

Generative AI for Smart Legal Professionals

An Introduction To Large Language Models

(Second Edition)

Complimentary
copy from

The logo for 'ils' features a stylized 'i' with a circular graphic element above it, followed by the lowercase letters 'ls' in a bold, blue, sans-serif font.

By: John Tredennick and Dr. William Webber

TABLE OF CONTENTS

Introduction.....	3
PART 1: What is Generative AI and How Does it Work?	4
What is GPT?	5
What are LLMs?	6
LLMs: Modern Supercomputers	7
Training an LLM.....	7
Training Cutoff	8
If an LLM Has No Memory, How Does it Carry on a Conversation?	10
The Role of ChatGPT	10
The Importance of Context Windows.....	11
Carrying on a Conversation	11
Context Window Size is Limited	12
Enlargements in Context Window Sizes.....	12
Is the Data We Send to GPT Secure?	14
Can an LLM Share Confidential Information, Even by Accident?.....	16
What about Hallucinations?.....	18
Reducing Chances for Hallucination.....	19
DiscoveryPartner, An Advanced RAG System	20
PART 2: Using Generative AI and Large Language Models for Investigation and Discovery	22
Using LLMs to Read, Analyze, and Report on Documents	24
Using LLMs to Answer Questions About Deposition and Hearing Transcripts	32
Deposition Summaries.....	32
Searching Across Transcripts.....	34
Conclusion.....	42
Key GenAI Terms Smart Legal Professionals Should Know.....	43
About the Authors.....	45
About Merlin Search Technologies	46

Introduction

In November of 2022, ChatGPT upended our thinking about artificial intelligence with a new form of machine learning called Generative AI (GenAI). Since then, discussions about GenAI models like GPT have taken center stage in nearly every field, especially in the legal profession. Today, most legal publications feature articles about how GenAI will change the legal profession, and vendors at legal conferences tout their GenAI-powered software.

This Guide introduces Generative AI for Smart Legal Professionals. It is designed for the smart individuals in our profession who want to learn the basics of how GenAI models like GPT work and how they can improve discovery workflow. While the underlying algorithms may be complex, we can learn enough about their function to put them to use in our practices.

The book is divided into two parts. Part 1 explores the fundamentals of GenAI and Large Language Models (LLMs), including key concepts like training, context windows, data security, and the potential for hallucinations in AI-generated content. Part 2 focuses on practical applications for investigations and discovery to demonstrate how GenAI and LLMs can streamline tasks such as document review, analysis, and transcript summarization in discovery and investigation workflows.

Our goal is to teach smart legal professionals how GenAI models operate and how to use them for more efficient and effective work. Even if you are not focused on investigations and discovery, you should find this book interesting and helpful. While the examples in Part 2 are geared towards finding information in large document sets, the capabilities can be applied to a wide range of information needs. By understanding and harnessing the power of GenAI, professionals across various domains can unlock new levels of efficiency, insight, and innovation in their work.

Let's get going. We hope you enjoy the ride.

John Tredennick

Dr. William Webber

PART 1

What is Generative AI and How Does it Work?

Generative AI is a type of artificial intelligence capable of creating new content, such as text, images, videos, audio, code, and other data. It achieves this by learning the patterns and structures from the training data it is exposed to and then using that knowledge to generate novel and original content with similar characteristics.

At the heart of Generative AI are advanced neural networks called GenAI models. These models are typically trained on vast amounts of data, often comprising billions or even trillions of text examples and other forms of media. Through this training process, GenAI models learn to recognize and replicate the underlying patterns and structures in the data, enabling them to create content that closely resembles human-generated work.

In the following sections, we will delve deeper into the workings of LLMs, exploring key concepts such as training, context windows, and the relationship between ChatGPT and GPT. We will also address important considerations like data security and the potential for hallucinations in AI-generated content. By the end of Part 1, you will have a solid foundation in the fundamentals of Generative AI and LLMs, setting the stage for the practical applications discussed in Part 2.

What is GPT?

GPT stands for "Generative Pre-trained Transformer." It is a type of artificial neural network used in natural language processing tasks that uses deep learning techniques to generate human-like text. GPT models are trained on vast amounts of diverse text data, allowing them to learn patterns and structures in natural language. They literally took the world by storm because of their ability to answer questions, create poems, analyze and summarize documents and carry on human-like conversations.

The term itself was coined by OpenAI, which gave a shortened version of that name to the first Large Language model, which it called GPT. Not long after, the term generative pre-trained transformer became generic, used to describe these newly created Large Language Models that now represent the leading edge of artificial intelligence. All the current LLMs use variants of generative pre-trained architecture.

Different versions of these LLMs are referred to by numbers, e.g. GPT 3.0, GPT 3.5 or GPT 4.0. The more recent LLM versions come with added names and sometimes a reference to the size of their context window, e.g. GPT 4.0 Turbo (128k) or the somewhat confusingly called GPT-4o (128k). We will discuss the parenthetical numbers in a minute.



What are LLMs?

Because of the extensive training required for a GenAI model, they are often called Large Language Models (“LLMs”). GPT is an LLM but there are many others on the market today including Anthropic’s Claude, Google’s Gemini, Meta’s Llama, Falcon and Mistral. Indeed, there are hundreds, if not thousands of generative pre-trained transformer models in existence today.

Some LLMs are proprietary and available only from their publisher. Others are available through open-source licenses. In the latter case, you may need to host the model yourself, or use one that is available through a cloud provider like AWS. The proprietary models are SaaS based and typically accessed over the Internet through a secure API (Application Programming Interface).

While there is much to be said about the advantages and disadvantages of these competing forms of delivery, our focus is on using LLMs for discovery workflow rather than on which LLMs and which forms of delivery are better suited for your needs.



LLMs: Modern Supercomputers

LLMs require a massive amount of computing power and run on a large collection of expensive, specialized chips called GPUs or graphical processing units. One GPU chip suitable for LLM use may cost over \$100,000. The largest LLMs like GPT 4 or Claude Opus reportedly require tens of thousands of GPU chips. Microsoft reportedly used 25,000 NVIDIA chips for its GPT 4 implementation.

These expensive chips are needed because the prediction process that makes the LLM's output so valuable is mathematically intensive, requiring a huge amount of computing power. Efforts are underway to develop models that can run on a smaller number of servers or even on a laptop. Indeed, Apple is reportedly working on models that can run, either in whole or in part, on your mobile phone.

In addition to the hardware costs, there are significant operating expenses associated with running these chips daily. Although LLM providers don't publish exact cost figures, reports suggest that operating an LLM like GPT-4o can cost over a million dollars per day. Our suggestion? Don't plan on running one of the bigger LLMs at home, at least for the time being.

Ultimately, you can think of these LLMs as supercomputers, but with a level of depth, breadth, and power unprecedented in the history of computing.

Training an LLM

The primary goal for training an LLM is to enable the model to understand, generate, and reason with natural language in a way that is coherent, contextually relevant, and useful for a wide range of applications. That requires a two-stage training process.

First, the LLM must be trained on a massive amount of mostly Internet text including books, articles, websites and other textual sources. This is often called unsupervised learning as the neural network makes connections across the training examples and draws inferences it uses in creating responses. The process allows it to “understand” grammar, context, and a wide variety of topics.

We put the word “understand” in quotes because there is an ongoing debate about whether the LLM understands anything. Some critics call it a “stochastic parrot,” arguing that while LLMs can produce content that appears coherent and contextually relevant, their output is essentially the result of statistically processing and regurgitating the vast amounts of data they have been trained on, without true understanding or consciousness.

In most cases, the training is supplemented by thousands of hours of human interaction focused on asking the model questions and providing feedback on its answers. This process of supervised training is called reinforcement learning and is critical to the model’s fluency and effectiveness.

Training an LLM is expensive. The training cost for GPT-4 is estimated to be around \$100 million, including the cloud computing costs of renting a 25,000 GPU cluster from providers like Microsoft. If cloud costs were \$1 per A100 GPU hour, the cloud expenses alone would amount to around \$60 million for a typical four-month training period.

Ultimately, the simple goal in training is to teach the model to predict what the next word in a sentence should be, based on the words that have come before. As crazy as that sounds, this is what LLMs do. They simply predict what the next words in a sentence should be based on the questions asked and the words that have gone before.

Training Cutoff

One crucial aspect to understand about LLM training is the concept of a “cutoff” or endpoint. During the training process, the model learns from vast amounts of data, continuously updating its parameters to better understand and generate language. However, once the training is complete, the model's parameters are fixed, and it can no longer learn from new data.

This training cutoff has significant implications for the model's knowledge and capabilities. Everything the model has learned up until the cutoff point becomes its permanent knowledge base. It will not be able to adapt to or incorporate any changes, events, or new information that emerge after this point. For example, if an LLM's training data cutoff is January 2023, it will not have any knowledge of events or developments that occurred after that date.

It's essential for users to be aware of an LLM's training cutoff, as it directly influences the model's understanding of the world and its ability to provide up-to-date information. When interacting with an LLM, users should keep in mind that the model's knowledge is limited to the information available up until its training cutoff, and any queries or tasks related to post-cutoff events may yield outdated or inconsistent results.

It is also important to understand that an LLM has no memory. While it can communicate and provide responses to questions or prompts, it cannot remember your conversation once the answer is returned.

For that reason, we liken the trained model to a “brain in a jar” to reflect the fact that it has no memory and cannot learn from prompts or other information submitted to it. The LLM simply takes the information it is given and responds.



GPT is like a Brain in a Jar.

These limitations have important implications for LLM security. Since an LLM cannot use your prompt to broaden its knowledge base or remember previous conversations, it also cannot inadvertently pass prompt information to other users. This means that the information you provide to an LLM remains secure and is not shared or learned by the model.

In summary, while LLMs are powerful tools for generating human-like text, their knowledge is limited by their training cutoff, and they do not have the ability to remember or learn from interactions. These characteristics, although they may seem like limitations, actually contribute to the security and privacy of the information shared with LLMs, as users can be assured that their data is not being stored or shared by the model.

If an LLM Has No Memory, How Does it Carry on a Conversation?

That is the next obvious question. Many people have experienced carrying on a conversation with ChatGPT, and the initial experience can be eerie. ChatGPT seems to converse like a human, and some users engage in lengthy discussions with the software. If GPT, the LLM behind ChatGPT, has no memory, how can that happen?

The Role of ChatGPT

First, you need to grasp the difference between ChatGPT and GPT. ChatGPT is a software application designed to facilitate communication between users and GPT, the underlying Large Language Model (LLM) that analyzes and responds to questions. It provides a browser-based interface where users can enter their queries, known as "prompts."

When a prompt is submitted, ChatGPT sends it to GPT and then returns the generated answer to the user. The "Chat" part of ChatGPT saves the conversation history, allowing users to reopen previous discussions and continue the dialogue as if no time had passed. When a new request is made within an existing conversation, ChatGPT resends all the prior communications to GPT, enabling the LLM to maintain context and provide coherent responses throughout the interaction.

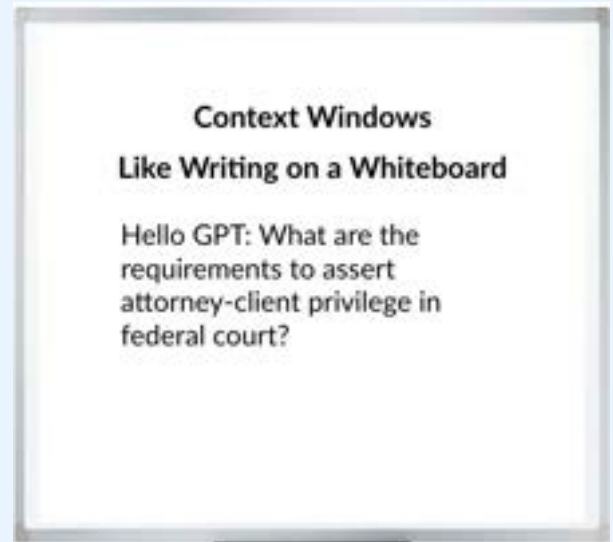
ChatGPT automatically saves your conversations, which means you can easily return to a previous discussion at any time. This feature allows you to pick up where you left off, even if you've had other conversations or taken a break in between. The saved conversations maintain the context of your earlier interactions, enabling GPT to provide coherent and contextually relevant responses when you revisit the dialogue. If you prefer not to keep a record of your conversations, ChatGPT also offers the option to delete your chat history. This gives you control over your data and ensures that your discussions with GPT remain private if desired.

The Importance of Context Windows

GPT communicates with the separate Chat application through what is called a "context window." In our discussions, we liken it to a whiteboard, one that exists outside the "brain in a jar" but is accessible to GPT and ChatGPT.

ChatGPT starts a conversation by sending the text you enter to GPT via the context window. Put another way, it writes your prompt on the virtual "whiteboard."

GPT can read what is written on the whiteboard and write its answer back. The Chat application reads the answer and returns it to us by displaying it on the browser window. Once that answer is passed back to ChatGPT, it is erased, much like a computer's RAM memory is erased when you turn it off.



A Simple Context Window (Whiteboard).

Carrying on a Conversation

Now you understand how GPT can carry on an extended conversation even though the whiteboard is erased after each response. ChatGPT keeps track of your conversation and sends the earlier parts back to GPT each time you make a new request. GPT views the entire conversation (or as much of it as can fit on the whiteboard) and uses it to carry on the discussion.

Understanding the relationship between ChatGPT, GPT, and the context window is crucial to grasping how an LLM can carry on a conversation despite lacking inherent memory. Software applications like ChatGPT (and many others) keep track of your conversation and send it to GPT (or any other LLM) so that the LLM can provide relevant and coherent responses, creating the illusion of a continuous discussion.

Context Window Size is Limited

The most important thing to know about context windows is that the amount of text you can place on them (prompt plus answer plus conversation history) is limited.

When GPT 3.5 (the original engine for ChatGPT) was first released, the context window was 4,096 tokens, which translates to about 3,000 words. (Tokens include punctuation, and some words will be split into more than one token for technical reasons beyond the scope of this book). Thus, your conversation with GPT—including both questions and answers—was limited to the size of the context window. When your conversation grew larger than the window allowed, ChatGPT would cut out the first part of the conversation so you could continue to ask new questions. If the conversation continues beyond the size of the context window, GPT will begin forgetting aspects of your earlier conversation.

It's easy to imagine that a system which can only analyze 3,000 words of text would have practical limitations. You certainly couldn't ask it to read and comment on a book or even a lengthy article. You might ask GPT about a complex tax provision but certainly not about the tax code itself. Likewise, you couldn't and still can't ask GPT to read and analyze millions (or even thousands) of your discovery documents.

Thus, don't confuse GPT or any Large Language Model with a search engine. Modern search engines can find information across millions of documents in milliseconds, but they can't analyze the documents they have found. In contrast, GenAI engines can analyze documents you bring to them but only as many as can fit in the context window.

Enlargements in Context Window Sizes

In short order, LLM context windows have increased, moving from 4 to 8 to 16 and even 32k context windows. Recently, Anthropic (founded by people from OpenAI) released a 200k version of its LLM called Claude 3 and touted its ability to read the entirety of *The Great Gatsby*, not to mention the scripts for all nine versions of the Star Wars movies. This development generated excitement as OpenAI responded with GPT 4 Turbo (128k) and then GPT-4o (also 128k). Anthropic reports that Claude 3 has a context window of

one million tokens, but we have not had access to test it yet. Google similarly reports that its LLM, called Gemini, also has a one million token context window.

These were great advances from GPT's early days (literally just months before), but there are strong suggestions that increasing the context window to substantially larger sizes may not be feasible, either technically or due to cost considerations. Even if the windows can be made larger (which they undoubtedly will be), there is currently concern that the models cannot remember everything read in large context windows, which may mean that they will overlook important details when giving their answer.

At this point, all we can say is that the larger context windows open the door to using these powerful GenAI models for a variety of applications, including for investigations and discovery.



Is the Data We Send to GPT Secure?

When ChatGPT was first released, many raised concerns about whether sending client information to the program might breach confidentiality obligations or risk a waiver of attorney-client or work-product privileges. For legal professionals at least, this is a serious question. Lawyers have ethical obligations to preserve client confidentiality, and they have a parallel duty to protect against an inadvertent waiver of attorney-client or work-product privileges.

Much of the concern arose because ChatGPT was originally released as a free public beta. As a condition of the free license, OpenAI reserved the right to use information contained in the prompts for testing and to improve later models.

Not surprisingly, legal professionals became concerned about the risk of sending confidential information to GPT, particularly about the possibility that the information might be disclosed later, either inadvertently by GPT or through its use in training new models.



In seeming response, companies like OpenAI, Microsoft, and Anthropic began offering commercial licenses for their LLMs. These licenses include written promises that the information sent to the LLMs would not be used for any purpose other than your interaction with the LLM itself.

For example, Microsoft offers this statement regarding OpenAI, which it hosts on Azure:

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 (The models are stateless, unless you explicitly fine-tune models with your training data).
- Your fine-tuned Azure OpenAI models are available exclusively for your use.

The Azure OpenAI Service is fully controlled by Microsoft; Microsoft hosts the OpenAI models in Microsoft’s Azure environment and the Service does NOT interact with any services operated by OpenAI (e.g., ChatGPT, or the OpenAI API).”

OpenAI and Anthropic similarly promise not to review communications between the user (prompt) and the system’s response. Here, for example, is the controlling provision from OpenAI’s service agreement:

We do not use Content that you provide to or receive from our API (“API Content”) to develop or improve our Services. We may use Content from Services

other than our API (“Non-API Content”) to help develop and improve our Services.

API stands for “Application Programming Interface.” It is a software intermediary that allows two programs to talk to each other without human involvement. Commercial applications typically communicate with LLMs through a secure, encrypted API.

Can an LLM Share Confidential Information, Even by Accident?

From our earlier discussions, you know that the answer is no. The LLM itself cannot learn from the prompt information you send into its knowledge base, and it cannot remember your conversations. Thus, the only risk here is that the LLM provider intercepts your communications and uses them for an illicit purpose.

In that regard, sending prompt information to a company like Microsoft is not materially different from storing files and email in Microsoft 365. The same is true for using Google office applications or storing file with Box, ShareFile, etc. While a legal discussion is beyond the scope of this book, courts have consistently held that using third party systems like these does not waive privilege so long as there is a “reasonable expectation of privacy.”¹

For example, courts have repeatedly taken the position that unencrypted email communications, even on a company server, do not result in a waiver of privilege so long as the person sending the communication had a “reasonable expectation of privacy.” E.g., *Twitter, Inc. v. Musk, C. A.* 2022-0613-KSJM (Del. Ch. Sep. 13, 2022) (Musk used Tesla/SpaceX email servers for Twitter-related legal communications); *Stengart v. Loving Care Agency, Inc.*, 990 A.2d 650 (2010) (personal legal communications made on work server). The courts did so notwithstanding the fact that Tesla and SpaceX explicitly reserved the right to inspect company emails for any purpose (including, presumably, abuse of the email privileges).

¹ We have written an article and given several webinars on the subject, one for U.S. audiences (with Professor William Hamilton of the University of Florida Law School) and one for Londoners (joined by Thomas Leyland of Dentons). You can find a copy of our August 2023 article [Are LLMs Like GPT Secure? Or Do I Risk Waiving Attorney-Client or Work-Product Privileges?](#) (Law 360, 8/17/23). You can watch [both the U.S. and London webinars here](#).

Likewise, the Standing Committee on Ethics and Professional Responsibility for the American Bar Association has repeatedly affirmed that email communications did not waive the privilege so long as the communicator had a “reasonable expectation of privacy” in the communication. [ABA Comm. on Ethics & Prof’l Responsibility, Formal Op. 17-477](#) (leaving an open question about message boards and cell phone use).

The same undoubtedly holds true for the use of web hosting services like those offered by Microsoft, Google, or AWS as well as litigation support providers. All have access to your data at one point or another, but they are under contractual obligations not to exercise that privilege except to protect their systems from abuse or misuse.

Concerns about confidentiality and waiver are understandable with any new technology. Similar fears arose with cell phones, the internet, email, and cloud productivity suites. But so long as agreements and practices support a reasonable expectation of privacy, using LLMs via a commercial license should not create meaningful privilege risk, especially with the strong security measures in place for data sent to LLMs like GPT and Claude. The tremendous potential benefits should not be sacrificed for undue concerns.

So, are you risking a breach of confidentiality or privilege waiver by using an LLM through a commercial provider? Our answer is no, at least not if you are using a commercial license for the service. Providers like Microsoft, Anthropic, and other major large language model companies include solid non-disclosure and non-use provisions in their commercial contracts. They are easily as strong as the ones included in your Office 365 licenses. And, they provide the same reasonable expectation of privacy you have when you store email and office files in Azure or AWS.

What about Hallucinations?

The last thing to cover in this part is a discussion of hallucinations. Many of us have heard the term “hallucinations” in the context of Generative AI but what are they? The concept is easy to understand but it has proven a bit unsettling for legal professionals to contemplate.

At its most basic, a hallucination is where an LLM confidently gives a detailed answer that seems plausible but is simply not true. Hallucinations typically arise because of the way LLMs are trained. LLMs are trained to predict the next word in a sequence based on the words that come before it. The LLM isn't concerned with whether the statement is true, only that it writes a convincing continuation of whatever text it is given.

As a result, LLMs can sometimes generate content that is fluent and plausible but not actually grounded in reality. The model may combine snippets of information from various sources in a way that seems coherent but is ultimately incorrect or fabricated.

A striking example of this occurred when a lawyer cited fake cases generated by ChatGPT to a court. In the filing, the lawyer cited at least six cases that did not actually exist, with fake judicial decisions, bogus quotes, and bogus internal citations.

The revelation came when opposing counsel couldn't find the cases and requested more information. Ultimately, the offending lawyer was called into court to explain the situation and faced the possibilities of sanctions (not awarded). However, since then there have been at least two other lawyers who made similar false filings and have received monetary sanctions.

These incidents highlighted the risks of lawyers relying on AI chatbots for legal research and writing without verifying the accuracy of the information provided. A simple antidote for this potential problem is to check key case citations before using them. Our lawyer author here received many memos from associates during his days as a litigation partner. He can't remember a time he made a filing or appeared in court citing an important case he had not personally read. The lesson? Always check the source material whether you receive a memo from GPT or your crack associate.

Reducing Chances for Hallucination

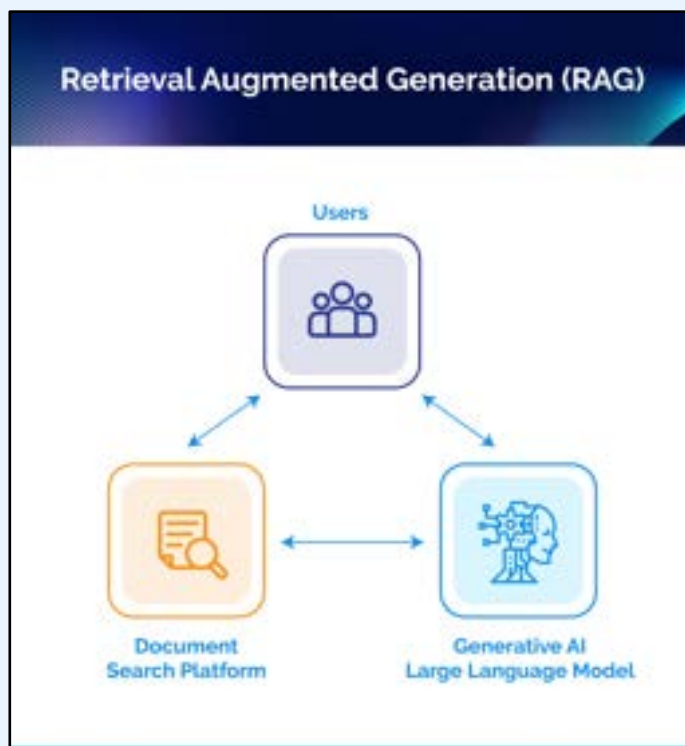
Hallucinations often occur when an LLM is asked to answer a question based on its internal training. If the training data doesn't include the specific information requested, the LLM may sometimes generate a plausible-sounding but ultimately fabricated response to fill in the gaps. This is problematic, as it can lead to the dissemination of inaccurate or misleading information.

To address this issue, developers have explored alternative approaches to using LLMs, one of which is called RAG, or Retrieval Augmented Generation. RAG combines the power of the LLM's language understanding and generation capabilities with a more targeted and controlled input process.

A RAG system looks something like this diagram.

In a typical RAG system, when the user asks a question, the first step is to search a collection of documents to find passages that are most relevant to answering it. Those passages are then inserted into the LLM's context window along with the user's original question.

The key difference between RAG and traditional LLM usage is that the LLM can be explicitly instructed to generate an answer based only on the provided passages, rather than relying on its general knowledge gained from training. By constraining the LLM to only the information present in the relevant passages, RAG helps to ensure that the generated answers are grounded in factual, verifiable information.



DiscoveryPartner, An Advanced RAG System

Merlin's DiscoveryPartner is an advanced RAG system and works much like described above. Not surprisingly, we take several steps to reduce or eliminate the potential for hallucinations including the following:

1. For every prompt sent to an LLM, we instruct the system to base its answers solely on the documents provided to it and to never answer based on its internal information.
2. We give that instruction twice, once at the beginning of the prompt and once at the end, to ensure that the LLM does not forget the instruction in crafting its answer.
3. We also instruct the LLM to back up every important point in its answer with a link to one or more source documents. With a click, the user can access the source document summary which, in turn, has links to the source document itself. This allows the user to quickly check the source for all of the LLM's answer.

While RAG systems can significantly reduce the occurrence of hallucinations, they are not infallible. Incorrect answers can still occur if the retrieved passages themselves contain inaccurate information. However, by employing strict anti-hallucination methodologies like those used in DiscoveryPartner, the risk of hallucinations can be minimized even further.

While RAG is not a foolproof solution to eliminating hallucinations, it significantly reduces their occurrence by limiting the LLM's response to the content of the provided passages. This approach promotes more accurate and trustworthy outputs, as the LLM is less likely to introduce fabricated information when its knowledge base is restricted to the relevant, curated content.

As RAG and similar techniques continue to evolve, we can expect to see further improvements in the accuracy and reliability of LLM-generated responses. By staying at the forefront of these developments and continuously refining our approaches, we can ensure that tools like DiscoveryPartner remain indispensable assets for legal professionals in their quest for truth and justice.

As we conclude Part 1, we have gained a deeper understanding of the fundamentals of Generative AI and Large Language Models, including their potential to revolutionize legal processes and the importance of addressing concerns like hallucinations. By leveraging advanced techniques like Retrieval Augmented Generation and strict anti-hallucination methodologies, legal professionals can harness the power of LLMs while maintaining the high standards of precision and reliability demanded by the legal profession. With this foundational knowledge, we are now well-equipped to explore the practical applications of Generative AI and LLMs in discovery and investigation workflows in Part 2.

PART 2

Using Generative AI and Large Language Models for Investigation and Discovery

In Part One, we explored the fundamentals of Generative AI and Large Language Models, gaining a deeper understanding of how these technologies work and their potential implications for the legal profession. With this foundation in place, we can now delve into the heart of our subject: Using GenAI and LLMs to improve investigation and discovery workflows.

Imagine a scenario where a discovery professional is tasked with reviewing thousands of documents to identify key information for a complex case. Traditionally, this process would involve hours of manual review, searching for relevant keywords and phrases, and carefully analyzing each document for potential relevance. But what if there was a way to streamline this process, leveraging the power of AI to quickly identify the most pertinent information and provide valuable insights?



In Part Two, we will explore several real-world examples that demonstrate how GenAI and LLMs can be used to enhance efficiency, accuracy, and cost-effectiveness in investigation and discovery. From automated document classification and summarization to advanced search and analysis capabilities, we will showcase the transformative potential of these technologies in the hands of smart legal professionals.

While the possibilities for applying GenAI and LLMs in legal practice are vast, our focus will be on the most impactful and practical applications in the context of investigation and discovery. By providing concrete examples and step-by-step guidance, we aim to equip you with the knowledge and inspiration needed to start implementing these tools in your own work.

The insights gained from these examples extend far beyond the realm of discovery. The principles and strategies discussed in this section can be adapted and applied to a wide range of legal tasks, from due diligence and contract review to legal research and analysis. By understanding how GenAI and LLMs can be leveraged to streamline and enhance these processes, legal professionals across various practice areas can unlock new levels of efficiency and value for their clients.

So, whether you're a seasoned discovery professional looking to stay ahead of the curve or a legal practitioner seeking innovative ways to improve your workflow, this section will provide you with the insights and practical guidance needed to harness the power of GenAI and LLMs in your practice. Let's dive in and explore the future of investigation and discovery in the age of artificial intelligence.

Using LLMs to Read, Analyze, and Report on Documents

Let's begin our discussion by looking at how an LLM can assist us in one of the most basic components of a discovery process—reading, analyzing, and reporting on discovery documents. This is a process normally reserved for carbon-based units, often those holding a law degree. Before the advent of LLMs, few thought AI would advance to where it could match humans for these kinds of tasks.

That is no longer the case. LLMs like GPT, Claude, and others can positively assist in this process, saving both time and legal fees in the bargain. While many in our industry remain skeptical about using LLMs for document analysis and review, we can show you how the process works, and the results which can be obtained using DiscoveryPartner, Merlin's LLM-integrated platform for investigations and discovery.

The Assignment

Assume our investigation is focused on this topic, one which has been used during several of the annual text retrieval conferences (TREC) put on by our National Institute of Standards and Technology (NIST).

Eminent Domain—All documents concerning the legality or morality of expropriating land in Florida for commercial development.

Further assume that our investigation is focused on several million emails made public by Governor Jeb Bush during his two terms as governor of Florida. Lastly, assume that we have identified several hundred documents that seem to be relevant for our investigation.

How might we determine if the documents are indeed relevant and, if so, how they relate to our investigation? In the past, the answer was to have legal professionals review the documents and report on their findings. This was a tedious process at best, taking time and costing substantial amounts in legal fees proportional to the number of documents needing to be reviewed.

With an LLM-integrated platform like DiscoveryPartner, this work can be done in minutes rather than hours or days, at a fraction of the cost for human review. Here is how the process might work:

1. Create a Prompt: A prompt is the starting point for assigning work to an LLM. While we could write an entire book on the art of prompting, let's just focus on this simple prompt which we can use to set the wheels in motion:

We are interested in learning more about this topic: Eminent Domain—All documents concerning the legality or morality of expropriating land in Florida for commercial development.

What are the key issues pro or con being debated in the documents regarding this topic, and who is raising these issues?

2. Read, Analyze, Rank, and Summarize the Documents: When we run the prompt, the system identifies the top 100 most relevant documents from our collection and tasks an LLM with reading, analyzing, ranking and summarizing each. Here is an LLM-generated summary of one of the more highly ranked documents. We are using GPT-4o for this exercise. The unedited, LLM-created text is in blue.

Bush920808 | Anna Current's Email on Riviera Beach Eminent Domain Controversy |

Rel: 100%

The document is an email sent by Anna Current on December 1, 2005, to a list of recipients including Earl Mallory, Elizabeth Moser, Frank Schnidman FAU, Garry Schickedanz, Greg Kino, Heinz and Ingrid Goerke, Jack Craft, Jay & Louise Morrish, Jim Burling, Jimmy Burg, John Parsons, John Slattery, John Taylor, Josh and Melissa White, with Jeb Bush, Jeff Atwater, and Joe Negron cc'd [¶1]. The subject of the email is "EMINENT DOMAIN--For the Well-to-do!" [¶1].

The email references an article from the Los Angeles Times about Riviera Beach, Florida, where the local government plans to displace about 6,000 residents to build a yachting and residential complex [¶2, ¶6]. The plan, led by Mayor Michael D. Brown, aims to "forever change the landscape" and boost the local economy [¶9]. However, opponents argue that it is a government-sanctioned land grab benefiting private developers and the wealthy [¶10-¶12].

State Rep. Ronald L. Greenstein and Carol Saviak, executive director of the Coalition for Property Rights, criticize the plan as unfair and morally wrong [¶12, ¶14]. Scott G. Bullock from the Institute for Justice highlights the project's large scale and the number of people it displaces [¶15]. The email also mentions a recent U.S. Supreme Court ruling that has led to legislative efforts to restrict eminent domain seizures [¶16].

In Florida, local officials can take private land for redevelopment if it is deemed "blighted" [¶17]. Martha Babson, a resident, conducted her own survey and found errors in the city's official study that classified properties as dilapidated or vacant [¶18-¶19]. Other residents like Rene Corie and Dee Cunningham express distress and opposition to the plan, with Cunningham comparing the situation to living under constant threat [¶20-¶22, ¶26-¶27].

The redevelopment project involves significant changes, including moving U.S. Highway 1 and creating an artificial lagoon [¶29]. Viking Inlet Harbor Properties has been chosen as the master developer [¶30]. Affected residents are supposed to receive new homes and compensation, but uncertainties remain [¶31-¶32]. Business owners like Bill Mars and Mike Mahoney accuse local leaders of neglecting city codes to justify redevelopment [¶32-¶35].

Some residents have attended seminars to learn how to fight the plan, while others have accepted offers or retained lawyers [¶36-¶37]. Babson hopes that legislative action and public interest will halt the developers [¶38-¶39].

Key issues debated include the morality and fairness of displacing residents for commercial development, the accuracy of blight designations, and the adequacy of compensation for affected residents. Proponents argue the plan will boost the local economy, while opponents see it as benefiting the wealthy at the expense of the poor.

The LLM-generated summary is remarkably comprehensive and accurate, especially considering the speed and cost at which it was produced. In mere seconds and at a fraction of the cost of human labor, the LLM has not only effectively distilled the key points of the underlying email but also seamlessly integrated links to the original text.

This allows for easy navigation between the summary and the source material, as demonstrated in the following screenshot.

The redevelopment project involves significant changes, including moving U.S. Highway 1 and creating an artificial lagoon [¶29]. Viking Inlet Harbor Properties has been chosen as the master developer [¶30]. Affected residents are supposed to receive new homes and compensation, but uncertainties remain [¶31-¶32]. Business owners like Bill Mars and Mike Mahoney accuse local leaders of neglecting city codes to justify redevelopment [¶32-¶35].

Some residents have attended seminars to learn how to fight the plan, while others have accepted offers or retained lawyers [¶36-¶37]. Babson hopes that legislative action and public interest will halt the developers [¶38-¶39].

Key issues debated include the morality and fairness of displacing residents for commercial development, the accuracy of blight designations, and the adequacy of compensation for affected residents. Proponents argue the plan will boost the local economy, while opponents see it as benefiting the wealthy at the expense of the poor.

[View Text](#) | [View Document](#)

2024-06-09 19:27 | GPT 4o [Copy](#)

- ¶36 Some foes of the redevelopment plan have attended seminars in Washington organized by property-rights advocates to learn how to better fight to save their homes.
- ¶37 Some residents have accepted offers from developers and moved out; others have retained lawyers to try to get a better price from the city. Still others are waiting to see what happens, noting the troubled history of local redevelopment efforts. "This is the fourth eminent domain CRA plan I've seen since I've been here," said Mars. "I survived those, and I may survive this one too."
- ¶38 Babson said she was counting on the Florida Legislature, as well as public interest kindled by the recent Supreme Court case, to halt the developers.
- ¶39 "We're definitely in Tiananmen Square: one little guy in front of all of those tanks," Babson said. "We've slowed them down, but we haven't stopped them."

3. Have the LLM Synthesize the Documents, and Generate a Report

Once the documents have been reviewed, ranked and summarized, the next step is to prepare a report, synthesizing information contained in the top 100 documents. Here is the report created in response to our initial prompt. Again, no human editing has taken place for the following. The BegControl references are linked to the source for viewing.

The topic of eminent domain in Florida, particularly concerning the legality and morality of expropriating land for commercial development, has generated significant debate. The key issues being debated, along with the individuals and organizations raising these issues, are as follows:

Key Issues

1. Legality and Constitutionality:

- **Public Use vs. Private Development:** A central issue is whether eminent domain should be used for private commercial development or strictly for public use. Critics argue that using eminent domain for private gain violates the Fifth Amendment, which stipulates that private property can only be taken for public use with just compensation (Bush1472500, Bush1649676, Bush1250438).
- **Broad Definitions of Blight:** The broad and often vague definitions of "blight" used to justify property seizures are criticized for enabling abuse. For example, in Florida, areas can be deemed blighted for reasons such as "diversity of ownership," which can be easily manipulated (Bush852127, Bush1757638).

2. Morality and Ethics:

- **Impact on Homeowners and Small Businesses:** The moral implications of displacing homeowners and small businesses for commercial projects are a significant concern. Many argue that this practice disproportionately affects less affluent communities and benefits wealthy developers (Bush1649676, Bush1538656, Bush920808).
- **Fair Compensation:** There is debate over whether the compensation offered to property owners is fair and reflects true market value. Critics argue that assessed values are often lower than market values, leading to inadequate compensation (Bush1250438, Bush1203365).

3. Economic and Social Impact:

- **Community Displacement:** The social impact of displacing entire communities for commercial development is a major issue. Projects like the Riviera Beach redevelopment plan, which aimed to displace 6,000 residents, highlight the potential for significant community disruption (Bush1461380, Bush920808).
- **Economic Development vs. Property Rights:** Proponents of using eminent domain for commercial development argue that it can lead to economic revitalization and increased tax revenues. However, opponents contend that these benefits do not justify the infringement on property rights (Bush1649676, Bush1575796).

4. Legislative and Judicial Responses:

- **Legislative Measures:** Various legislative efforts have been made to curb the abuse of eminent domain. For instance, House Bill 1567 was signed to restrict the use of eminent domain for transferring private property to another private entity (Bush1649676, Bush1539144, Bush1575796).
- **Judicial Precedents:** The U.S. Supreme Court's decision in *Kelo v. City of New London*, which allowed the use of eminent domain for private development, has been a focal point of criticism and has spurred legislative responses at the state level (Bush1649676, Bush1250438, Bush1538656).

Key Individuals and Organizations

1. **Governor Jeb Bush:** As the Governor of Florida, Jeb Bush has been a central figure in the debate, expressing concerns about the misuse of eminent domain and supporting legislative measures to protect property rights (Bush1649676, Bush1539144, Bush1575796).
2. **Warren Davis:** Representing the Office of Citizens' Services, Davis has frequently communicated on behalf of Governor Bush, addressing concerns raised by citizens and outlining legislative actions taken to curb eminent domain abuse (Bush1649676, Bush1539144, Bush1575796).

- 3. Institute for Justice:** This libertarian public-interest law firm has been a vocal critic of eminent domain abuse, documenting numerous cases and advocating for stronger property rights protections (Bush1757638, Bush1203361).
- 4. Local Government Officials:** Various local officials, such as those in Riviera Beach and Hollywood, have been involved in controversial eminent domain cases, often facing criticism for prioritizing commercial development over community interests (Bush1461380, Bush897090).
- 5. Affected Residents and Property Owners:** Individuals like Sharon White, Wendy Zamecnik, and Robert Barratt have raised personal concerns about the impact of eminent domain on their properties and communities, highlighting the human cost of such practices (Bush1250438, Bush1203365, Bush1649682).

In summary, the debate over eminent domain in Florida centers on the balance between economic development and the protection of property rights, with significant concerns about the legality, morality, and social impact of using eminent domain for commercial purposes. Key figures and organizations have raised these issues, advocating for legislative and judicial reforms to safeguard property owners.



This report is only the start of the process. Once you have identified documents for review, an LLM can quickly create different types of reports that can help you better understand your subject matter.

Here is the top portion of a timeline/chronology created in seconds:

Sherlock LLM

Here is the timeline of key events in table format:

Date	Person Involved	Summary of Event	Document References
May 6, 2003	Peter H. Colt	Discusses misuse of eminent domain in Daytona Beach, FL, where properties are targeted for private development.	Bush1757638
May 8, 2003	Brian James Tesoriero	Expresses concerns about potential bias in educational standards at the University of South Florida.	Bush1756788
June 23, 2005	Munsell McPhillips, Ph.D.	Expresses concern over the Supreme Court decision in Kelo v. City of New London and urges legislative action.	Bush1207821
July 12, 2005	Raquel Rodriguez	Discusses need for legal changes to restrict eminent domain for private development, referencing Hollywood, FL.	Bush897090
July 29, 2005	Wendy Zamecnik	Details personal experience with eminent domain abuse in Palm Beach County for a golf course development.	Bush1203365
August 3, 2005	Kathy Stone	Expresses concern about eminent domain for private development and urges preventive measures similar to Alabama.	Bush1241689
December 2, 2005	Sharon White	Criticizes misuse of eminent domain in Riviera Beach, FL, and urges legislative action to protect landholders' rights.	Bush1250438
April 14, 2006	"jefevolcan"	Opposes Vanderbilt Beach Road Extension in Collier County, arguing it benefits developers.	Bush1453592
April 24, 2006	Chanta G. Combs	Provides update on legislative status of eminent domain issues in Florida House and Senate.	Bush1575796
May 11, 2006	Governor Jeb Bush	Signs House Bill 1567 to restrict use of eminent domain for transferring private property to another private entity.	Bush1649676 , Bush1539144 ,

Each of the blue links will quickly open the supporting summary and provide quick access to the underlying source documents.

The total time taken to identify, read, analyze, and report on this 100-document investigation sample was about two minutes. While there is a lot more to be said about the utility of using an LLM to review documents, you can see from this example that

LLMs are both fast and powerful. They can be quite helpful in analyzing and reviewing documents for both investigation and discovery.

Using LLMs to Answer Questions About Deposition and Hearing Transcripts

LLMs can also be a game changer for reviewing and summarizing deposition and hearing transcripts. With a properly configured discovery platform, LLMs can not only summarize transcript testimony, but they can also answer questions across hundreds of transcripts and do so in seconds. Current transcript software can run keyword searches across transcripts, but it cannot answer your questions about witness testimony.

Deposition Summaries

The traditional approach for dealing with transcript testimony is to create a summary, typically in a Word or PDF format. These projects are often given to associates or senior legal assistants who read the transcript and summarize it as they go. One of the authors regularly created deposition summaries when he was a trial lawyer, dictating key points of the testimony along with page and line numbers.

A typical deposition summary might look something like this:

Leaking Barrels and Chemical Exposures at NECC	
36:23-37:1	Samples were taken during the filtering process and sent to the lab to ensure the product met the required quality before proceeding to the next stage.
37:2-37:11	After filtering, the reserve salt went through a drying process on a conveyor belt, similar to a potato chip dryer, and came out as a powder that was put into drums and shipped to customers for use as a metal stripper.
37:16-38:3	Mr. Nadeau recalls Metro-Atlantic also made other products like 40-5 that went through a similar filtering process as reserve salt minus the drying, and products for waterproofing fabrics under the brand name Rane-Pel.
38:4-38:10	Various raw materials were used to make the waterproofing products, including fish oil, methanol, potash, waxes, caustic acid, and formaldehyde.
38:19-38:23	Mr. Nadeau acknowledges his memory may be somewhat contaminated from hearing other people's testimony, but most of it is his own memory.
39:1-40:3	Referring to his affidavit, Mr. Nadeau confirms that incoming barrels were stored on the ground at New England Container, some emitting chemical odors, and materials from the barrels leaked onto the soil.
40:23-41:19	The barrels were typically stacked on their sides in long rows, sometimes stacked on top of each other. NECC would receive hundreds of dirty barrels from various places for reconditioning.
42:2-42:7	Mr. Nadeau himself perceived and observed the leaks from the barrels, standing in the leaked material.

With an LLM, you can extend the process further, having the LLM include an overall summary of the deposition as we show here:

MERLIN[☆]

Deposition Summary

Joseph Nadeau

Case name: Emhart Industries, Inc. vs. New England Container Company, Inc., Providence Washington Insurance Company and Travelers Casualty & Surety Company, Consolidated with Emhart Industries, Inc. vs. United States Department of the Air Force, et al.

Date: Wednesday, June 12, 2013

Witness: Joseph Nadeau

Location: Offices of Duffy & Sweeney, LTD, 1800 Financial Plaza, Providence, Rhode Island

Overall Summary

In his deposition, Joseph Nadeau recounts his employment at New England Container Company (NECC) and Metro-Atlantic Chemical Company in the 1960s. He worked summers at NECC reconditioning drums that contained chemical residues, which often caused skin irritation and nosebleeds. Nadeau also worked briefly at Metro-Atlantic, where he was involved in manufacturing corrosive products like reserve salt using hazardous raw materials.

Nadeau describes the processes at both companies, including the use of dangerous chemicals, the generation of waste, and the disposal practices. At NECC, leaking barrels were stored on the ground, and workers were exposed to the leaks while handling them. The drum reconditioning process involved burning off residues, sandblasting, and steam cleaning. At Metro-Atlantic, the manufacturing processes generated black sludge waste that was partially hosed into floor drains leading to the canal. Leaks and spills from pipes and equipment were routinely washed down the drains into the "tailrace" as part of daily operations.

Nadeau also discusses the layout of the facilities, identifying various buildings and their functions using aerial photographs and maps. He mentions a small outbuilding called the Texas Tower at Metro-Atlantic, which he believes was used to produce a specialty product, possibly hexachlorophene, though he had little direct involvement with it.

Throughout his testimony, Nadeau emphasizes the pervasive chemical odors, the hazardous working conditions, and the potential environmental impact of the waste disposal practices at both NECC and Metro-Atlantic during his employment in the 1960s.

Deposition of Joseph Nadeau (publicly available on the EPA website)

We can then provide a hyperlinked table of contents with summaries and statements that distill key information from the deposition and provide instant links to the source testimony.

Chemical Raw Materials and Leaks at Metro-Atlantic

Raw materials arrived at Metro-Atlantic in various forms and were stored in above-ground tanks outside the building. The tanks were connected to the building via a network of pipes and valves, which allowed fluids to be transferred inside. Leaks and spills occurred from the pipes due to corrosion, and the spilled material was hosed into floor drains leading to the tailrace.

69:3-69:17	Raw materials arrived at Metro for use in manufacturing chemicals via delivery truck or tanker in various packaging, including bulk pumped out of tankers, 55-gallon metal and fiber drums, and bags.
69:18-70:25	There were several above-ground storage tanks (three to six) at Metro on the outside wall over the wet area/tailrace, connected to the building via pipes. The tanks varied in size and some had large, heavy submarine tank covers that locked down with huge locking nuts.
71:1-71:20	Fluids were transferred from storage tanks to the Metro building via a pipe system with valves that allowed flow or transfer of tank contents to the desired destination, sometimes under pressure or gravity feed.
71:21-72:11	The storage tanks were connected to the Metro building by steel pipes with valves (quick valves and heavy turn valves) and flex pipes. The pipes allowed directing product from one point to another through a series of valves.
72:12-73:1	Leaks and spills occurred from the pipes, typically at unions, valves, or elbows, due to wear and corrosion from the harsh products like reserve salt. When leaks were noticed, the system would be shut down and a maintenance man would replace the leaking section.
73:2-73:18	The pipes eroded from the harshness of the contents like reserve salt, which would eat holes in tanks and mixing blades. Eroded pipes were observed where leaks occurred.
73:21-74:13	Leaks occurred inside the Metro building from pipes transporting raw materials from outside storage tanks. The leaks were noticeable when the product was moved under air pressure. Management was observant of costly leaks.

Deposition and hearing summaries are a standard way to extract information from their associated transcripts, but they are costly and time consuming to prepare. The summary excerpts you see above were created by Claude 3 Opus in minutes, at a fraction of the cost for a human to do the job.

Searching Across Transcripts

Once we mastered creating traditional deposition summaries, the next goal was to allow a user to search across multiple transcripts to get answers to questions about

witness testimony. Our DiscoveryPartner system will allow users to create prompts that ask questions across hundreds of transcripts and get answers rather than search hits.

Based on the sections retrieved, the LLM will provide answers to questions about testimony during the deposition or hearing. Here is an example using two volumes of the Nadeau transcripts which provided the basis for the earlier deposition summaries.

Here is a more complicated prompt we will use for our investigation (which was, in part, created by an LLM itself).

[Prompt] I am investigating the environmental and safety practices related to the handling and disposal of industrial drums and barrels at the Metro-Atlantic and New England Container Company (NECC) facilities. These practices have been highlighted in witness testimony from Joseph Nadeau, a former employee, who provided detailed accounts of how these facilities managed unusable drums and the disposal of hazardous materials. Key points from Nadeau's testimony include:

- The sale of un-reconditionable drums as “burn barrels” for burning trash, indicating a potential environmental hazard.
- The observation of liquid waste from French drains at the Metro-Atlantic facility leading outside, with uncertainty about its final destination.
- The designation of specific areas within the facilities for drum storage, painting, washing, and burning, suggesting structured but potentially unsafe handling practices.
- The existence of a dump area where drums may have been improperly discarded.
- The disposal practices for residue from drum cleaning processes and the handling of plastic liners removed from drums at NECC.

Given these points, I am looking for additional information, documents, or testimonies that can shed light on:

1. Any environmental impact assessments or reports related to the practices at these facilities.

2. Regulatory actions, fines, or citations issued against Metro-Atlantic or NECC for environmental or safety violations.
3. Testimonies or statements from other employees, residents, or environmental experts regarding the impact of these practices.
4. Measures taken by Metro-Atlantic or NECC to address environmental and safety concerns, if any.
5. Historical context on the operation periods of these facilities and any known incidents of pollution or health hazards.

This research aims to understand the broader implications of the drum and barrel handling and disposal practices at Metro-Atlantic and NECC, including their compliance with environmental regulations and the potential risks posed to public health and safety.

In seconds, GPT-4o provided this answer:

Comprehensive Report on Environmental and Safety Practices at Metro-Atlantic and NECC Facilities

Introduction

This report investigates the environmental and safety practices related to the handling and disposal of industrial drums and barrels at the Metro-Atlantic and New England Container Company (NECC) facilities. The practices have been highlighted in witness testimony from Joseph Nadeau, a former employee, who provided detailed accounts of how these facilities managed unusable drums and the disposal of hazardous materials.

The key points from Nadeau's testimony include:

- The sale of unreconditionable drums as “burn barrels” for burning trash, indicating a potential environmental hazard.
- The observation of liquid waste from French drains at the Metro-Atlantic facility leading outside, with uncertainty about its final destination.
- The designation of specific areas within the facilities for drum storage, painting, washing, and burning, suggesting structured but potentially unsafe handling practices.

- The existence of a dump area where drums may have been improperly discarded.
- The disposal practices for residue from drum cleaning processes and the handling of plastic liners removed from drums at NECC.

Given these points, this report seeks to shed light on:

- Any environmental impact assessments or reports related to the practices at these facilities.
- Regulatory actions, fines, or citations issued against Metro-Atlantic or NECC for environmental or safety violations.
- Testimonies or statements from other employees, residents, or environmental experts regarding the impact of these practices.
- Measures taken by Metro-Atlantic or NECC to address environmental and safety concerns, if any.
- Historical context on the operation periods of these facilities and any known incidents of pollution or health hazards.

Key Issues and Practices

1. Sale of Unreconditionable Drums as “Burn Barrels”

Joseph Nadeau testified that drums that could not be reconditioned were sold as "burn barrels" for burning trash. This practice indicates a potential environmental hazard due to the improper disposal of industrial drums. The sale of these drums for burning could lead to the release of hazardous substances into the environment, posing risks to air quality and public health (JN_00000001 §5, JN_00000001 §6).

2. Liquid Waste from French Drains

Nadeau observed that liquid waste from French drains at the Metro-Atlantic facility led outside, but he was uncertain about its final destination. This raises concerns about the potential contamination of surrounding areas, including soil and water sources. The lack of clarity on where the waste ended up highlights the need for proper waste management and environmental monitoring (JN_00000001 §7, JN_00000002 §8).

3. Designation of Specific Areas for Drum Handling

The facilities had designated areas for drum storage, painting, washing, and burning. While this suggests a structured approach, the practices were potentially unsafe. For instance, the paint station, burner area, and wash station were all part of the drum handling process. The structured handling practices, although organized, may not have adhered to safety and environmental standards, leading to potential hazards (JN_00000001 §5, JN_00000002 §14).

4. Existence of a Dump Area

Nadeau mentioned a dump area where drums may have been improperly discarded. This area was referred to as the "dump area" by others, though Nadeau himself did not use this term. The existence of such an area suggests that there may have been improper disposal practices, leading to potential environmental contamination and health risks (JN_00000001 §6, JN_00000002 §17).

5. Disposal Practices for Residue and Plastic Liners

The residue from drum cleaning processes and the handling of plastic liners removed from drums at NECC were disposed of in ways that could pose environmental risks. For example, plastic liners were sometimes temporarily dumped on the ground before being collected. The improper disposal of these materials could lead to soil and water contamination, posing risks to the environment and public health (JN_00000001 §8, JN_00000002 §10).

Additional Testimonies and Observations

Handling of Hazardous Materials

Nadeau mentioned the use of various hazardous materials at the Metro-Atlantic facility, including sulfuric acid, caustic, and nitrobenzene. These materials were used in the production of products like "reserve salt," a metal stripper. The handling of these chemicals posed significant safety risks to workers and the environment. The use of such hazardous materials necessitates strict adherence to safety protocols and environmental regulations to prevent accidents and contamination (JN_00000002 §10).

Incidents and Safety Concerns

Nadeau recounted several incidents that highlighted the safety concerns at the facilities. These included a major fire involving a methanol tank, smaller fires, and an explosion that damaged the building. These incidents underscore the potential hazards associated with the handling and disposal practices at the facilities. The occurrence of such incidents indicates the need for improved safety measures and emergency response protocols (JN_00000002 §13).

Waste Disposal Practices

The waste disposal practices at the facilities included the use of French drains to dispose of liquid waste. Nadeau observed that the liquid waste from the cleaning processes flowed into these drains and eventually led to a "wet spot" or "tailrace" at the back of the property. Additionally, the disposal of plastic liners from drums was not well-managed, with liners sometimes being temporarily dumped on the ground before being collected. These practices highlight the need for proper waste management systems to prevent environmental contamination (JN_00000002 §8, JN_00000002 §10).

Recommendations for Further Investigation

1. Conduct Environmental Impact Assessments

Comprehensive assessments should be conducted to evaluate the environmental impact of the handling and disposal practices at the Metro-Atlantic and NECC facilities. These assessments should include soil and water testing to identify potential contamination and its sources.

2. Review Regulatory Actions

Investigate any regulatory actions, fines, or citations issued against the facilities to understand the extent of compliance with environmental and safety regulations. This review should include an examination of historical records and any enforcement actions taken by regulatory agencies.

3. Gather Additional Testimonies

Collect testimonies from other employees, residents, and environmental experts to gain a broader perspective on the impact of the practices. These testimonies can provide valuable insights into the day-to-day operations and potential risks associated with the facilities.

4. Implement Safety Measures

Ensure that the facilities implement appropriate safety measures to address the environmental and safety concerns identified in the testimonies. This may include upgrading waste management systems, improving handling practices for hazardous materials, and enhancing emergency response protocols.

5. Historical Analysis

Conduct a historical analysis of the operation periods of the facilities and any known incidents of pollution or health hazards to provide context for the current investigation. This analysis should include a review of historical documents, incident reports, and environmental studies.

Conclusion

The environmental and safety practices at the Metro-Atlantic and NECC facilities, as described by Joseph Nadeau, indicate significant concerns regarding the handling and disposal of industrial drums and hazardous materials. The lack of regulatory oversight and measures to address these concerns further exacerbates the potential risks to public health and the environment. Further investigation, including additional testimonies and detailed environmental impact assessments, is necessary to fully understand the implications of these practices and ensure compliance with environmental regulations.

By addressing these recommendations, a more comprehensive understanding of the environmental and safety practices at the Metro-Atlantic and NECC facilities can be achieved, ultimately leading to improved compliance and reduced risks to public health and the environment.

Key People and Organizations

- **Joseph Nadeau:** Former employee and key witness providing detailed testimony.
- **Metro-Atlantic and New England Container Company (NECC):** Facilities involved in the handling and disposal of industrial drums and barrels.
- **Emhart Industries, Inc.:** Involved in a legal case related to the facilities (JN_00000002 §1).

Dates

- **1960s:** Period during which Joseph Nadeau worked at the facilities (JN_00000001 §1, JN_00000002 §1).

This comprehensive report is based on the provided document summaries and focuses on the key issues, practices, and testimonies related to the environmental and safety practices at the Metro-Atlantic and NECC facilities. Further detailed documents and testimonies would be required to gain a complete understanding of the situation.

Once again, this summary of information was created in minutes, this time based on information quickly obtained across 150 transcripts.

Conclusion

Throughout this book, we have explored the fundamentals of Generative AI and Large Language Models, delving into their inner workings, capabilities, and limitations. Our primary goal was to provide smart legal professionals with the knowledge and tools necessary to harness the power of GenAI and LLMs in their investigation and discovery practices.

By focusing on several discovery workflow examples, we sought to demonstrate the transformative potential of GenAI in streamlining and enhancing critical tasks such as document review, analysis, and transcript review. From automating the classification and summarization of documents to extracting key insights from vast amounts of data, the integration of LLMs like GPT into discovery workflows marks a significant advancement for the profession.

As we have seen, LLMs can dramatically improve the efficiency and accuracy of these tasks, enabling legal teams to quickly identify relevant information and make better use of their time and resources. This, in turn, allows legal professionals to devote more attention to high-value activities such as developing trial and settlement strategies, exercising professional judgment, and providing sound advice to clients.

The promise of GenAI extends far beyond simply making existing processes more efficient. This technology has the potential to fundamentally reshape the very nature of legal work, opening new possibilities and redefining the contours of the profession. As we stand at the threshold of this new era, it is up to smart legal professionals to seize the opportunities presented by GenAI.

By embracing the capabilities of LLMs and continuously exploring innovative applications, legal professionals can not only adapt to this paradigm shift but also actively drive the transformation of legal practice. Leveraging the power of GenAI, we can unlock new levels of efficiency, insights, and better outcomes for clients. Together, let us boldly shape the future of legal technology and redefine the boundaries of what is possible in the realm of discovery and beyond.

Key GenAI Terms Smart Legal Professionals Should Know

Here are several terms smart people should know about Generative AI. These concepts are at the heart of this new form of artificial intelligence and will help you better understand our subject.

- **API (Application Programming Interface):** A set of protocols, routines, and tools for building software applications. In the context of GenAI, APIs allow users to interact with and access the capabilities of LLMs through a defined set of commands and inputs.
- **Generative AI (GenAI):** A type of artificial intelligence that can generate new content, whether it's text, images, music, or other forms of media, based on its training and the input it receives. This is accomplished through machine learning models that have been trained on large datasets, enabling them to recognize patterns, styles, or structures in the data.
- **GPT (Generative Pretrained Transformer):** A form of GenAI designed to understand, process, and generate human-like text based on the input it receives. As a legal professional, think of it as an advanced legal assistant or associate that can help with some pretty complex reading, analyzing, and writing tasks.
- **ChatGPT:** A web-based application that allows users to interact with GPT (i.e., send information through prompts) and receive answers. It runs on GPT but is not the same as GPT. Think of it as a front-end gateway, but not the only gateway to GPT.
- **Large Language Model (LLM):** GenAI systems (often called models) like GPT, Claude, Bard, Llama, and now hundreds of others that are specifically designed to understand, generate, and interact with human language. These models are "large" both in terms of the size of their neural network architecture (the complex web of

interconnected nodes that process and store information) and the volume of data they have been trained on.

- **Neural Network Architecture:** The structure and organization of an artificial neural network, which consists of interconnected nodes (neurons) arranged in layers. This architecture allows the network to learn and process information by adjusting the strength of the connections between nodes based on the input data and desired output.
- **Prompt:** The initial input or instruction given to the GenAI model to elicit a specific response or output. Prompts can range from simple questions, commands, or statements to more complex scenarios or instructions, depending on the desired outcome. For example, a prompt could be "Write a summary of the key arguments in the *Smith v. Johnson* case."
- **Reinforcement Learning:** A type of machine learning where the AI model learns to make decisions or take actions based on feedback in the form of rewards or penalties. In the context of LLMs, reinforcement learning involves human interaction, such as asking the model questions and providing feedback on its answers to improve its performance.
- **Token:** A unit of data sent to or received from an LLM during the course of performing its services. A token may be a word, part of a word, punctuation, or a mix of the above and is on average approximately four characters in length. A rough guide is that 750 words equate to about 1,000 tokens.
- **Unsupervised Learning:** A type of machine learning where the AI model learns to identify patterns and structures in data without explicit guidance or labeled examples. In the context of LLMs, unsupervised learning involves training the model on vast amounts of text data, allowing it to learn language patterns and relationships on its own.

About the Authors

John Tredennick (JT@Merlin.Tech) is the CEO and founder of Merlin Search Technologies, a software company leveraging generative AI and cloud technologies to make investigation and discovery workflow faster, easier, and less expensive. Prior to founding Merlin, Tredennick had a distinguished career as a trial lawyer and litigation partner at a national law firm.



With his expertise in legal technology, he founded Catalyst in 2000, an international ediscovery technology company that was acquired in 2019 by a large public company. Tredennick regularly speaks and writes on legal technology and AI topics and has authored eight books and dozens of articles. He has also served as Chair of the ABA's Law Practice Management Section.

Dr. William Webber (wwebber@Merlin.Tech) is the Chief Data Scientist of Merlin Search Technologies. He completed his PhD in Measurement in Information Retrieval Evaluation at the University of Melbourne under Professors Alistair Moffat and Justin Zobel, and his post-doctoral research at the E-Discovery Lab of the University of Maryland under Professor Doug Oard.



With over 30 peer-reviewed scientific publications in the areas of information retrieval, statistical evaluation, and machine learning, he is a world expert in AI and statistical measurement for information retrieval and ediscovery. He has almost a decade of industry experience as a consulting data scientist to ediscovery software vendors, service providers, and law firms.

About Merlin Search Technologies

Founded in 2019, Merlin Search Technologies is a leader in AI and secure cloud technology, harnessing the transformative power of Generative AI to reengineer search, investigations, and discovery workflows for the digital age. As industry trailblazers, Merlin offers the first-ever On/Off Cloud Utility Pricing, allowing clients to reduce hosting costs, support green computing, and make discovery more secure, flexible, and cost-effective.

The Merlin team brings two decades of experience designing and building advanced discovery platforms for the largest companies and law firms in the world. This deep experience with traditional discovery workflows has positioned us to create a groundbreaking solution that sets a new standard for fast, efficient, and cost-effective discovery. Driven by an unflinching commitment to innovation, and backed by the newest AI and cloud technologies, Merlin's team is poised to redefine the future of search, investigations, and discovery.

Merlin's mission is to redefine how legal professionals handle their most complex and data-intensive challenges. By combining the power of Generative AI with secure, flexible, and cost-effective cloud technologies, we're enabling our clients to find critical information faster, more efficiently, and more affordable than ever before.

Say hello at <https://merlin.tech>.