve skepticism.

Motion Practice Support

Testimony analysis supports motion practice by enabling rapid identification of testimony supporting factual assertions, systematic citation to testimony in briefs and statements of fact, impeachment through identified contradictions, and credibility arguments based on consistency analysis.

Integration requires outputs formatted for legal writing—properly formatted citations, quotable excerpts with context, comparative analysis showing patterns, and synthesis suitable for incorporating into briefs.

Effective integration also includes verification workflows ensuring that testimony cited in motions is accurately represented and comprehensively supported. The last thing a practitioner wants is citing testimony in a brief only to have opposing counsel demonstrate the citation misrepresents what the witness actually said.

Trial Preparation

Trial preparation demands comprehensive command of testimony across all witnesses. Integration here includes organizing cross-examination material by witness and topic, preparing impeachment with precise transcript citations, developing examination outlines based on testimony analysis, and creating exhibits showing testimony comparisons or contradictions.

© 2025 Merlin Search Technologies, Inc.

The integration challenge is making comprehensive testimony analysis accessible during trial when rapid retrieval is essential. Solutions include searchable testimony databases enabling instant retrieval, organized cross-examination outlines with embedded citations, impeachment books with AI-generated comparison analyses, and real-time trial support where AI helps identify relevant testimony as issues arise during trial.

Cost-Effectiveness and ROI

Transcript analysis using AI must demonstrate cost-effectiveness to justify adoption. Understanding economics helps make informed decisions about when and how to deploy these capabilities.

Direct Cost Analysis

Direct costs for AI transcript analysis include LLM API charges proportionate to transcript length and analysis complexity, platform and infrastructure costs for systems managing transcript processing, storage for transcripts and analysis outputs, and vendor fees for commercial transcript analysis platforms.

A typical deposition transcript of 200 pages might cost $100 for a comprehensive summarization, with reduced prices for higher volumes. In contrast, manual transcript summarization might cost $1,500-3,500 per transcript depending on whether performed by senior paralegals or junior associates. Sophisticated comparative analysis or cross-examination preparation might cost $10,000-15,000 in attorney time. AI analysis at $100 per transcript shows dramatic cost advantages.

However, complete cost accounting includes human review time for verification, correction time for errors, training time for staff learning to use systems, and system administration and maintenance. Even with these overhead costs, AI analysis typically shows an even stronger ROI for transcript-heavy litigation.

Efficiency Gains

Beyond direct cost savings, AI transcript analysis offers efficiency benefits that may exceed cost savings in value. Traditional transcript analysis might take days or weeks for complex multi-witness matters. AI analysis might compress this to hours or overnight.

Timeline compression enables earlier strategic decisions based on comprehensive testimony understanding, faster motion practice with systematic testimony support, more thorough cross-examination preparation, and flexible adaptation as cases develop.

These efficiency gains have value beyond cost reduction. Earlier strategic decisions might enable better settlement outcomes. More thorough cross-examination might strengthen the trial position. Comprehensive testimony analysis might reveal case strengths or weaknesses that inform resource allocation.

Strategic Value

© 2025 Merlin Search Technologies, Inc.

Some benefits resist quantification but provide strategic value. Systematic contradiction analysis might reveal impeachment opportunities that would be missed in traditional review. Comprehensive multi-witness synthesis might identify fact patterns supporting summary judgment. Timeline analysis might reveal witness memory failures affecting credibility.

For high-stakes matters where case outcome depends significantly on testimony, these strategic benefits may dwarf cost considerations. An automated system to analyze thirty depositions will show exceptional ROI if it reveals contradictions enabling effective summary judgment motion or identifies vulnerabilities enabling favorable settlement.

Looking Ahead

This chapter examined how AI transforms transcript and testimony analysis—from basic summarization through sophisticated agentic investigations of testimony patterns. The applications described here demonstrate how legal professionals can leverage AI to understand testimony more comprehensively, prepare more effectively, and identify insights that manual analysis might miss.

Chapter 12 builds on these capabilities to examine timeline and chronology construction—one of the most valuable but tedious discovery tasks. While this chapter showed timeline construction from testimony, Chapter 12 explores comprehensive timeline development combining testimony, documents, and other evidence sources into unified chronologies showing how events unfolded. The techniques for systematic temporal analysis and gap identification demonstrate how AI enables discovery professionals to develop factual understanding that would be impractical through traditional methods.

© 2025 Merlin Search Technologies, Inc.

MERLIN

12

Timeline and Chronology Construction

© 2025 Merlin Search Technologies, Inc.

CHAPTER TWELVE

Timeline and Chronology Construction

Understanding when events occurred, in what sequence, and with what temporal relationships stands at the center of most litigation and investigations. Did the executive know about the defect before approving the product launch? When did the company first learn about regulatory violations? What sequence of communications led to the disputed contract terms? These questions demand comprehensive chronologies showing not just what happened but when.

Traditional timeline construction is tedious. An attorney or paralegal reviews thousands of documents and testimony, extracts date references, organizes events chronologically, cross-checks conflicting dates, identifies gaps, and synthesizes findings into coherent timeline formats. This manual process consumes days or weeks, proves error-prone as references are missed or misinterpreted, and becomes quickly outdated as new evidence emerges.

Generative AI transforms timeline construction by systematically extracting temporal information from all available sources, organizing events chronologically while preserving uncertainty, identifying temporal contradictions deserving investigation, recognizing gaps where documents would be expected but don't appear, and maintaining dynamic timelines that update as new evidence emerges. This chapter examines how AI enables discovery professionals to build comprehensive chronologies that would be impractical through manual methods.

The Timeline Challenge

Temporal information in discovery materials comes in many forms, each presenting distinct challenges. Documents contain explicit dates—email headers, meeting invitations, contract effective dates. They also contain temporal references within content—"we discussed this last month," "the meeting scheduled for next Tuesday," "approximately two weeks after the incident." Testimony includes witness recollections of when events occurred, often imprecise or contradictory.

These temporal references are scattered across thousands of sources. A significant litigation matter might involve emails spanning years, contracts with various effective dates, meeting minutes from dozens of sessions, testimony from multiple witnesses recalling overlapping events, and documents creating chains of temporal dependencies. Understanding the complete chronology requires synthesizing all these sources.

Temporal information also arrives in varying precision levels. Some events have exact dates and times—an email sent at 2:47 PM on March 15, 2023. Others are approximate—"sometime in early 2023" or "a few weeks after the meeting." Still others are relative—"before we learned about the problem" or "after the Board approved the plan." Effective timelines must handle this precision variation without forcing false exactness.

Temporal contradictions frequently appear. Document A suggests Event X occurred in March

while Document B implies it happened in April. Witness testimony conflicts with documentary evidence about timing. Documents dated April reference meetings that witnesses recall occurring in May. These contradictions might reflect errors, different perspectives, or deliberate deception, but all deserve investigation.

Timeline construction serves multiple purposes beyond understanding event sequences. Timelines reveal patterns—when activity intensified, when communications ceased, when decisions clustered. They identify gaps suggesting missing documents or unexplored time periods. They enable comparison between what participants knew at particular times versus what they claim to have known. They provide frameworks organizing evidence for presentation to courts, clients, or fact finders.

Technical Foundations for Timeline Construction

Before examining specific applications, understanding how AI processes temporal information helps explain both capabilities and limitations.

Temporal Information Extraction

AI systems extract temporal information through several mechanisms. Explicit dates are straightforward—email headers, document metadata, contract dates. The system identifies these structured temporal markers and associates them with the documents or events they describe.

Temporal references within text require more sophisticated processing. The system must understand that "last Tuesday" refers to a specific date relative to the document's date, that "two weeks after the meeting" creates a temporal relationship between events, and that "early 2023" specifies an approximate timeframe. This understanding draws on the LLM's training in natural language, enabling it to interpret varied temporal expressions.

The extraction process produces structured temporal data from unstructured sources. Each extracted temporal reference includes the date or timeframe (with precision indicators), the event or content associated with that time, the source (document, testimony, or other evidence), and confidence levels reflecting certainty about the temporal information.

This structured extraction enables chronological organization while preserving important context. The system knows not just that Email A was sent on March 15 but that it discussed Product Launch and was sent from Executive to Board. It understands not only that the witness testified about a meeting but that the witness said it occurred "probably in March, maybe early April."

Handling Temporal Ambiguity

Much temporal information in discovery is ambiguous or imprecise. Effective timeline systems must handle this ambiguity rather than forcing false precision.

When sources provide approximate dates, the system represents this appropriately. A witness saying "sometime in early 2023" generates a timeline entry spanning January through March rather than arbitrarily selecting a specific date. A document saying "approximately two weeks

© 2025 Merlin Search Technologies, Inc.

after the incident" creates a temporal range rather than precise date.

Relative temporal references require resolving relationships between events. If Document A says an event occurred "after the Board meeting" and the Board meeting date is known, the system can establish minimum bounds for the event date. If multiple documents establish a sequence (A happened before B, B happened before C), the system can order events even without knowing precise dates.

Conflicting temporal information should be preserved rather than suppressed. If one source dates an event to March while another suggests April, the timeline should reflect this conflict rather than arbitrarily choosing one source over another. These conflicts often reveal important issues—memory failures, document dating errors, or deliberate misrepresentations.

Temporal Reasoning and Inference

Beyond extracting explicit temporal information, AI systems can perform temporal reasoning—drawing inferences from available evidence about when events likely occurred or identifying temporal relationships that aren't explicitly stated.

For example, if Email A references "the meeting we had yesterday" and Email A is dated March 16, the system can infer that a meeting occurred on March 15. If Contract B has an effective date following signatures by three weeks and all signatures are dated by April 3, the system can infer the effective date as April 24.

This reasoning must remain conservative—drawing only well-supported inferences and clearly distinguishing inferred dates from explicit temporal information. Aggressive temporal reasoning risks creating false chronologies based on flawed assumptions. The goal is enriching timelines with justified inferences while avoiding speculation.

Non-Agentic Timeline Applications

We'll start with non-agentic applications where humans direct specific timeline construction tasks and AI executes them systematically.

Single-Source Timeline Extraction

The most straightforward application extracts temporal information from individual sources—a set of emails, a deposition transcript, a collection of contracts. The system processes the source systematically, identifying all temporal references, extracting associated events or content, organizing chronologically, and producing timeline outputs.

For email collections, this might extract all emails in chronological order, identify key communications around critical dates, note when particular topics first appear in communications, and track when activity intensifies or diminishes. The resulting timeline shows communication patterns over time.

For testimony, timeline extraction identifies all events the witness describes, dates or timeframes the witness provides for those events, and temporal relationships the witness establishes. This creates testimony-based chronologies showing events according to witness recol-

© 2025 Merlin Search Technologies, Inc.

lection.

For contracts, extraction captures effective dates, termination dates, milestone dates, and any temporal obligations or deadlines the contracts specify. This produces contractual timelines showing when obligations arise and expire.

The workflow is straightforward—process the source through temporal extraction prompts, organize extracted information chronologically, format outputs appropriately, and verify accuracy through sampling. This single-source approach provides focused chronologies showing temporal dimensions of particular evidence types.

Multi-Source Timeline Integration

More sophisticated timeline applications integrate temporal information across multiple sources—combining emails, documents, testimony, and other evidence into unified chronologies. This integration reveals how different evidence types converge to establish event sequences.

The integration process extracts temporal information from each source type, reconciles varying date formats and precision levels, identifies where different sources address the same events, organizes all temporal information chronologically, and produces integrated timelines with citations to all supporting sources.

Consider a timeline showing communications about a product defect. Email extraction identifies when engineers first discussed quality issues. Document analysis reveals when formal defect reports were filed. Testimony extraction shows when executives claim they learned about the problem. Meeting minutes indicate when the defect was discussed with the Board. Integrating all these sources creates a comprehensive chronology showing information flow about the defect.

The integrated timeline reveals patterns invisible in single sources. Perhaps emails show engineers discussing quality concerns in January, but executives testify they didn't learn about problems until March, yet Board meeting minutes from February reference quality discussions. This pattern suggests either memory failures or potentially deliberate misrepresentations about knowledge timing.

Quality assurance for integrated timelines focuses on verification that sources are correctly integrated, that events are properly attributed to their sources, that temporal relationships are accurately represented, and that precision levels are appropriately indicated. Sampling protocols check significant timeline entries against source documents to ensure accuracy.

Temporal Gap Identification

Beyond extracting temporal information, timeline systems can identify gaps—periods where activity would be expected but evidence doesn't appear. These gaps often indicate missing documents, unexplored time periods, or areas deserving investigation.

Gap identification requires understanding normal activity patterns. If emails flow regularly between parties for months but show a two-week silence during a critical period, that gap deserves attention. If financial records exist for all months except one, that absence is significant.

© 2025 Merlin Search Technologies, Inc.

If testimony covers events before and after a particular date range but doesn't address what happened during that range, the omission warrants investigation.

The gap identification workflow analyzes timeline patterns to establish baseline activity levels, identifies periods with unexpectedly low document volumes or activity, evaluates whether gaps correlate with significant events, and flags gaps deserving investigation or suggesting incomplete discovery.

For example, analyzing email communications between executives might reveal that messages flow steadily except for a three-week period immediately before a crucial decision. That gap might reflect deleted communications, lost documents, or simply a period when communications occurred through other channels. Regardless of explanation, the gap deserves investigation.

Gap identification also helps assess production completeness. If the requesting party sought documents from January through December but productions show abundant documents for most months and sparse documents for March and August, those gaps suggest potentially incomplete production or targeted withholding.

Temporal Pattern Recognition

Timeline analysis often reveals patterns that inform case strategy. Activity intensifies around certain events. Communications shift from particular topics. Participants who communicated frequently suddenly stop interacting. These patterns provide context for understanding how events unfolded.

AI systems can systematically identify temporal patterns by analyzing communication frequency over time, tracking when particular topics appear and disappear, identifying when relationships between participants change, and recognizing when activity patterns shift significantly.

A pattern analysis might reveal that discussions of a merger proceeded calmly for months, then communications suddenly intensified in a two-week period before the announcement. Or that a key executive who regularly communicated with subordinates about operations stopped sending emails about a particular project after a specific date. Or that mentions of regulatory concerns appeared sporadically until a particular month when they suddenly dominated communications.

These patterns don't necessarily prove anything—they raise questions deserving investigation. Why did activity intensify during that period? What changed that caused the executive to stop discussing the project? What event prompted sudden concern about regulatory issues? Temporal pattern analysis identifies these questions systematically.

Comparative Timeline Analysis

Different sources often produce different chronologies for the same events. Comparing these timelines reveals discrepancies that might indicate memory failures, document issues, or inconsistent accounts.

Comparative analysis extracts timelines from different source types, identifies events that ap-

© 2025 Merlin Search Technologies, Inc.

pear in multiple timelines, compares how different sources date those events, flags discrepancies exceeding expected variation, and generates reports highlighting temporal contradictions.

For example, comparing document-based timelines to testimony-based timelines might reveal that witnesses consistently date events earlier than documentary evidence suggests, perhaps reflecting telescoping where witnesses compress timeframes in memory. Or comparison might show that one witness's chronology conflicts with all other evidence, suggesting either unique knowledge or inaccurate recollection.

These comparisons prove particularly valuable for impeachment. If a witness testifies that Event A occurred in March but contemporaneous documents establish it occurred in May, that discrepancy undermines the witness's credibility. If a party's interrogatory responses date a decision to January but documents show the decision wasn't made until April, that inconsistency deserves exploration.

Agentic Timeline Applications

Agentic systems enable more sophisticated timeline applications where AI autonomously pursues comprehensive chronology construction across all available evidence.

Comprehensive Timeline Investigation

Traditional timeline construction requires humans to identify what events matter, search for relevant evidence, extract temporal information, and organize chronologically. Agentic systems can pursue much of this work autonomously within defined investigation goals.

Given an investigation goal—"construct comprehensive timeline of events leading to the product recall decision"—an agentic system can identify what types of events are relevant, search across all available sources for information about those events, extract temporal references from discovered sources, organize findings chronologically, identify gaps and contradictions, and generate comprehensive timeline reports with full source citations.

The investigation proceeds through multiple phases. Initial broad searches identify documents and testimony addressing the investigation topic. Temporal extraction from those sources produces preliminary chronology. Analysis of that chronology identifies gaps or unclear sequences prompting additional targeted searches. Newly discovered sources are integrated into the evolving timeline. The process iterates until the system determines the chronology is comprehensive within available evidence.

This autonomous investigation compresses work that might require weeks of human effort into hours or days. The system works systematically through available evidence without the memory limitations and fatigue that affect human timeline construction. It maintains consistency in how temporal information is extracted and organized. It identifies patterns and gaps that might be missed in manual review.

The key difference from non-agentic approaches is autonomy—the system determines what searches to run, evaluates whether results are adequate, decides what additional investigation is needed, and refines the timeline iteratively. Humans establish the investigation goal and boundaries but the system executes the investigation autonomously.

© 2025 Merlin Search Technologies, Inc.

Timeline-Driven Investigation Strategy

In some investigations, chronology itself drives strategy. Understanding event sequences reveals what happened when, suggesting what evidence should exist, what witnesses would have knowledge, and what explanations are plausible.

Agentic systems can pursue timeline-driven investigations autonomously. The system constructs initial chronology from available evidence, analyzes that chronology to identify what additional evidence should exist based on events identified, searches for expected evidence and evaluates whether it's present, investigates gaps where expected evidence doesn't appear, and refines chronology iteratively as new evidence is discovered.

For example, if timeline analysis reveals that executives met to discuss quality issues on March 15, the system might autonomously search for meeting agendas, minutes, or follow-up communications from that date. If those documents aren't found, the gap gets flagged for investigation. If found, the system extracts temporal information from them, potentially revealing subsequent meetings or decisions that require further investigation.

This investigative approach uses chronology not just as output but as tool—the developing timeline guides what evidence to seek next. The system pursues leads suggested by temporal patterns, investigates gaps systematically, and builds increasingly comprehensive chronologies through iterative evidence gathering.

Event Causation and Temporal Dependency Analysis

Understanding not just when events occurred but how they're temporally related—what preceded what, what might have caused what, what depended on what—provides deeper insight than simple chronologies.

Agentic systems can analyze temporal dependencies autonomously. The system identifies key events in the timeline, analyzes what preceded each event (potential causes or precursors), evaluates what followed each event (potential effects or consequences), identifies decision points where temporal paths diverged, and constructs narrative explanations of how events unfolded based on temporal relationships.

This analysis goes beyond simple before-after relationships to evaluate plausibility of causal connections. If Event A occurred and then Event B occurred, that's temporal sequence. If Event B was a predictable consequence of Event A, that's a causal relationship. If documents show Event A prompted discussions leading to decisions resulting in Event B, that's supported causal chain.

Consider analyzing decisions leading to a product launch. Simple chronology might show: quality concerns raised (January), testing completed (February), approval meeting held (March), launch occurred (April). Causal analysis examines whether testing addressed the quality concerns, whether the approval meeting considered test results, whether the launch proceeded despite unresolved issues, and whether the temporal sequence suggests adequate deliberation or rushed decisions.

This causation analysis doesn't prove anything legally—causation remains subject to legal standards and human judgment. But it systematically organizes temporal evidence bearing

© 2025 Merlin Search Technologies, Inc.

on causation questions, identifies what evidence supports or undermines causal theories, and provides factual foundation for causation arguments.

Witness Timeline Comparison and Credibility Analysis

Different witnesses often provide different chronologies for the same events. Systematic comparison of witness timelines reveals consistency patterns affecting credibility assessment.

Agentic systems can pursue witness timeline comparison autonomously. The system extracts timeline from each witness's testimony, identifies events discussed by multiple witnesses, compares how different witnesses date those events, analyzes whether discrepancies exceed normal variation, evaluates whether discrepancies suggest coordinated accounts or independent recollections, and generates credibility analysis based on temporal consistency.

If five witnesses independently describe the same meeting as occurring in March, that convergence suggests reliable recollection. If one witness dates it to January, that outlier deserves investigation. If all witnesses provide identical unusual temporal details (not just "March" but "March 17 at 2:30 PM in Conference Room B"), that precision might suggest coaching rather than independent memory.

The analysis also evaluates witness chronologies against documentary evidence. Witnesses whose temporal recollections consistently align with contemporaneous documents show more reliable memories than witnesses whose chronologies conflict with documentary timelines. These reliability patterns inform how much weight to give witness testimony on timing questions.

Dynamic Timeline Maintenance

Discovery is dynamic—new evidence emerges, testimony occurs, productions arrive. Static timelines become outdated quickly, requiring manual updates to incorporate new information.

Agentic systems can maintain dynamic timelines that update automatically as new evidence becomes available. The system monitors for new documents or testimony entering the evidence collection, extracts temporal information from new sources, integrates new information into existing timeline, identifies where new evidence confirms, contradicts, or supplements existing timeline, and flags significant changes requiring human attention.

This dynamic maintenance ensures timelines remain current throughout litigation. When new depositions occur, testimony-based timeline updates automatically. When supplemental productions arrive, document-based chronologies incorporate new materials. The system identifies where new evidence changes understanding of event sequences or reveals previously unknown events.

Notifications alert legal teams to significant timeline changes. If new documents reveal events during a previously identified gap, the team learns immediately. If new testimony contradicts the established chronology, that conflict gets flagged. If evidence patterns shift based on new information, those changes receive attention.

© 2025 Merlin Search Technologies, Inc.

Specialized Timeline Applications

Beyond general chronology construction, timeline techniques apply to specific discovery challenges requiring temporal analysis.

Knowledge Timeline Construction

Many disputes hinge on when parties knew particular facts. Product liability cases ask when manufacturers knew about defects. Securities cases examine when executives knew about financial problems. Employment cases explore when employers knew about discrimination or harassment.

Knowledge timelines track not what happened but when parties learned what happened. The construction process identifies all events potentially affecting party knowledge, extracts evidence showing when parties actually learned about events, distinguishes between potential knowledge (information that existed) and actual knowledge (information party possessed), and tracks knowledge progression as information flowed through organizations.

For example, a knowledge timeline might show that engineers identified a defect in January, quality assurance documented the defect in February, reports circulated to management in March, and executives discussed the defect at an April Board meeting. This timeline doesn't just show when the defect existed—it shows how knowledge of the defect progressed through the organization.

Knowledge timelines often reveal gaps between when information existed and when key decision-makers claim they learned it. These gaps raise questions about information flow, organizational communication, or credibility of claimed ignorance.

Decision Timeline Construction

Understanding decision-making processes requires chronologies showing deliberation, information gathering, stakeholder consultation, and final decisions. Decision timelines organize evidence showing how decisions were made.

The construction process identifies decisions requiring timeline analysis, extracts evidence showing decision processes (meetings, discussions, analyses, consultations), tracks information available to decision-makers at relevant times, identifies when decisions were actually made, and analyzes whether decision processes show adequate deliberation.

A merger decision timeline might show initial discussions, due diligence processes, financial analyses, board deliberations, stakeholder consultations, and final approval—with dates for each phase. This chronology reveals whether decisions followed appropriate processes or were rushed without adequate information.

Decision timelines prove particularly valuable for showing that decisions were made with knowledge of particular facts. If executives claim they approved a strategy without knowing about problems, but the decision timeline shows problem reports circulated before the approval meeting, that temporal evidence undermines claimed ignorance.

© 2025 Merlin Search Technologies, Inc.

Communication Network Evolution

Understanding how communication patterns evolved over time reveals relationship changes, information flow patterns, and organizational dynamics. Communication evolution timelines track these patterns.

The construction process analyzes communications over time, identifies who communicated with whom during different periods, tracks when communication patterns changed, recognizes when participants entered or left communication networks, and visualizes how communication networks evolved.

For example, analysis might reveal that Executive A regularly communicated with Engineering about a project until a particular date, after which all communications went through Executive B instead. This shift suggests relationship changes or information control worth investigating. Or analysis might show that certain topics were openly discussed across an organization until a specific date, after which discussions became restricted to a small group—suggesting concern about sensitivity.

These temporal patterns in communication networks provide context for understanding events and relationships that static network analysis might miss.

Regulatory Compliance Timelines

Cases involving regulatory violations require timelines showing when regulations took effect, when parties knew about requirements, when compliance efforts occurred, and when violations happened. Compliance timelines organize this temporal evidence.

The construction process identifies relevant regulatory requirements and effective dates, tracks when parties learned about requirements, documents compliance efforts and their timing, identifies when violations occurred relative to knowledge and requirements, and analyzes whether compliance efforts were timely and adequate.

A compliance timeline might show that regulations took effect in January, the company learned about requirements in February, began compliance planning in March, implemented controls in May, but violations continued through April and into June. This chronology reveals both the compliance timeline and the violation timeline, supporting analysis of whether the company acted reasonably after learning about requirements.

Quality Assurance for Timelines

Timeline quality assurance ensures chronologies accurately represent temporal evidence while acknowledging uncertainty appropriately.

Temporal Accuracy Verification

The foundation of timeline QA is verifying that dates and temporal relationships accurately reflect source evidence. Verification sampling checks timeline entries against sources to confirm dates are correctly extracted, events are accurately described, temporal relationships are properly represented, and precision levels are appropriately indicated.

© 2025 Merlin Search Technologies, Inc.

Verification intensity should focus on critical timeline periods and events. High-stakes decisions, disputed events, or periods where chronology significantly affects case theories deserve close verification. Less critical periods might receive lighter sampling.

Verification also checks that timeline entries properly cite sources. Every temporal claim should reference supporting evidence—documents, testimony, or other sources establishing dates. This citation verification ensures timeline entries can be defended if challenged.

Completeness Assessment

Beyond accuracy, timelines should be comprehensive—including significant temporal information rather than arbitrary selections. Completeness assessment evaluates whether timelines capture events that matter for case purposes.

Assessment approaches include comparing timelines to case chronologies developed by legal teams to identify omissions, having attorneys review timelines and note missing events they know occurred, checking whether all key decision points and significant events are represented, and evaluating whether timelines provide sufficient detail for their intended purposes.

Completeness failures often reflect scope limitations rather than extraction failures. If timeline instructions didn't specify certain event types, they won't appear. Iterative refinement based on completeness feedback improves timeline scope and utility.

Consistency Checking

Timelines integrating multiple sources should show internal consistency—the same event shouldn't be dated differently in different sections. Consistency checking identifies and resolves these discrepancies.

Checking involves identifying duplicate or overlapping events across sources, comparing dates assigned to the same events, investigating discrepancies to determine which source is more reliable, and either reconciling discrepancies or explicitly noting conflicting temporal information.

Some inconsistencies reflect genuine source conflicts rather than timeline errors. When Document A dates an event to March while Document B suggests April, the timeline should reflect this conflict rather than forcing consistency. These conflicts themselves provide valuable information about source reliability or event complexity.

Gap Analysis Verification

Identified gaps should reflect genuine absences of evidence rather than extraction failures. Gap verification ensures that flagged gaps are real and significant.

Verification checks that gap periods actually lack evidence (confirming no relevant documents were missed), that gaps are meaningful given context (distinguishing between suspicious absences and normal low-activity periods), and that gap identification considers all available sources (not just selected source types).

© 2025 Merlin Search Technologies, Inc.

False gap identification—flagging periods that actually have adequate evidence—reduces timeline credibility. Verification sampling helps ensure gap analysis is reliable before using gaps to suggest incomplete production or missing evidence.

Integration and Presentation

Timelines serve multiple audiences and purposes, requiring flexible presentation formats and integration with other discovery workflows.

Format Adaptation

Effective timeline presentation adapts to audience and purpose. Trial timelines emphasize clarity and visual impact. Attorney work product timelines prioritize comprehensive detail. Client presentations balance comprehensiveness with accessibility.

Format options include chronological lists with detailed entries and citations, visual timelines showing events on temporal axes, narrative chronologies integrating events into flowing text, comparative formats showing multiple chronologies side-by-side, and interactive formats enabling filtering and exploration
.
The presentation format should match the use case. Cross-examination timelines need quick reference to specific dates with citations. Motion practice timelines require formatted citations suitable for legal writing. Jury presentations need visual clarity without overwhelming detail.

Discovery Platform Integration

Timelines don’t stand alone—they integrate with document review platforms, case management systems, and other discovery tools. Integration enables users to move seamlessly from timeline entries to source documents.

Integration approaches include linking timeline entries to source documents in review platforms, enabling timeline filtering based on document properties or coding, incorporating timeline information into document metadata, and providing timeline views within review platforms alongside traditional document views.

This integration makes timelines active tools rather than static reports. Attorneys reviewing documents can see temporal context. Case strategists can filter timelines to focus on relevant periods or events. Researchers can move from timeline entries to underlying evidence instantly.

Collaboration Support

Timeline construction often involves multiple team members contributing temporal information from different sources. Collaboration features enable distributed timeline development.
Collaboration support includes allowing multiple contributors to add timeline information, tracking who contributed each entry for accountability, providing review workflows where contributors can validate each other’s entries, and maintaining version history showing how timelines evolved.

© 2025 Merlin Search Technologies, Inc.

This collaborative approach distributes timeline construction across team members while maintaining quality through peer review. Paralegals might extract dates from documents, associates might analyze testimony for temporal references, and senior attorneys might validate significant entries before timelines are used in motion practice or trial.

Cost-Effectiveness and ROI

Timeline construction using AI must demonstrate value justifying costs and implementation effort.

Direct Cost Analysis

Direct costs for AI timeline construction include API charges proportional to document volume and analysis complexity, platform costs for systems managing timeline construction, storage for timeline data and supporting evidence, and vendor fees for commercial timeline tools.

For a matter with 50,000 documents and 20 deposition transcripts, comprehensive timeline construction might cost $100 in GenAI charges, depending on sophistication of analysis required. Commercial timeline platforms might charge additional subscription or per-use fees.

Traditional timeline construction might consume 40-80 hours of paralegal or junior associate time at $250 to 750 per hour, yielding costs in the tens of thousands. AI-assisted timeline construction show strong cost advantages even accounting for verification overhead.

Efficiency and Strategic Value

Beyond direct costs, AI timeline construction offers efficiency benefits enabling earlier strategic assessment, faster motion practice supported by systematic chronological evidence, more comprehensive trial preparation, and identification of temporal patterns informing case strategy.

Strategic value often exceeds cost savings. A comprehensive timeline might reveal that key executives knew about problems earlier than they claimed, supporting fraud allegations. Or timeline analysis might identify gaps suggesting incomplete production, informing discovery disputes. These strategic insights can fundamentally affect case outcomes.

For high-stakes matters where chronology drives liability or damages, comprehensive timeline construction proves valuable regardless of cost considerations. The investment in systematic temporal analysis yields strategic advantages that may determine case success.

Looking Ahead

This chapter examined how AI transforms timeline and chronology construction—enabling comprehensive temporal analysis that would be impractical through manual methods. The applications described show how discovery professionals can systematically extract temporal information, organize complex chronologies, identify patterns and gaps, and generate timelines serving diverse purposes from trial presentation to strategic analysis.

© 2025 Merlin Search Technologies, Inc.

Timeline construction represents foundational work for trial preparation. Understanding what happened when, who knew what at each point, and how events unfolded creates the factual framework that supports all trial advocacy. But having a comprehensive timeline is only the beginning. The next challenge is transforming that chronological understanding into persuasive narratives that convince judges and juries.

Chapter 13 examines how generative AI assists with this transformation—taking the factual foundations built through timeline analysis and marshaling them into compelling opening and closing arguments. We'll explore how AI can synthesize vast trial records, identify the most persuasive evidence for specific legal theories, and generate sophisticated argument drafts that maintain professional quality while dramatically reducing preparation time. The chapter demonstrates how these capabilities change trial preparation while preserving the strategic judgment and advocacy skills that define excellent lawyering

© 2025 Merlin Search Technologies, Inc.

MERLIN

13

Using GenAI to Write Powerful and Persuasive Opening and Closing Arguments

© 2025 Merlin Search Technologies, Inc.

CHAPTER THIRTEEN

Using GenAI to Write Powerful and Persuasive Opening and Closing Arguments

Trial advocacy has always demanded the synthesis of vast amounts of evidence, the ability to construct compelling narratives, and the skill to marshal facts into persuasive arguments. For decades, these tasks required teams of lawyers working countless hours to review transcripts, analyze exhibits, and craft arguments that could sway judges and juries.

Generative AI has fundamentally changed this landscape. Modern Large Language Models can now process thousands of pages of trial materials in minutes, identify the most relevant evidence for specific legal theories, and generate sophisticated first drafts of opening and closing arguments that previously would have taken weeks to develop. This chapter demonstrates how smart discovery professionals can leverage these capabilities to transform trial preparation while maintaining the professional judgment and strategic thinking that defines excellent advocacy.

The question is no longer whether AI can assist with trial preparation. The question is how legal professionals will adapt to these powerful tools and integrate them into their practice to serve their clients more effectively.

The Challenge: Creating Arguments from Complex Trial Materials

Opening and closing arguments represent the pinnacle of trial advocacy. They require lawyers to:

- Synthesize massive volumes of testimony and exhibits
- Identify the most compelling evidence for their theory of the case
- Construct narratives that acknowledge weaknesses while emphasizing strengths
- Anticipate and counter opposing arguments
- Present complex technical information in accessible ways
- Adapt arguments to the specific perspectives of different parties in multi-party litigation

Traditional preparation for these arguments involves teams of lawyers spending days or weeks reviewing trial transcripts, organizing exhibits, drafting argument sections, and revising until the final product emerges. Even with dedicated teams, the time pressure and volume of material make it difficult to ensure that every relevant piece of evidence has been considered and integrated into the argument.

Generative AI transforms this process by enabling comprehensive analysis of the entire trial record in a fraction of the time, while producing sophisticated first drafts that lawyers can refine and personalize.

The BP Oil Spill Trial: A Case Study in AI-Assisted Advocacy

The Deepwater Horizon oil spill trial provides an ideal case study for demonstrating the power

of GenAI in trial preparation. On April 20, 2010, an explosion on the Deepwater Horizon drilling rig killed 11 workers and triggered an oil spill that released approximately 4.9 million barrels of crude oil into the Gulf of Mexico over 87 days. The Phase One trial focused on liability, with multiple defendants facing complex questions of shared responsibility.

The trial generated thousands of pages of testimony and exhibits. Multiple parties needed to construct arguments from different strategic perspectives, each requiring deep analysis of the same evidence to support divergent theories of liability. This complexity makes it an excellent test case for evaluating GenAI's capabilities in legal analysis and argument construction.

Using the Merlin Alchemy platform, we demonstrated how GenAI could analyze this massive trial record and generate sophisticated closing arguments for each party: TransOcean, Halliburton, the government plaintiffs, and BP itself. Each argument required the AI to adapt to different legal positions while maintaining rigorous analytical standards and comprehensive use of the available evidence.

The Methodology: A Systematic Approach to AI-Assisted Argument Development

The process for using GenAI to create trial arguments follows a systematic methodology that combines AI's analytical power with human legal expertise. This approach has been refined through extensive testing and real-world application.

Step One: Document Integration and Repository Creation

The foundation of effective AI-assisted argument development is a comprehensive, well-organized repository of trial materials. This includes:

Trial Transcripts: Complete testimony from all witnesses, organized by date and witness name. The transcripts should be processed to enable both semantic and keyword-based searching.

Trial Exhibits: All documents, communications, technical reports, and other materials admitted into evidence. Each exhibit should be indexed and linked to the testimony that references it.

Document Structuring: Materials should be divided into logical sections that enable the AI to locate and retrieve relevant evidence efficiently. For the BP trial, we divided the materials into over 9,000 sections of testimony and exhibit text.

Metadata Integration: Each document section should include metadata indicating the source, date, witness, and other relevant contextual information that helps the AI understand the significance of the evidence.

This preparation work is essential because the quality of AI-generated arguments depends directly on the quality and organization of the underlying materials. A well-structured repository enables the AI to find and integrate the most relevant evidence for any given argument.

Step Two: Developing the Strategic Outline

© 2025 Merlin Search Technologies, Inc.

Before generating the actual argument text, the first step is to create a comprehensive outline that establishes the strategic framework. This outline development process leverages the AI's ability to analyze the entire trial record and identify the strongest arguments available.

The prompt for outline development should include:

Role Definition: Clear instructions about which party the AI is representing and the overall strategic objective. For example: "You are the lead trial lawyer for British Petroleum in the Deepwater Horizon Oil Spill Phase One Liability Trial. You have years of experience defending large corporations in complex disasters."

Strategic Direction: Specific guidance about the theory of the case and key themes to emphasize. For BP's defense, this included: "Your goal will be to push responsibility for this terrible tragedy to the other defendants and to show that BP should not be held liable if at all possible. Tell the story that BP was not the operator of the well and wasn't responsible for the key elements that failed here."

Output Requirements: Clear instructions about the format and content of the outline, including the level of detail needed for each section.

The AI processes the entire trial record through sophisticated algorithmic search methods to identify evidence that supports the requested strategic approach. For the BP trial, we used three complementary search techniques:

1. **Natural Language Search:** Semantic, vector-based searching that identifies contextually relevant evidence even when specific keywords are not present

2. **Keyword Algorithm:** Traditional keyword searching to ensure important specific terms and phrases are captured

3. **Machine Learning Classification:** Trained algorithms that learn from initial seed rankings to identify similar relevant sections

This multi-method approach ensures comprehensive coverage of the trial record, catching evidence that any single method might miss.

The resulting outline provides a detailed roadmap for the argument, organized into major sections with supporting subsections and specific citations to testimony and exhibits. For BP's closing argument, the outline included seven major sections:

- Introduction: The Tragedy in Context
- BP Was Not the Direct Operator of the Well
- Technical Failures Beyond BP's Direct Control
- Critical Decision-Making and Operational Failures
- BP's Safety Culture and Risk Management
- Shared Responsibility Among Multiple Parties
- Conclusion: A Fair Assessment of Responsibility

Each section included detailed subsections with specific document citations, creating a com-

© 2025 Merlin Search Technologies, Inc.

prehensive framework for the final argument.

Step Three: Generating the Argument Text

With the outline in place, the next step is to generate the actual argument text. For lengthy arguments, this is best done in sections to maintain quality and enable iterative refinement. The prompt for argument generation should include:

Reinforced Role Definition: Restate the lawyer's role and expertise to maintain consistency with the outline phase.

Specific Instructions: Clear guidance about format, tone, and objectives. For example: "Make your argument in narrative form as if you were giving it to the Judge in this case. Include citations to testimony and exhibits that support your arguments."

Outline Section: The specific portion of the outline being addressed in this generation run.

Citation Requirements: Explicit instructions to support all factual assertions with specific citations to the trial record.

For the BP closing argument, we divided the outline into three groupings and generated the argument in three separate runs, each processing the relevant sections against the 9,000+ document sections in the repository. This approach allows the AI to focus on each portion of the argument while maintaining access to the entire trial record for evidence selection.

The AI generates narrative argument text that follows the outline structure, integrates the most relevant evidence from the trial record, and includes specific citations linking each factual assertion to its supporting testimony or exhibit. The result is a sophisticated first draft that reads like an experienced trial lawyer's argument, complete with rhetorical devices, strategic emphasis, and comprehensive evidence integration.

Step Four: Combining and Refining the Sections

The final technical step is combining the separately-generated sections into a unified argument. Modern LLMs output their text in Markdown format, which can be easily integrated into word processing applications. Google Docs, for example, includes a "paste from Markdown" option that automatically formats the text with proper headers, emphasis, and structure. Once combined, the argument should be reviewed for:

Narrative Flow: Ensuring smooth transitions between sections and logical progression of arguments.

Consistent Tone: Verifying that the different sections maintain a consistent voice and strategic approach.

Citation Accuracy: Checking that all citations are properly formatted and accurately reference the trial record.

Strategic Emphasis: Confirming that key arguments receive appropriate prominence and that

© 2025 Merlin Search Technologies, Inc.

the overall structure effectively supports the theory of the case.

This review process is where human legal expertise becomes essential. The AI provides a comprehensive first draft, but skilled lawyers must refine the argument to reflect their judgment, emphasize their strategic priorities, and ensure that the final product represents their best advocacy.

The Results: Sophisticated Arguments in Hours, Not Weeks

The BP trial series demonstrated remarkable results across multiple parties and strategic approaches.

TransOcean's Closing Argument

For TransOcean, the drilling rig operator, the argument needed to emphasize the complex contractual relationships in deepwater drilling while shifting primary responsibility to BP as the well owner and lessee. The AI generated a comprehensive argument that:

- Acknowledged TransOcean's role as rig operator while emphasizing BP's ultimate authority
- Highlighted specific instances where BP overrode TransOcean's safety recommendations
- Integrated technical testimony about drilling operations and well control
- Constructed a narrative of shared responsibility with primary fault lying with BP

Halliburton's Closing Argument

Halliburton's position as the cementing contractor required a different strategic approach, emphasizing the limited scope of their work and BP's control over well design decisions. The AI constructed an argument that:

- Clearly delineated Halliburton's contractual boundaries
- Documented BP's comprehensive control over operations and decisions
- Presented evidence of BP systematically prioritizing cost over safety
- Identified specific examples of BP overriding contractor safety recommendations

Government Plaintiffs' Closing Argument

The government's position required demonstrating gross negligence and willful misconduct by the defendants, with a focus on establishing BP's primary liability. The AI developed an argument that:

- Synthesized evidence across multiple parties to show coordinated failures
- Emphasized the environmental and economic devastation caused by the spill
- Documented repeated warnings ignored by the defendants
- Constructed a narrative of reckless decision-making driven by profit over safety

© 2025 Merlin Search Technologies, Inc.

BP's Closing Argument: The Ultimate Challenge

The most difficult assignment was creating BP's defense argument. Most legal observers viewed BP's position as nearly indefensible given the overwhelming evidence of their responsibility. The court ultimately ruled that BP was guilty of gross negligence and willful misconduct, apportioning 67% of blame to BP.

Yet the AI successfully constructed a legitimate defense strategy based on shared responsibility in deepwater drilling operations. The argument:

- Acknowledged the tragedy while establishing the complex nature of deepwater drilling
- Emphasized BP's limited operational control as non-operator of the well
- Highlighted technical failures by other parties, particularly the blowout preventer maintained by TransOcean
- Identified cementing failures by Halliburton
- Documented critical operational decisions made by TransOcean's crew
- Presented evidence of BP's safety commitment and reliance on contractor expertise
- Argued for fair allocation of responsibility among all parties

The final argument spanned 60 pages of sophisticated legal reasoning supported by specific citations to the trial record. What's remarkable is that this entire project was completed in about two hours on a Sunday afternoon, at a cost less than what many would spend on a nice dinner.

Key Benefits of AI-Assisted Trial Argument Development

The BP trial series demonstrated several transformative advantages of using GenAI for trial preparation:

Comprehensive Evidence Analysis

The AI processed thousands of pages of trial materials in minutes, ensuring that no relevant evidence was overlooked. Traditional preparation methods, with lawyers manually reviewing transcripts and exhibits, inevitably miss some relevant material due to time constraints and the sheer volume of information.

The systematic algorithmic approach catches evidence that human reviewers might miss, including testimony that uses different terminology but addresses the same concepts, exhibits referenced in multiple contexts throughout the trial, and connections between different witnesses' accounts that support key arguments.

Multiple Strategic Perspectives

The ability to generate arguments for different parties using the same evidence demonstrates how GenAI can help legal teams explore alternative theories of liability and defense strategies. Rather than committing to a single approach early in trial preparation, lawyers can quickly generate and evaluate multiple strategic options.

This flexibility is particularly valuable in complex multi-party litigation where the optimal stra-

tegic approach may not be immediately obvious. The AI can generate arguments from different perspectives, allowing the legal team to assess the strengths and weaknesses of each approach before committing resources to full development.

Rapid Iterative Development

Traditional argument preparation follows a linear process: review documents, draft sections, revise, incorporate new evidence, revise again. This process can take weeks, and significant strategic pivots late in the process can require starting over.

GenAI enables rapid iteration. If new evidence emerges or strategic priorities shift, the AI can quickly generate a revised argument incorporating the new direction. This flexibility allows legal teams to remain responsive to developments throughout trial preparation.

Cost Efficiency

Perhaps the most striking advantage is the dramatic reduction in cost. Generating sophisticated first drafts in hours rather than weeks produces substantial savings on legal fees. More importantly, it allows senior lawyers to focus their expensive time on strategic refinement and personalization rather than initial research and drafting.

For the BP arguments, the entire project cost less than a nice dinner for two. While senior trial lawyers would never present AI-generated arguments verbatim, having a comprehensive first draft that would take associates weeks to prepare provides enormous value.

Quality Control and Citation Accuracy

Every point in an AI-generated argument is linked to specific evidence in the trial record. This comprehensive citation enables efficient verification of factual assertions and ensures that the argument is built on the actual trial record rather than assumptions or generalizations.
The systematic approach also helps ensure consistency in how evidence is characterized and reduces the risk of inadvertent misrepresentation that can occur when multiple team members work on different sections without complete coordination.

Practical Applications Beyond Major Trials

While the BP trial provides a dramatic example of GenAI's capabilities, the same methodology applies to cases of all sizes and complexity levels.

Routine Civil Litigation

For standard civil cases, the ability to quickly analyze depositions and discovery materials and generate argument outlines can dramatically reduce preparation time for summary judgment motions, pretrial hearings, and trial preparation.

Criminal Defense

Criminal defense attorneys can use GenAI to analyze police reports, witness statements, fo-

© 2025 Merlin Search Technologies, Inc.

rensic evidence, and trial testimony to identify inconsistencies, generate alternative theories, and construct comprehensive defense arguments.

Appeals

Appellate advocacy requires synthesizing trial records, legal precedent, and policy arguments. GenAI can analyze trial transcripts to identify preserved issues, locate supporting testimony, and generate comprehensive statements of facts that accurately represent the record.

Arbitrations and Administrative Proceedings

These proceedings generate substantial documentary records that must be analyzed and integrated into closing arguments. GenAI provides the same benefits in these contexts as in traditional litigation.

Limitations and Considerations

Despite its impressive capabilities, GenAI is not a replacement for human legal expertise. Smart discovery professionals must understand its limitations and use it appropriately.

Professional Judgment Remains Essential

AI can analyze evidence and generate arguments, but it cannot exercise professional judgment about strategic priorities, assess the credibility implications of different argument choices, or evaluate how arguments will resonate with specific judges or juries. These remain fundamentally human skills.

The AI-generated argument is a first draft, not a final product. Experienced trial lawyers must review, refine, and personalize the argument to reflect their strategic judgment and advocacy style.

Ethical Obligations Apply

Lawyers remain responsible for the accuracy and propriety of all arguments they present, regardless of how those arguments were prepared. This includes:

- Verifying that all citations accurately represent the evidence
- Ensuring that characterizations of testimony are fair and not misleading
- Confirming that legal arguments are supported by applicable law
- Maintaining client confidentiality when using AI platforms
- Complying with all court rules regarding disclosure of AI assistance

Context and Nuance Matter

AI can miss subtle contextual factors that experienced lawyers immediately recognize. This includes understanding the relative credibility of different witnesses, appreciating the importance of demeanor evidence that doesn't appear in transcripts, and recognizing tactical implications of emphasizing certain arguments over others.

© 2025 Merlin Search Technologies, Inc.

Human review must ensure that the AI-generated argument reflects appropriate contextual judgment.

Repository Quality Determines Output Quality

The AI can only work with the materials it is given. Incomplete document repositories, poor document organization, or inadequate metadata will result in inferior arguments. Significant upfront investment in proper document preparation is essential for optimal results.

Best Practices for Implementation

Legal professionals looking to integrate GenAI into their trial preparation should follow these best practices:

Start with Document Organization

Invest time in creating a well-structured repository of trial materials. This includes:

- Complete transcripts divided into logical sections
- All exhibits properly indexed and linked to relevant testimony
- Metadata that helps the AI understand document context
- Clear organization that enables efficient retrieval

Develop Clear Strategic Direction

The quality of AI-generated arguments depends heavily on the clarity of the strategic direction provided in prompts. Take time to think through:

- The theory of the case you want to pursue
- Key themes that should be emphasized
- Arguments you want to avoid or minimize
- The tone and style appropriate for the audience

Use Iterative Refinement

Don’t expect perfection on the first generation. Use the AI’s output to identify areas that need additional evidence, alternative framings, or different strategic approaches. Generate multiple versions and compare them to identify the strongest approach.

Maintain Human Control

Always review AI-generated content carefully before incorporating it into final work product. Verify citations, assess strategic implications, and ensure the argument reflects your professional judgment. The AI is a powerful assistant, not a replacement for legal expertise.

© 2025 Merlin Search Technologies, Inc.

Document Your Process

Maintain clear records of how AI was used in argument preparation, including the sources analyzed, prompts used, and how AI-generated content was reviewed and revised. This documentation may be important for ethical compliance and responding to questions about your preparation process.

Stay Current on Ethical Requirements

The legal profession is still developing ethical guidelines for AI use. Stay informed about rules in your jurisdiction regarding disclosure of AI assistance, data security requirements, and limitations on AI use in legal practice.

Looking Forward: The Future of Trial Advocacy

The integration of GenAI into trial preparation represents a fundamental shift in legal practice. This technology is not coming in the future. It is here now and capable of handling even the most complex aspects of trial preparation.

The question for legal professionals is not whether AI will transform their practice, but how quickly they will adapt to leverage these powerful tools. Forward-thinking attorneys who embrace AI capabilities while maintaining the professional judgment and strategic thinking that defines excellent advocacy will deliver better results for their clients while working more efficiently.

The future of legal practice depends on thoughtful integration of artificial and human intelligence. GenAI amplifies legal expertise rather than replacing it. By combining AI's analytical power with human judgment, strategic thinking, and advocacy skills, lawyers can enhance both their efficiency and effectiveness while maintaining the high standards our profession demands.

Trial advocacy will always require the human qualities that define great lawyers: strategic thinking, persuasive communication, ethical judgment, and dedication to clients. GenAI simply provides better tools for applying those qualities to serve clients more effectively. Smart discovery professionals who master these tools will be better positioned to deliver outstanding results in an increasingly competitive legal marketplace.

Conclusion

This chapter demonstrated how generative AI transforms trial preparation—analyzing complex litigation materials and producing compelling opening and closing arguments that integrate comprehensively with underlying evidence. The BP Oil Spill trial series showed that AI can generate sophisticated arguments from multiple strategic perspectives while maintaining the quality and evidentiary support that effective advocacy requires.

The ability to generate trial arguments in hours rather than weeks represents significant advancement. Yet trial argument generation is one capability among many that AI brings to discovery practice. Document analysis, transcript review, timeline construction, relationship mapping, production assessment, and argument generation each deliver substantial value

© 2025 Merlin Search Technologies, Inc.

individually. The real transformation comes when these capabilities work together.

Chapter 14 examines how individual AI capabilities integrate into comprehensive AI-augmented discovery practice. We'll explore how organizations build institutional capabilities supporting AI adoption, what skills and competencies discovery professionals need as AI capabilities advance, how to measure success and ROI across AI-augmented workflows, and where this technology is heading. The chapter synthesizes lessons from the applications we've examined, moving beyond viewing AI as isolated tools toward understanding it as fundamental enhancement to how discovery professionals work—augmenting capabilities, accelerating workflows, and enabling analyses that would be impractical through traditional methods.

© 2025 Merlin Search Technologies, Inc.

MERLIN

14 Timeline and Chronology Construction

© 2025 Merlin Search Technologies, Inc.

CHAPTER FOURTEEN

Building AI-Augmented Discovery Practice

The preceding chapters demonstrated how AI transforms specific discovery workflows—document analysis, transcript review, timeline construction, relationship mapping, and production assessment. Each application shows substantial value individually. But the true power emerges when these capabilities combine into comprehensive AI-augmented discovery practice where technology and human expertise work in concert throughout the discovery lifecycle.

This chapter synthesizes lessons from Part Three's applications, examining how individual AI capabilities integrate into cohesive discovery workflows, how organizations build institutional capabilities supporting AI adoption, what skills and competencies discovery professionals need as AI capabilities advance, how to measure success and ROI across AI-augmented workflows, and where AI in discovery is heading as technologies continue evolving.

The goal is moving beyond viewing AI as collection of isolated tools toward understanding it as fundamental enhancement to how discovery professionals work—augmenting capabilities, accelerating workflows, and enabling analyses that would be impractical through traditional methods.

Integrated Discovery Workflows

Individual AI applications provide value, but integrated workflows where multiple AI capabilities work together deliver transformative results.

From Documents to Understanding

Traditional discovery workflows move linearly through stages—collection, processing, review, analysis, and production. AI-augmented workflows remain sequential in some respects but enable iteration and integration across stages that wasn't previously practical.

An integrated AI workflow might begin with initial assessment where AI rapidly processes samples of collected documents to understand population characteristics—key custodians, relevant time periods, major topics, document types present. This assessment informs search and culling strategy, suggesting productive keywords and date ranges while identifying custodians deserving priority review.

Document analysis then proceeds with AI assistance—categorization and metadata enrichment, privilege screening, issue tagging, and relevance assessment. But rather than treating these as separate sequential tasks, integrated workflows enable parallel processing where documents receive multiple analyses simultaneously. A single document might be assessed for privilege, tagged for relevant issues, categorized by type, and have entities and dates extracted—all in one processing pass.

© 2025 Merlin Search Technologies, Inc.

Timeline construction draws on document analysis results, extracting temporal information systematically while document analysis is ongoing. As new documents are processed, timeline updates automatically. This dynamic approach means understanding of event chronology develops progressively rather than waiting until document review completes.

Relationship mapping similarly builds on document analysis, extracting communication patterns and network structures as documents are processed. Combined with timeline data, this reveals not just who communicated with whom but how those relationships evolved temporally.

Production review applies these same analytical techniques to opposing parties' productions, using the same AI capabilities that assisted with analyzing your own documents to assess production completeness systematically.

This integration means that by the time document review completes, teams have comprehensive timelines, detailed relationship maps, identified patterns and gaps, and thorough understanding of case facts—not as separate follow-on projects but as natural byproducts of AI-augmented review workflows.

Iterative Investigation Workflows

Beyond linear integration, AI enables iterative investigation workflows where findings from one analysis inform subsequent investigations, creating virtuous cycles of discovery and understanding.

An investigation might begin with broad questions about key events or decisions. Initial AI-powered document searches retrieve relevant materials. Analysis of those materials reveals key participants, critical timeframes, and unexpected issues. These findings inform targeted follow-up searches—focusing on identified participants during relevant timeframes, searching for newly identified issues, and examining communications that earlier analysis suggested should exist.

This iteration continues through multiple cycles. Each analytical pass refines understanding, suggests new questions, and identifies gaps deserving investigation. Timeline gaps prompt targeted searches for missing periods. Relationship analysis reveals key players who should be deposed. Testimony from those depositions identifies documents that should exist, prompting production challenges or supplemental searches.

Traditional workflows struggled with this iteration because manual analysis was too slow and expensive for multiple cycles. AI makes iteration practical—rapid analysis enables quick pivots as understanding develops, low marginal costs support exploratory searches that might prove unproductive, and systematic processing ensures consistency across iterative cycles.

The result is discovery that adapts dynamically to evolving understanding rather than following rigid plans based on incomplete initial information. Cases develop organically as evidence reveals which theories have support and which don't.

© 2025 Merlin Search Technologies, Inc.

Multi-Source Synthesis

Comprehensive case understanding requires synthesizing information across documents, testimony, timelines, relationships, and other evidence types. AI excels at this synthesis, combining insights from diverse sources into coherent narratives.

A synthesis workflow might combine document-based timelines showing when things happened according to contemporaneous records, testimony-based timelines showing when witnesses claim things happened, relationship analysis showing who was involved and how they communicated, gap analysis identifying what's missing from the evidentiary record, and comparative analysis highlighting where sources agree or conflict.

This synthesis reveals patterns invisible when sources are examined separately. Documents might show executives discussing problems in March, but testimony claims they didn't learn about issues until May. Relationship analysis might reveal extensive communications about problems during the disputed April period. Gap analysis might identify missing documents from key players during critical times. Synthesizing these findings paints comprehensive picture of what actually happened versus what parties claim happened.

The synthesis also identifies what additional investigation would be most valuable. If synthesis reveals that all sources converge on certain facts, those facts seem established. If synthesis shows conflicts across sources, those conflicts deserve investigation. If synthesis identifies gaps where key information should exist but doesn't, those gaps justify targeted discovery efforts.

Building Organizational Capabilities

Successfully deploying AI in discovery requires more than purchasing technology—it requires building organizational capabilities supporting effective adoption.

Technical Infrastructure and Platform Selection

Organizations need infrastructure supporting AI-augmented workflows. This infrastructure combines existing discovery platforms with AI capabilities, either through platforms with built-in AI features, third-party AI tools integrating with platforms via APIs, or custom-developed AI solutions for specialized needs.

Platform selection should consider several factors. Integration with existing discovery infrastructure minimizes disruption and enables leveraging current investments. Scalability ensures platforms can handle matter sizes organizations typically encounter. Security and confidentiality protections must meet professional obligations and client expectations. Cost structures should align with usage patterns—some organizations prefer subscription models while others want usage-based pricing.

Critically, platforms should provide transparency into how AI analysis works and what decisions AI makes. Black box systems that provide outputs without explaining reasoning resist the verification that ethical AI use requires. Systems that show their work enable the human oversight that maintains professional responsibility.

© 2025 Merlin Search Technologies, Inc.

Technical infrastructure also includes data management capabilities—storing and organizing the large volumes of documents, analysis results, timelines, relationship data, and other information AI-augmented workflows generate. This data management must enable retrieval, update, and long-term retention as matters progress.

Process Design and Workflow Integration

Technology alone doesn't transform practice—organizations must design processes integrating AI capabilities into discovery workflows effectively.

Process design starts with mapping current workflows to understand existing practices, identifying pain points, and recognizing where AI could add value. Not every workflow step benefits from AI augmentation. Some tasks remain better suited to human judgment or traditional approaches. Effective integration deploys AI where it provides clear advantages while preserving human control over matters requiring professional judgment.

Workflow integration also requires defining handoffs between AI and human work. When does AI analysis pass to human verification? What triggers human intervention during autonomous AI investigations? Where do humans provide input or direction to AI processes? Clear handoff definitions prevent both gaps where work falls through cracks and redundancy where AI and humans duplicate efforts.

Process design should also accommodate iteration and learning. Initial AI deployments might require intensive human oversight. As organizations gain experience and confidence, oversight can lighten for routine applications while remaining intensive for novel uses or high-stakes matters. Process flexibility enables this adaptation.

Documentation of AI-augmented processes serves multiple purposes—providing audit trails supporting defensibility if methods are challenged, enabling training of new staff on AI-augmented workflows, facilitating quality improvement as organizations learn what works, and demonstrating professional competence if ethical obligations are questioned.

Training and Skills Development

Staff at all levels need training on working effectively with AI-augmented workflows. Training needs vary by role but all personnel require foundational understanding.

Attorneys need to understand what AI can reliably do and what remains uncertain, how to verify AI outputs appropriately, when AI applications are appropriate versus when traditional methods work better, and how to explain AI-assisted work to clients, courts, and opposing parties. They need enough technical understanding to make informed decisions about AI deployment while not requiring them to become AI experts.

Case managers and discovery coordinators need deeper technical knowledge—how to configure AI workflows for specific matters, how to monitor AI performance during processing, how to troubleshoot when AI systems produce unexpected results, and how to coordinate between AI capabilities and traditional review platforms.

Document reviewers need practical skills—understanding what AI has already analyzed and

© 2025 Merlin Search Technologies, Inc.

why, recognizing when AI assessments seem incorrect and should be overridden, and providing useful feedback when AI performance seems suboptimal.

Training should be ongoing rather than one-time. AI capabilities evolve rapidly. Organizations adopting AI should expect regular training updates as capabilities advance and as organizational experience reveals best practices.

Skills development also includes cultivating judgment about AI use. Technology makes many things possible—the professional skill is knowing what should be done. Training should develop not just technical proficiency but wisdom about when and how to deploy AI capabilities responsibly.

Quality Management and Continuous Improvement

Organizations should implement quality management systems ensuring AI-augmented work meets professional standards consistently.

Quality management includes verification protocols appropriate to stakes and AI application types, error tracking and pattern analysis identifying where AI performs well versus where human oversight is essential, user feedback collection revealing what works and what frustrates practitioners, and cost and value monitoring ensuring AI investments deliver returns justifying expenses.

Continuous improvement builds on quality management by using performance data to refine AI deployments—adjusting prompts and configurations based on error patterns, modifying verification intensity based on demonstrated reliability, expanding applications that prove valuable, and retiring approaches that don't justify costs.

Organizations should treat AI deployment as ongoing learning process rather than one-time implementation. Early deployments inform later ones. Experience reveals what works in organizational context. As confidence builds, AI adoption can expand into new applications.

Quality management and improvement require leadership commitment. Organizations must dedicate resources to monitoring, provide mechanisms for feedback, take corrective action when problems emerge, and reward learning from both successes and failures.

Critical Competencies for Discovery Professionals

As AI capabilities advance, discovery professionals need evolving competency sets balancing technical understanding with enhanced professional judgment.

AI Literacy Without Requiring Expertise

Discovery professionals need AI literacy—practical understanding of capabilities and limitations without requiring deep technical expertise. This literacy includes understanding how LLMs work at conceptual level to make informed decisions about deployment, recognizing hallucination risks and mitigation strategies to ensure appropriate verification, understanding context windows and processing constraints to design feasible AI workflows, and appreciating both what AI can accomplish and what remains beyond current capabilities.

© 2025 Merlin Search Technologies, Inc.

AI literacy also includes practical knowledge about implementation—what questions to ask vendors claiming AI capabilities, how to evaluate whether AI features actually provide value, what assurances about accuracy and reliability are meaningful, and what infrastructure and processes AI deployment requires.

This literacy develops through combination of education and experience. Formal training provides foundation. But hands-on experience working with AI systems builds intuitive understanding of what works, what doesn't, and why.

Importantly, discovery professionals need not become AI experts to use AI effectively. They need working knowledge comparable to their understanding of other technology—sufficient to deploy it appropriately, recognize when problems arise, and know when expert consultation is needed.

Enhanced Analytical and Synthesis Capabilities

As AI handles more routine analysis, human value increasingly lies in synthesis, judgment, and strategic thinking that AI cannot replicate. Discovery professionals should develop enhanced capabilities in these areas.

Synthesis skills involve integrating findings across multiple AI analyses into coherent understanding, recognizing patterns and relationships AI might identify separately but not connect, evaluating conflicting information from different sources to determine what's reliable, and constructing narratives explaining how events unfolded based on disparate evidence.

Strategic thinking includes determining what investigations are worth pursuing given available resources, deciding when gathered information is sufficient for strategic decisions versus when additional investigation is needed, evaluating what theories evidence supports versus what theories should be abandoned, and advising clients on risks and opportunities based on discovery findings.

These synthesis and strategic capabilities have always been important but become more critical as AI handles more analytical heavy lifting. The discovery professional's comparative advantage shifts from processing information to interpreting it wisely.

Verification and Quality Assurance Skills

With AI handling substantial analysis, verification skills become essential professional competencies. Discovery professionals need systematic approaches to verifying AI outputs without rechecking everything manually.

Verification skills include sampling strategies appropriate to stakes and AI reliability, recognizing red flags suggesting AI outputs deserve closer scrutiny, tracing AI conclusions to source evidence efficiently, and conducting targeted verification focusing on high-risk areas rather than comprehensive rechecking.

Quality assurance extends beyond individual output verification to systematic evaluation of AI performance over time—tracking accuracy trends, identifying drift where AI performance degrades, recognizing when AI struggles with particular content types, and knowing when AI

© 2025 Merlin Search Technologies, Inc.

approaches need refinement or replacement.

These verification and QA skills develop through practice. Early AI adoption typically involves intensive verification. As reliability is demonstrated, verification can become more targeted while remaining rigorous for high-stakes applications.

Communication and Explanation Abilities

Discovery professionals must explain AI-assisted work clearly to diverse audiences—clients who need to understand how conclusions were reached, courts that may question methodology, opposing parties who might challenge AI use, and colleagues who need to build on AI-assisted work.

Communication skills include explaining AI capabilities and limitations in accessible terms without oversimplifying, describing verification approaches ensuring accuracy, acknowledging uncertainty appropriately rather than overstating confidence, and relating AI methodology to familiar concepts making it comprehensible.

These explanations should be transparent without being defensive. The message is that AI enables more thorough, systematic analysis than traditional methods while appropriate oversight ensures reliability. Organizations using AI to improve discovery shouldn't apologize for technological sophistication—they should confidently explain how it enhances quality.

Explanation abilities also include responding to skepticism or concern about AI use. Some judges, opposing counsel, or clients may be uncertain about AI reliability. Discovery professionals should be prepared to address concerns directly with specific information about how AI is used, how accuracy is verified, and why AI-assisted approaches are reliable.

Ethical Reasoning in Technology Context

Ethical frameworks established in Part One apply throughout AI use, but discovery professionals need skills applying those frameworks to novel situations AI creates.

Ethical reasoning includes recognizing when AI deployment raises professional responsibility questions, evaluating competing considerations when ethics guidance is uncertain, making defensible decisions about appropriate AI use and oversight, and documenting reasoning supporting ethical decisions.

As AI capabilities advance, novel ethical situations will arise without settled guidance. Discovery professionals who can reason through these situations using established ethical principles while adapting to new technological contexts will navigate uncertainty more successfully than those waiting for explicit rules covering every situation.

Ethical reasoning also includes recognizing limits—knowing when situations require consultation with ethics counsel, when risks of AI use outweigh benefits, and when traditional approaches are more appropriate despite efficiency disadvantages.

© 2025 Merlin Search Technologies, Inc.

Measuring Success and ROI

Organizations deploying AI in discovery need frameworks for measuring whether investments deliver value.

Metrics That Matter

Success measurement should include multiple dimensions capturing different value types. Efficiency metrics track time savings—case assessment, document review, transcript analysis, timeline construction, and relationship mapping all complete faster with AI assistance. But time savings alone don't capture full value. Quality improvements—more thorough analysis, patterns identified that manual review missed, and better-supported conclusions—provide value beyond speed.

Cost metrics compare total costs including AI charges, infrastructure, training, and verification against traditional workflow costs. But cost analysis should be sophisticated—accounting for quality differences and considering strategic value alongside direct cost savings.

Strategic value metrics attempt quantifying benefits like earlier case assessment enabling better settlement decisions, more comprehensive evidence supporting stronger litigation positions, production challenges backed by solid gap analysis, and credibility maintained through systematically verified work.

User satisfaction metrics capture whether legal professionals find AI tools valuable—time saved, frustration reduced, confidence increased, and willingness to use AI for similar tasks. Poor user satisfaction suggests AI implementations aren't delivering practical value regardless of theoretical capabilities.

Comparative Analysis and Benchmarking

Measuring AI value requires baselines for comparison. Organizations should establish pre-AI baselines for key metrics—typical time for case assessment, document review costs per document, timeline construction effort, and investigation comprehensiveness.

Comparing AI-augmented workflows to these baselines quantifies improvements. But comparison should account for differences beyond just AI—AI adoption often coincides with other workflow improvements, case mix might vary across periods, and organizational learning curves affect performance independent of technology.

Industry benchmarking provides additional perspective. How do AI-augmented costs and timelines compare to industry norms? Are improvements in line with what other organizations achieve? Benchmarking helps calibrate expectations and identify whether organizations are extracting typical value from AI investments or whether performance suggests implementation problems.

Long-Term Value Assessment

Some AI benefits emerge over time rather than immediately. Organizations should evaluate both near-term returns and long-term value.

© 2025 Merlin Search Technologies, Inc.

Long-term value includes institutional learning as organizations become more sophisticated about AI deployment, process refinement improving efficiency as workflows optimize around AI capabilities, competitive advantage from capabilities enabling better client service, and risk mitigation from systematically verified work reducing error risks.

Conversely, long-term costs include ongoing infrastructure and subscription expenses, training as staff turns over and capabilities evolve, maintenance and system administration, and adaptation costs as AI technologies change.

Comprehensive value assessment accounts for full lifecycle costs and benefits rather than just initial implementation results.

The Trajectory of AI in Discovery

Understanding where AI in discovery is heading helps discovery professionals position themselves advantageously as capabilities continue advancing.

Near-Term Developments

Over the next few years, expect several developments building on current foundations. Model capabilities will continue improving—better accuracy reducing hallucination risks, larger context windows enabling processing of longer documents and more comprehensive analysis in single passes, faster processing reducing wait times for AI analysis, and improved multimodal capabilities handling images, audio, and video alongside text.

Specialization will increase as AI models are trained or fine-tuned specifically for legal applications—models that understand legal terminology and concepts better, specialized applications for particular discovery tasks, and platforms integrating multiple AI capabilities into cohesive discovery workflows.

Automation will expand as more discovery tasks can be handled autonomously with appropriate oversight—routine document categorization and tagging, preliminary timeline construction from documents and testimony, initial relationship mapping from communications, and basic production completeness assessment.

Integration will deepen as AI capabilities embed more thoroughly in discovery platforms—seamless workflows where AI assistance is natural rather than add-on, consistent user experiences across different AI applications, and unified data models supporting multiple AI capabilities.

Medium-Term Possibilities

Looking further ahead, several developments seem probable though timing remains uncertain.

Agentic capabilities will mature beyond current implementations—more sophisticated planning and reasoning enabling complex investigations with minimal human direction, better error correction allowing systems to recognize and fix mistakes autonomously, improved verification abilities where AI systems check their own work before presenting to humans, and

© 2025 Merlin Search Technologies, Inc.

coordination across multiple specialized agents handling different aspects of complex matters. Real-time analysis may become practical where AI processes documents as they're collected rather than batch processing—enabling truly continuous case assessment, dynamic timelines and relationship maps updating automatically as evidence emerges, and proactive alerting when important patterns or documents are identified.

Predictive capabilities might emerge where AI suggests what evidence should exist based on patterns learned from similar matters, predicts likely document completeness issues based on production characteristics, estimates litigation outcomes based on evidence patterns, and recommends strategic approaches supported by historical patterns.

Natural interaction will improve as AI systems understand more nuanced instructions and questions—conversational interfaces replacing structured queries, systems that clarify ambiguous requests rather than guessing intent, and adaptive interfaces that learn user preferences and adjust accordingly.

Fundamental Uncertainties

Despite clear near-term trajectories and probable medium-term developments, fundamental uncertainties remain about AI's ultimate role in discovery.

Will AI capabilities plateau or will they continue advancing toward general intelligence that would transform legal practice beyond recognition? Current consensus suggests continued advancement but probably not rapid achievement of artificial general intelligence.

How will courts and regulators respond to increasingly sophisticated AI in discovery? Will they embrace capabilities enabling more thorough discovery at lower cost, or will they impose restrictions limiting AI use? Current signs suggest general acceptance with appropriate oversight but this could evolve.

What new risks might emerge as AI capabilities advance? More capable AI might create novel problems we haven't anticipated. Organizations and the profession must remain vigilant for unexpected issues.

How will AI availability affect discovery economics and access to justice? Will AI-enabled efficiency reduce discovery costs broadly, or will it primarily benefit well-resourced parties who can afford sophisticated implementations? The answer significantly affects justice system equity.

These uncertainties counsel both enthusiasm about AI potential and humility about prediction limits. Discovery professionals should prepare for continued evolution while remaining adaptable to surprises.

Positioning for the AI-Augmented Future

Discovery professionals can position themselves advantageously for continued AI advancement by developing capabilities that remain valuable as AI handles more routine work.

© 2025 Merlin Search Technologies, Inc.

Cultivating Judgment and Strategic Thinking

As AI handles more analysis, human comparative advantage increasingly lies in judgment and strategic thinking. Discovery professionals should cultivate these capabilities deliberately.

Judgment development involves seeking diverse experiences exposing you to different matter types and challenges, learning from both successes and failures through systematic reflection, studying how experienced practitioners make difficult decisions, and practicing decision-making in contexts where feedback reveals accuracy.

Strategic thinking develops through understanding not just what happened in cases but why particular strategies succeeded or failed, recognizing patterns across matters suggesting generalizable lessons, considering multiple perspectives and alternative interpretations, and thinking several steps ahead about how decisions cascade through matter lifecycles.

These capabilities develop over time through deliberate practice rather than passively accumulating. Discovery professionals should actively work on judgment and strategic thinking rather than assuming they develop automatically with experience.

Building Technology Acumen Without Losing Focus

Discovery professionals need sufficient technical understanding to deploy AI effectively without becoming technologists who've lost focus on legal applications.

The balance involves maintaining curiosity about technological developments relevant to discovery, developing working knowledge of AI capabilities and limitations, learning enough about implementation to make informed vendor selections and deployment decisions, but resisting the temptation to become focused primarily on technology rather than legal applications.

Discovery professionals are legal professionals who use technology, not technologists who happen to work in legal contexts. The distinction matters because it keeps priorities straight—technology serves legal practice rather than being interesting for its own sake.

Developing Adaptive Capacity

Perhaps most importantly, discovery professionals should develop adaptive capacity—ability to learn continuously, adjust to changing tools and practices, maintain effectiveness despite disruption, and position for opportunities rather than just defending against threats.

Adaptive capacity includes intellectual humility recognizing that today's best practices might not remain optimal as capabilities evolve, openness to experimentation trying new approaches despite uncertainty about outcomes, resilience maintaining effectiveness despite setbacks or mistakes, and growth mindset viewing challenges as learning opportunities.

Organizations should foster cultures supporting adaptation—rewarding learning from failures rather than just celebrating successes, encouraging experimentation within appropriate boundaries, providing resources for training and development, and modeling adaptive behavior at leadership levels.

© 2025 Merlin Search Technologies, Inc.

Discovery professionals with strong adaptive capacity will thrive regardless of how AI capabilities evolve because they'll learn to leverage new capabilities as they emerge rather than clinging to approaches that become obsolete.

The Augmented Discovery Professional

The future of discovery isn't AI replacing discovery professionals—it's discovery professionals augmented by AI capabilities working together more effectively than either could work alone. Effective AI-augmented practice requires discovery professionals who understand enough about AI to deploy it appropriately while maintaining the professional judgment, ethical grounding, and strategic thinking that AI cannot replicate. It requires organizations that build institutional capabilities around AI while preserving professional standards and maintaining human accountability. It requires continued learning as AI capabilities evolve and as experience reveals what works in practice.

The chapters in Part Three demonstrated how AI transforms specific discovery workflows. This concluding chapter examined how those transformations integrate into comprehensive AI-augmented practice. Organizations and professionals embracing these capabilities while maintaining professional foundations will find themselves increasingly effective at serving clients, advancing justice, and building successful practices in an AI-augmented future.

The book concludes with Part Four, which provides practical resources supporting AI adoption—implementation frameworks, decision tools, prompt libraries, and guidance on evaluating AI vendors and platforms. These resources translate Part Three's application examples into actionable guidance supporting discovery professionals beginning their AI adoption journeys or expanding existing capabilities.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.

© 2025 Merlin Search Technologies, Inc.