



HOW SHOULD ESI PROTOCOLS ADDRESS GENERATIVE AI FOR DOCUMENT REVIEW

BEN SEXTON, SENIOR VICE PRESIDENT - INNOVATION & STRATEGY, JND eDISCOVERY

“The ability to simplify means to eliminate the unnecessary so that the necessary may speak.”

- Hans Hofmann

How should ESI Protocols evolve to address Generative AI (GenAI)? Are we entering a brave new world where lawyers will negotiate complex LLM (Large Language Models) ‘hyperparameters’ like “Top P” in order to ensure they’re using the technology properly? Or is there a simpler (better) answer?

Spoiler alert, it’s the latter. While GenAI is a new ‘black box’ for review automation, there’s nothing new or more complicated about how we’ll validate the results.

Background

Like the fossil record, ESI (Electronically Stored Information) Protocols are an evolutionary marker on a case. Email threading used to be cutting-edge. TAR (Technology Assisted Review) used to be way too complicated. Attachments used to actually be attached to emails. ESI Protocols have evolved to keep up with the rapidly changing landscape of discoverable information, as well as the advancing technologies at our disposal to address that information. Most well-crafted ESI Protocols today address topics like Email Threading, Short Messages, Ephemeral Data, Modern Attachments, and TAR validation. Soon, updates will be made to address GenAI.

By Way of Illustration

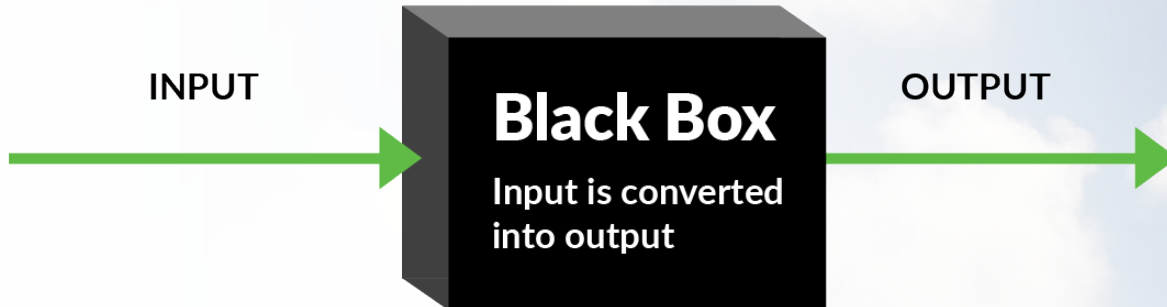
Let’s consider a real-world scenario. You work at a law firm representing a company in a class action lawsuit. Your eDiscovery platform offers GenAI for first-pass review. You perform targeted collections and, after objective culling, wind up with 100k documents subject to review. You craft a prompt based on the review protocol and then cut the GenAI loose, reviewing the 100k documents. 12 hours later, the GenAI has completed first pass review, and 10k docs are coded as responsive. You batch out the responsive set and assign it to an eyes-on team for second pass review.

You’re clearly a busy person who deserves a raise, but let’s pause here and zoom out. GenAI’s work is done. What just happened? You took a lot of something, applied an operation, and found less of something to send downstream. Again, specific operations aside, does that sound like a brand-new process, or did we just invent a better funnel?

Similarly, plaintiffs may use GenAI to automate their review of the documents that were produced. While their legal objectives are different, the operation they’re performing, of separating the ‘wheat from the chaff’, is essentially the same.

Just A New Black Box?

Ultimately, GenAI is just a new black box that addresses the same age-old problem: separating the wheat from the chaff. GenAI may be show-stoppingly better at the job, but it's the same job formerly held by search terms and TAR.



Measuring Success

This all goes to say, the question of “How do we make it defensible?” can be answered in the same way as it is for TAR, keywords, or any other operation that results in a “good pile” and “bad pile”. Do we negotiate the index dimensionality of a TAR model? No, we simply validate the results using random sampling to measure recall, precision, and elusion rate (others if desired). These terms will become table stakes in the coming months, so let’s quickly review what they mean.

- **Recall asks the question, “Did you get everything?”**
More precisely, recall is the % of responsive/relevant documents that you were able to find. This is the most important validation marker for all parties.
- **Precision asks, “How much extra stuff did you have to pull in in order to get everything?”**
This is more relevant for the producing party, since it measures how much time they’ll spend reviewing non-responsive docs in order to arrive at a production set consisting of only responsive docs.
- **Elusion Rate asks, “What % of the ‘discard pile’ is actually responsive?”**
This estimates the % of responsive documents missed by whatever “culling” operation was deployed.

As we’re all aware, perfection is not the standard for defensibility. Instead, thresholds are set for recall, precision, and elusion rate, to establish criteria for success (e.g. 85% recall).

1. Role of Recall

Generally, recall metrics receive more attention from parties than precision metrics. In *Lawson v. Spirit AeroSystems*, the defendant used TAR to produce with a recall of approximately 85 percent, which the court confirmed was reasonable and within a typical range for TAR matters.²⁰² The plaintiff had demanded that the defendant switch from keyword searching to a TAR methodology, which the defendant ultimately agreed to do, subject to filing a motion to shift costs based on proportionality.²⁰³ An initial review using TAR 2.0 achieved a 68.5 percent recall rate of responsive documents, but the plaintiff insisted that the process be repeated until a 75 to 85 percent recall rate was achieved. The defendant agreed to 80 percent recall and then, after stopping its review, determined that it had reached 85 percent recall.²⁰⁴

Sedona Conference TAR Case Law Primer, Second Edition, 2023, p. 34

So, How Should We Address GenAI in ESI Protocols?

The same way you currently address TAR. Negotiate the minimum 'accuracy' thresholds that must be met to use the results. For example:

For TAR, we may say that: "Parties agree that, to the extent that TAR is deployed to accelerate the review, a minimum Recall of 85% must be achieved in order to utilize the results."

For GenAI, we can modify it to: "Parties agree that, to the extent that TAR, AI or other 'review automations' are deployed to accelerate the review, a minimum Recall of 85% must be achieved in order to utilize the results."

It's a simple yet robust solution. Occam would be proud.

Everyone Wins?

It's often cited that keywords are able to recall 30-50% of responsive documents. TAR improves averages to around 80-85%. However, 85% recall still means 15 out of every 100 responsive documents are neither reviewed nor produced. In the coming months, we'll see more hard figures coming out around GenAI's "accuracy," but between several of our own tests and larger-scale tests already published by law firms, recall rates for GenAI seem to hover in the 90-96% range. That means that, on the receiving side, you're likely receiving ~10% more responsive documents, which is a good thing.

In other words, a better funnel is better for everyone.

In Conclusion

While the technology is game changing, the way we validate the results is business as usual.

Ben Sexton

Senior Vice President - Innovation & Strategy, JND eDiscovery



About JND eDiscovery

JND eDiscovery is an innovative legal technology and eDiscovery services provider supporting law firm, government and corporate clients in the areas of: Generative AI, Litigation Readiness, Identification, Forensic Collection and Preservation, Early Case Assessment and Data Processing, Hosted Review, Managed Review, Analytics and Technology-Assisted Review (TAR), and Production.

JND is a subsidiary of JND Legal Administration (JNDLA), founded by Jennifer Keough (CEO), Neil Zola (Executive Managing Director), and David Isaac (Executive Managing Director). The company's eDiscovery service line was formed in 2016 with the acquisition of Minneapolis-based Alloy Group (now JND eDiscovery). JND eDiscovery's leadership and key personnel have been working side by side for more than 10 years. JND continues to operate out of Minneapolis and is led by EVP of eDiscovery, Scott Lombard and Ben Sexton, SVP of Innovation and Strategy.

CONTACT: [JNDLA.com](https://www.jndla.com) 800.207.7160 info@jndla.com

CONNECT: [in](https://www.linkedin.com/company/jnd-legal-administration) /jnd-legal-administration

JND