COMPLETE DISCOVERY SOURCE

GENERATIVE AI IN THE LEGAL WORLD:

Integrating LLMs Into Your eDiscovery Workflow



INDEX

03 Executive Summary

03 Introduction

04 Recapping TAR

05 What is Generative AI?

06 Generative AI for Document Review

07 Incorporating LLMs into Traditional TAR Workflows

09 Using aiR Classifications to Drive Other Workflows

11 How do the LLMs see their future in eDiscovery?

12 About CDS

Executive Summary

Generative AI, particularly tools like GPT, is transforming legal workflows by offering capabilities far beyond traditional chatbots. These AI models can intake extensive text and provide detailed, relevant content, aiding tasks such as document summarization and complex data analysis.

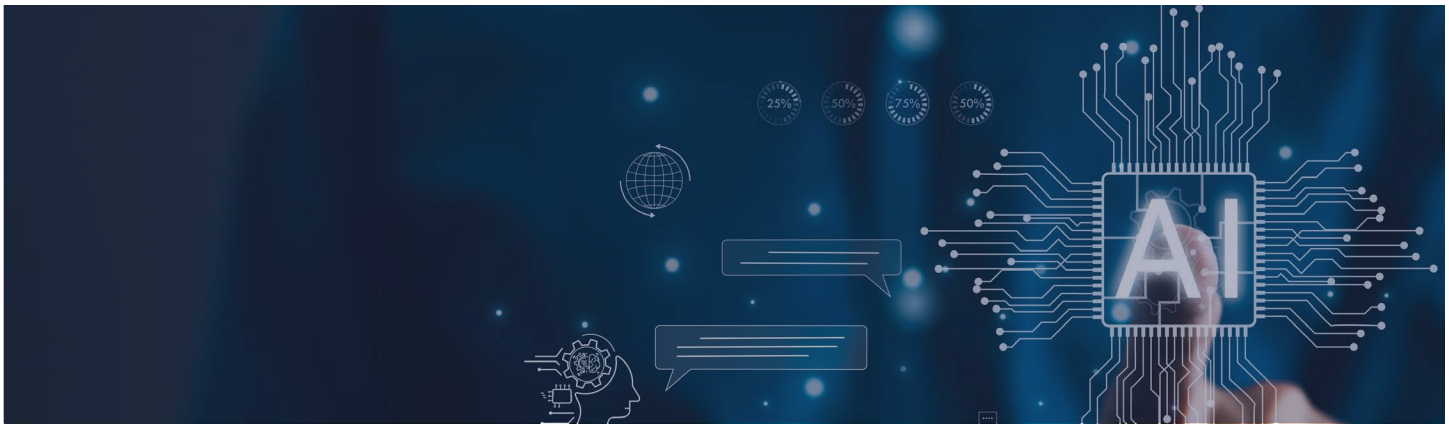
However, as impressive as these tools seem, they come with limitations, including data currency restrictions, potential for generating incorrect information ("hallucination"), and issues of data bias and security.

Generative AI has also expanded the already bountiful array of capabilities in the key cost saving - and Proportionality enhancing - legal use cases of search and review options. While the new choices might make selecting the best technology approach a little more challenging to identify, and the risks associated with using the technology need new mitigation strategies, the benefits of this software will likely be a major win for the eDiscovery industry going forward, especially in regard to time savings. This paper will discuss how law firms can effectively integrate Generative AI with existing eDiscovery platforms and workflows, including technical and strategic considerations that must be made.

Introduction

Anyone involved in eDiscovery has probably overheard battle-hardened veterans reminiscing about the old days, when the litigation support world was tangible– attorney teams sifting through banker's boxes to (hopefully) find the relevant documents, paralegals photocopying and manually Bates-stamping every page. However, with the advent of the digital age, electronically stored information (ESI) replaced paper documents and "e-" became forever attached to discovery.

Technology Assisted Review (TAR) software introduced attorneys to the potential power of machine learning, but its adoption has been slow, and it has remained much the same over the last decade. Until now. After 2024, everything is poised to change again, but this time those changes look more impactful, and the potential represented by the new technology is even more disruptive. Why? Generative AI. LLMs. ChatGPT. A new wave of generative AI solutions—including Relativity aiR and DISCO's Cecilia—are coming to market. The hype surrounding these products has been immense, prompting teams to ask: Will generative AI replace TAR? Will it replace "first level" human review entirely?



What is Generative AI? Assisted Review (TAR)

TAR has existed in virtually the same two formats since it emerged as a defensible document review process for large data sets in 2012 after Judge Andrew Peck's seminal opinion in Da Silva Moore. Judge Peck and numerous outspoken judges made it clear that the bench recognized that discovery had to change and that attorneys had to look to technology to better address the insurmountable challenges of electronic evidence (Electronically Stored Information or "ESI") in the context of discovery obligations under federal and state court rules. The US legal system has always taken great pride in the adversarial dispute resolution system and the right of litigants to have their "day in court" with relevant probative evidence to support or refute the claims and defenses of their cases and the cases brought against them. Computers both capture information that might be important to a matter and create new challenges of gargantuan data volumes and complexities in collecting and reviewing that data without diminishing its usability and authenticity.

TAR 1.0:

In this workflow, attorneys review a random sample of documents known as a "control set." Attorneys then review a few samples of documents, typically about 1,000 to 5,000 in total, used to train the TAR software to distinguish relevant from irrelevant documents. The software generates scores for the unreviewed documents, and using the control set for measurement, chooses a cut-off score. Documents above the cut-off score are produced or reviewed, while documents below the cut-off score are discarded.

TAR 2.0:

In this workflow, the software works the same way, but the review workflow is different. Rather than building a model and then proceeding with review or production, the TAR scores are used to sort the documents. Documents with the highest scores are reviewed first, and the software re-trains itself periodically with newly reviewed documents. At some point, relevant documents start to run out, and (hopefully) the review can end before the full set has been reviewed manually. A decision to stop is typically made by reviewing an illusion sample – a random sample of the remaining documents.

What is Generative AI?

Generative Artificial Intelligence (AI) is a type of technology that uses machine learning to generate novel content in response to user prompts. Generative AI actually dates back to the 1960s when the first chatbot, Eliza, was developed at MIT. While the concept has been around for decades, generative AI hasn't made much news outside the computer science community until recently. The lack of prior fanfare was appropriate – people could draw images and perform music that were objectively better and cheaper than the software, so generative AI wasn't creating value.



Most generative software used today works interactively, meaning the user types or says an instruction using natural language, and the software outputs an attempt to fulfill the request – human language text, programming code, image, audio, video, or 3D models. In the last few years, generative AI content quality has become extremely realistic. It is now capable of creating images that look like photographs, songs that sound like Drake, and countless other deepfakes that continue making headlines. This ability to interact with and mimic humans is exactly why GAI has grabbed so much attention and created such a sense of excitement and foreboding. Legal professionals and clients are both keen to understand GAI's use cases and insistent that the risks endemic to GAI are understood and managed.

Seeing the potential value, venture capital investment in generative AI was estimated at $4.8 billion in 2022, nearly 12x growth since 2018. Although the legal profession isn't in the business of generating deepfakes, law firms and legal departments generate enormous amounts of text, and will be directly impacted by the most well-known Generative AI software – ChatGPT.

Generative AI for Document Review

The ease with which legal professionals can interact with GAI carries over to legal document review. In fact, experienced document review attorneys and managers will quickly recognize the parallels in setting up a human "1L" document review and a GAI-powered review matter.

Here's a breakdown of what Generative AI does:

- A human user enters a prompt to Open AI's Chat GPT, Microsoft Co-Pilot, Google's Gemini, or some other interface.
- The chat interface submits the prompt to a **Large Language Model (LLM)** which generates a response. The responses have such high quality that it seems like the model understood the prompt, like a human.
- However, the model processes information much more quickly than a human ever could, and sometimes the responses include errors – termed "hallucinations" – that a human would not make.

So how can this technology help with document review?

Many eDiscovery platforms have developed integrations with LLMs using a technique known as Retrieval Augmented Generation (RAG). RAG combines a generative model prompt with access to a knowledge base of documents. This combination allows the model to ground its responses in that knowledge base, rather than its original training data.

In this example, AI extracts the information from the source material:



Look at the following link and provide a list of the names and titles of the executives at Complete Discovery Source: https://cdslegal.com/team/



The document lists the following executives at Complete Discovery Source:

- **Ron Ramnarayan, Founder and CEO**
- **Vinnic Budhram, Chief Operating Officer**
- **Nyi Htwe, Chief Techonology Officer**
- **Pete Smith, SVP, Sales**
- **William Belt, Managing Director, Consulting**
- **Regina Chepalis, Vice President, Sales**
- **Michael Milicevic, Esq., VP eDiscovery Products & Solutions**
- **Richard Das, Vice President of Operations**

Incorporating LLMs into Traditional TAR Workflows

LLMs have already been trained using massive volumes of information scraped from the internet, giving them the ability to understand natural language. They can also perform data classification, e.g., labeling the topics of news articles and sentiment analysis. With traditional TAR workflows, one of the barriers to adoption was the clunkiness of deploying TAR to the deadline-driven, time-sensitive document review project. The setup requirements involved meant that using TAR often meant the delay in getting a review team started. This is the heavy lifting that allows GAI to "sound human" and "look like they understand" input from humans. That intuitive accessibility reduces a key barrier to adoption.

And this "heavy lifting" can be incredibly powerful in the context of legal document review. Although legal document review may seem like a distinct type of classification, the only real difference is the subject matter expertise needed to make the distinction between a relevant and non-relevant, responsive and non-responsive document. We have learned that TAR can be an effective way to categorize data. Therefore, if the LLM can be trained on what a relevant and responsive document is, they could be useful for TAR workflows. Here's two ways to replicate or augment the TAR workflow using the new capabilities presented by LLM's:

Option 1: Fine Tuning

One option is to leave the process alone and use LLMs in a process called "fine tuning." In this workflow, new potentially relevant documents are ingested into the model. A subset of these documents could be reviewed by attorneys and labeled as relevant and not relevant, and a traditional classification algorithm added to create the same type of ranks that have been used in traditional TAR 1.0 and TAR 2.0 workflows.

While this approach has the advantage of utilizing a workflow that attorneys know will be defensible and accepted by regulators such as the DOJ and FTC in merger reviews, it has some limitations—two academic studies performed using pre-trained transformers for TAR workflows failed to find substantial improvement over existing classifiers.

However, the technology is rapidly evolving. LAER AI, a newcomer in the TAR market, conducted its own study and found that their models significantly outperformed traditional tools using a traditional TAR workflow.

Option 2: Train the LLM

A second possible workflow would involve prompting the LLM the same way attorneys hired to perform document review are trained by telling the LLM what is relevant for the review, potentially using a review protocol or natural language search. The LLM could then analyze each document and return with a score or a decision regarding whether or not it is relevant.

It is possible to prompt the LLM to classify a document as relevant or irrelevant, responsive or non-responsive to a set of instructions and requests:



In this workflow, it would still be possible to provide the software with examples of relevant documents, but it likely wouldn't be necessary or beneficial to provide more than a few for each issue. Validation could be the same as traditional TAR – the subject matter expert attorneys would review a random sample used to evaluate recall and precision.

This workflow has the advantage of requiring much less time reviewing documents to train the software by subject matter experts although it might be even more computationally expensive than fine tuning a model. In addition, it lacks a decade of approvals in court opinions.

Relativity aiR for Review features a much more robust and useful set of outputs than the one-word reply above, making it is possible to:

- Input a set of instructions detailing the type of content that is relevant to the review. The instructions can be in plain English, in a format that is nearly identical to a protocol written for a review team.

- Use aiR (via GPT-4o) to classify all the documents in the database on a scale of 0-4, with higher ranks likely being more relevant.

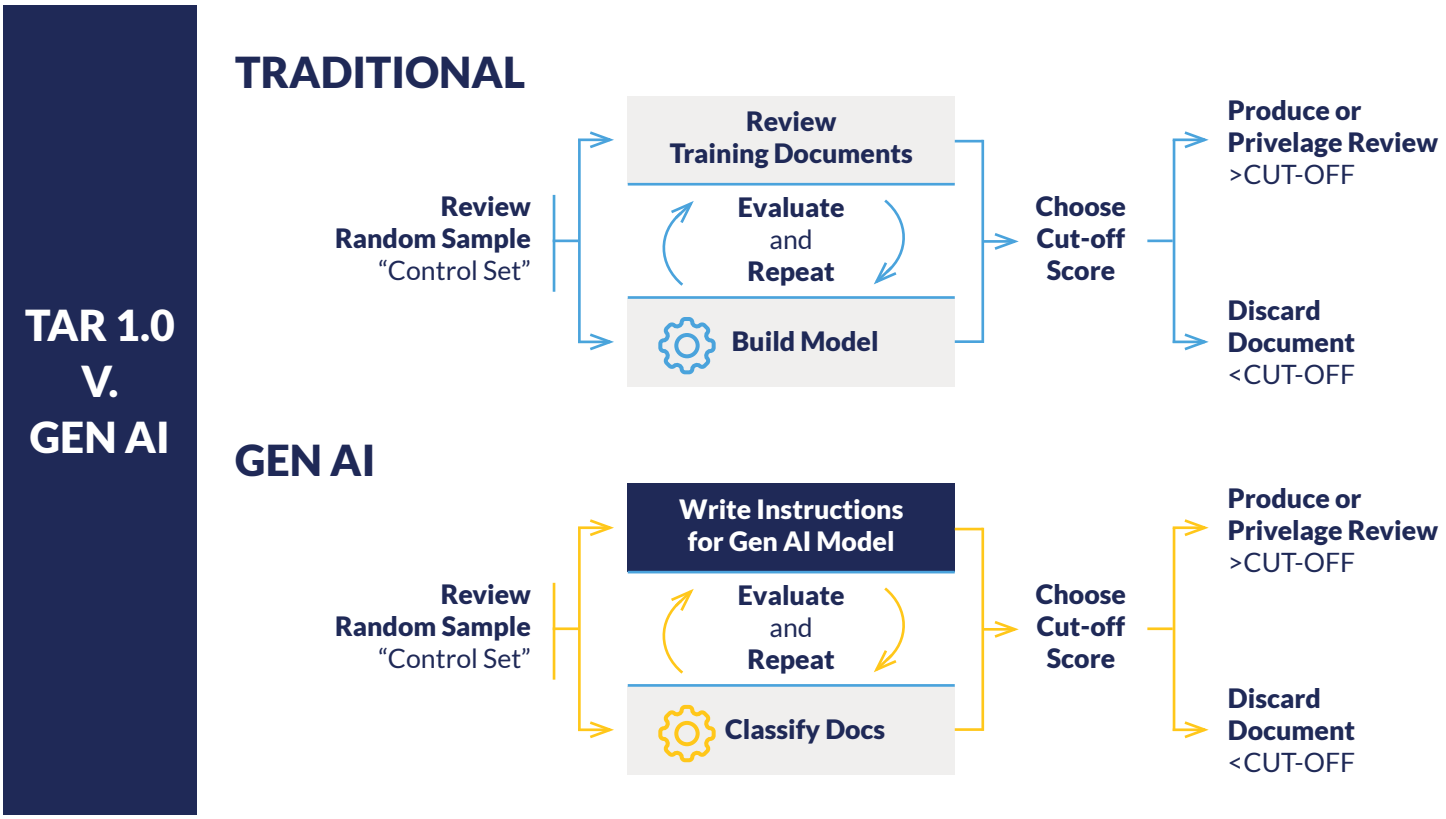
The output of the process is also more akin to the typical review workproduct we see in an eDiscovery first pass review, one that can flow into the QC reviews we typically deploy.

Using Relativity aIR Classifications to Drive Other Workflows

As with traditional TAR software, it is possible to use the classifications to drive different workflows:

TAR 1.0 Replacement

Using aiR as a replacement for relevance review in the same way TAR 1.0 has been applied most of the last decade is the easiest use case for this workflow—The Sedona Conference provides a deeper dive into this framework. In testing, aiR for Review has regularly achieved 85-95% recall, significantly better than the 80-85% typically realized with traditional TAR software. Although aiR for Review is much faster and less costly than manual review, it is more expensive than traditional TAR 1.0 software. Other potential downsides: it might only be helpful for outgoing production reviews and there is not yet any Da Silva Moore-type court approval of this workflow.

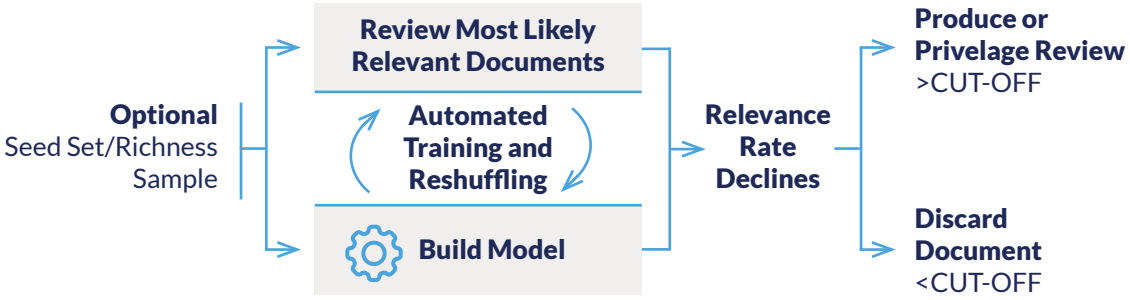


TAR 2.0 Replacement / Augmentation

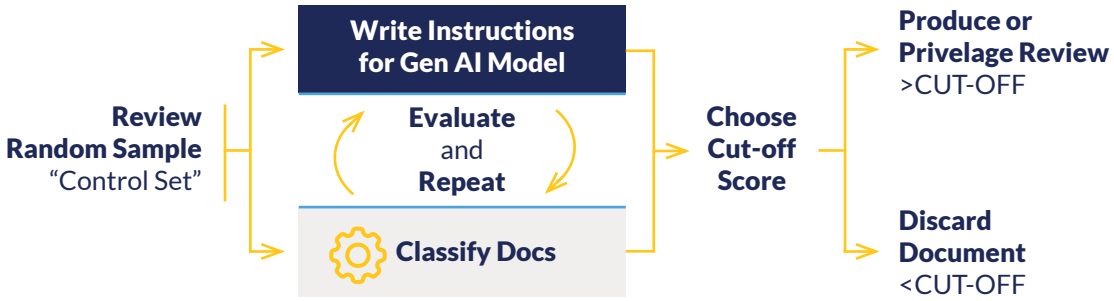
Although aiR for Review generates only four rankings, the rankings can be useful for prioritizing a review in the same way as traditional TAR software. It's also possible to turn on traditional TAR 2.0 software—such as Review Center—within each rank to more efficiently prioritize each subset. However, because this workflow encompasses review AND the heightened software fees, it will likely be most appealing for smaller data sets with more expensive attorney reviewers.

TAR 2.0 V. GEN AI

TRADITIONAL



GEN AI



One of the biggest benefits of aiR for Review over traditional TAR software is that it is not limited to a binary relevant/not relevant designation. Instead, it can classify for 10 different issues in a single analysis, which could be used after review to help with fact discovery and deposition preparation.

How do the LLMs See Their Future in eDiscovery?

How will LLMs be incorporated into eDiscovery TAR workflows? That remains to be seen. We asked three of the top generative AI chatbots to list the top five ways generative AI and LLMs are most likely to impact eDiscovery. Here's what they said:

Are you ready to find out more about how machine learning will impact your eDiscovery processes? CDS provides a full range of advisory services related to AI and eDiscovery workflows. To discuss how we can help your organization understand the benefits and challenges related to machine learning, contact us for a consultation today.

ChatGPT

- Automative document review
- Predictive coding
- Advanced data classification
- Enhanced search capabilities
- Cross-language document review

You.com

- Improved document review and analysis
- Automated summarization
- Better legal understanding
- Language translation
- Risk management

Gemini

- Improved search and filtering
- Automated document review
- Reduced costs
- Increased accuracy
- Improved transparency

About CDS

Complete Discovery Source (CDS) is a premier full-service eDiscovery provider with over two decades of experience helping corporations, law firms, and government agencies navigate complex legal and compliance challenges. Powered by real people, CDS combines industry-leading technology with expert consulting and project management to deliver efficient, cost-effective solutions tailored to litigation and investigations. With expertise in data migration, forensics, and managed review, CDS ensures seamless support at every stage of the eDiscovery process. Proprietary tools like Convert and Vision, along with premium Advisory and FedRAMP-Authorized Cloud Services, empower CDS to deliver real results that matter, no matter the size or scope of your case.

Headquartered in NYC and certified as a Minority Business Enterprise (MBE), CDS delivers scalable, tailored solutions to meet the demands of modern legal challenges. With offices and data centers across the globe, CDS is uniquely positioned to support clients worldwide. Learn more at cdslegal.com or email getstarted@cdslegal.com.

Learn more:
Visit cdslegal.com

Speak with an eDiscovery expert:
Call 855-813-0700

Request a Demo:
Email getstarted@cdslegal.com

