



PERSPECTIVES

**Generative AI in
eDiscovery: Now,
Soon, or Never**

Our perspectives feature the viewpoints of our subject matter experts on current topics and emerging trends.

INTRODUCTION

Having immersed themselves in an array of deep dive demonstrations of new Generative AI (GenAI) products for eDiscovery displayed during Legal Technology events in London and New York, Chantelle Jalland and Marybeth Kings reflect on the latest technology developments in eDiscovery. What is clear is that all leading document review platform providers have developed, or are in the process of developing, GenAI extensions to enhance their core product.

Each software developer has approached the opportunity of incorporating GenAI into their existing products in a slightly nuanced way.

MACHINE LEARNING

Before exploring new GenAI features, it is important to take a step back and look at the journey of AI adoption in the form of machine learning in eDiscovery.

Predictive coding (also known as technology-assisted review or TAR) is a technology that is able to utilise AI to identify and categorise documents as potentially relevant, learning from human decisions on other documents within a review pool. Originally advised for large cases containing over 50,000 documents, this advice evolved as the technology improved to automate parts of the workflow so that there can now be benefits on matters with as few as 1,000 documents.

It has been 15 years since predictive coding was released¹ as part of a document review platform for use in eDiscovery. It took three years until the use of predictive coding was accepted by Judge Andrew J. Peck in federal court in New York² and a further four years for the formal acceptance by the High Court in the UK³. That is not to say that predictive coding wasn't used before formal decisions were issued, as it was still possible to adopt the technology if privately agreed upon by the parties. That said, it is clear that adoption was slow, and lawyers were (and still are) cautious to utilise the technology.

The Civil Procedure Rules in the UK⁴ has the overriding objective of dealing with cases justly and at a proportionate cost. We understand that not every matter is suitable for predictive coding workflows, however, there is an argument that it should at the very least be considered for use in all cases.

Unfortunately, the reality is that predictive coding is not being used as much as it should, due to a lack of trust, education, and understanding in the technology. One study suggests that only 30% of people are using predictive coding in all or most of their cases⁵. All document review tools that incorporate predictive coding also include statistical validation techniques, which should in theory mitigate any trust issues with the technology.

GENAI

The term 'AI' became mainstream when 100 million users were introduced to OpenAI's chatbot, ChatGPT within two months of it becoming public⁶. Since this launch 18 months ago, there is a greater awareness of the availability of sophisticated AI technology across all industries. It is no surprise to the eDiscovery industry that existing technologies would expand to utilise GenAI in streamlining document review.

Here, it is important to pause and define what a Large Language Model (LLM) is. According to Merriam Webster, an LLM is a language model that utilizes deep methods on an extremely large data set as a basis for predicting and constructing natural-sounding text⁷. In other words, you can ask an LLM a question as you would a human, and it will return a natural human-sounding response. In the background, the LLM stores vast amounts of information with various static connections with probabilities between words. LLMs themselves don't learn when questions provide new information.

In eDiscovery, it is not as simple as saying 'let's use GenAI.' We need to understand how we can use GenAI as part of an end-to-end workflow within different use cases. Each document review tool has adopted GenAI in a different

¹ Accelerate Document Review Platform developed by Recomind was released in 2009 and is now owned by OpenText <https://www.opentext.com/products/accelerate>

² *Da Silva Moore v. Publicis Groupe & MSL Group, No. 11 Civ. 1279 (ALC) (AJP) (S.D.N.Y. Feb. 24, 2012)*

³ *Pyrrho Investments Ltd v MWB Property Ltd [2016] EWHC 256 (Ch)*

⁴ *Civil Procedure Rules, Rule 1.2.*

⁵ 2024 State of the Industry Report, 9 January 2024 <https://ediscoverytoday.com/2024/01/09/2024-state-of-the-industry-report-is-out-heres-how-to-get-it-ediscovery-trends/>

⁶ ChatGPT sets record for fastest-growing user base, 2 February 2023 <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

⁷ Merriam-Webster Online Dictionary <https://www.merriam-webster.com/dictionary/large%20language%20model>

way. Some technology companies openly advise that they are connecting to ChatGPT-4, some others withhold which LLM they connect to, and some connect to multiple different LLMs or create their own Small Language Models (SLMs). These SLMs are slimmed-down language models that are generally easier to train, fine-tune and deploy, as well as being cheaper to run⁸. There are some examples of SLMs that are built for very specific tasks, (for example, summarisation), where it is not possible to alter the prompt, and therefore leads to the question of validation. With so many options out there, each geared towards tackling a specific problem, we must first look at the unique challenges our data sets present to know which AI will best provide a solution.

Let's explore five GenAI applications below:

1. Review

The idea of this method is to replace first level linear review. This bottom-up approach requires a very detailed knowledge and up-front understanding of the case in order to create a detailed prompt before sending the documents through the LLM context window, which then processes each document one at a time. Initial case studies that have adopted this approach have seen a high level of accuracy using recall validation statistics. Therefore, the advantages include an increased level of consistency of coding as well as a reduction in the time to complete this stage of review.

2. Summarising

It goes without saying that law involves a lot of reading. Whether it is reviewing a contract with hundreds of pages or a 60,000-word judgement, this exercise takes a lot of time. If we can utilise technology to bypass this reading and provide a summary, would we take it? This technology is now available in the form of AI-powered summaries to deliver key insights into our evidence. In certain cases, this could be incredibly valuable, however there may be seemingly insignificant events excluded from the summary that could, in fact, turn a case on its head.

3. Timelines

In most legal cases, chronologies of events are helpful to better understand the timeline of key facts in a matter and build case strategies for clients. GenAI

may be notoriously bad at creating images with numbers; however, this is not the case when it comes to LLMs producing written accounts in order of dates. The timeline can also include references to sections in disclosed documents.

4. Privilege

One area that has long been a limitation of technology assisted review is the ability to recognise privileged material, because privilege decisions are often not found within the four corners of a document and may require context from other documents, such as cover emails and attachments. GenAI on the other hand, has the ability to analyse more than just the content contained within a single document, and therefore brings additional efficiencies for the creation of privilege logs.

5. Chatbot

New AI personal assistants built into document review platforms in the form of a chatbot (often named as a popular boys or girls name to make it feel human) add an additional avenue for interrogating your data. This top-down approach could be suitable in investigations that involve querying the database with general scoping questions. These chatbots have been shown to give robust responses based on the actual content of the database, and even include anecdotes and references that may be particularly interesting to the user.

REFLECTION

While these are simply five examples and by no means an exhaustive list, the variety of forms and use cases demonstrated by the above makes one wonder, what else? Could it be used by the receiving party to query what is new in the received production versus their own? Will it eventually make judgements on the likelihood of winning a case? Will there soon be a wider repository of case studies from which it can draw? For example, there could soon be a custom LLM that has the ability to reference actual cases and their judgements from within a review platform (one that isn't as subject to hallucinations and making up case studies⁹).

⁸ <https://thenewstack.io/the-rise-of-small-language-models/>

⁹ <https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=694204a77c7f>

It appears the real gap in the commentary is the lack of concrete evidence on comparing GenAI solutions to existing TAR solutions. Yes, there have been a few eDiscovery experiments that analyse the different approaches with reference to a case study. Generally, with this GenAI model having a recall of approximately 95% over TAR with approximately 85%. But arguably, the statistical advantage of GenAI needs to be evaluated in conjunction with costs to understand the true benefit.

The pricing point is complex, as most GenAI relies on third-party LLM token costs. Without going into too much technical detail, each LLM has a context window, essentially an input limit, from which the LLM generates a response. The amount of text that is input into an LLM is broken down into segments called tokens, which is used to calculate costs. Currently, doubling the size of a context window quadruples the cost, which is a particular pain point for the eDiscovery sector as we would hope to use entire databases of documents as context.

Technology providers are comparing the cost of adopting this approach against human review, though perhaps we should be comparing it against the cost of utilising existing technologies. Some are packaging the cost in a higher overall per GB rate, while others cost per prompt run or by mimicking the token cost. What we can say for certain, is that it would be cost-prohibitive to run 5 million documents through a GenAI model that links to ChatGPT-4 with its current pricing. Therefore, a workflow still needs to involve existing eDiscovery filtering techniques. Others are building their own SLMs that can lead to more economical cost models, but potentially at a cost to quality due to the smaller data pool that it draws from.

Whether exploring technology that connects to LLMs or SLMs, it is imperative to feel confident that sensitive client data is protected at all times, by asking your provider detailed privacy and security questions.

CONCLUSION

It is evident that technology providers are developing purpose-built technology solutions, and not simply utilising GenAI for the sake of using it.

The evolution of eDiscovery expertise seems to be shifting from analysis of appropriate keywords to prompt engineering. Or will prompt engineering soon be a term of the past as custom models are developed with pre-generated prompts?

The question is not how we can replace humans with GenAI, but rather how can we adopt LLMs alongside existing technologies in the most efficient way. Unsurprisingly, it comes down to what works best for your dataset and workflow.

So, what will we see in the next 12 months? Will GenAI replace first pass review? We certainly need more evidence of the effectiveness across a variety of matters. Our prediction is that it will not be a one-for-one overnight replacement – and not because of the technology, but because of current pricing. It will take time to understand the true benefit and ensure we are able to trust existing validation techniques. However, will it take 15 years to get to 30% adoption, as was the case with TAR? We don't think so.

Acknowledgments

We would like to thank Chantelle Jalland and [Marybeth Kings](#) for providing insight and expertise that greatly assisted this research.

More About J.S. Held's Contributor

[Marybeth Kings](#) is a Senior Director in the EMEA [Digital Investigations & Discovery](#) team at J.S. Held. Based in London, she has more than a decade of experience assisting government, legal, and corporate clients with their eDiscovery matters, from internal and regulatory investigations, arbitrations, and data subject access requests to pro-bono, charitable, and non-standard applications of eDiscovery software. Examples of projects she has worked on include data subject access requests, litigation, arbitrations, internal investigations, regulatory investigations, voluntary disclosure, as well as claims tracking and management.

Marybeth can be reached at marybeth.kings@jsheld.com or +1 44 20 4534 0415.

This publication is for educational and general information purposes only. It may contain errors and is provided as is. It is not intended as specific advice, legal or otherwise. Opinions and views are not necessarily those of J.S. Held or its affiliates and it should not be presumed that J.S. Held subscribes to any particular method, interpretation or analysis merely because it appears in this publication. We disclaim any representation and/or warranty regarding the accuracy, timeliness, quality, or applicability of any of the contents. You should not act, or fail to act, in reliance on this publication and we disclaim all liability in respect to such actions or failure to act. We assume no responsibility for information contained in this publication and disclaim all liability and damages in respect to such information. This publication is not a substitute for competent legal advice. The content herein may be updated or otherwise modified without notice.