JND

GENERATIVE AI AND PROTECTIVE ORDERS

JONATHAN MOODY, VICE PRESIDENT - SALES, JND eDISCOVERY

A recent conversation I had raised a new concern surrounding the use of Generative AI that is worth talking through. Will using Generative AI tools violate obligations surrounding the storage and review of documents designated as "Confidential" under protective orders in litigation.

Before diving into the discussion, it makes sense to lay out what a protective order is intended to do.

A simplified explanation is that a protective order lays out what protections exist for the data that is discoverable during a particular litigation, and what is required to maintain those protections. For example, a protective order might identify certain data to be highly confidential, which is designated to be documents that can only be viewed by attorneys actively working on the litigation in question, or their designated representatives.

So, what concerns might exist for leveraging Generative AI when a protective order is in place?

The short answer is disclosure of information in a way that would violate or revoke the designations given to a particular document.

The longer answer digs a little deeper into how the Generative AI tools work. A clearly improper disclosure of a highly confidential document would be leaking that document to the media for whatever purpose. But how could using a Generative AI tool be deemed analogous to a media leak?

Take, for example, an attorney for the makers of Dr. Thunder who decides to leverage ChatGPT to summarize a document produced to them by the makers of Dr. Pepper, which was deemed highly confidential by a protective order in litigation involving the makers of Dr. Pepper and Dr. Thunder. The attorney copies and pastes the text of the document into ChatGPT, preceded by the text "summarize the following document in 50 words or less." ChatGPT performs this task well, resulting in a highly complex scientific document being more easily understood by a litigator without a scientific background.

Great result, right?

But keep in mind that ChatGPT learns from inputs fed into the back-end LLM, which would potentially include information included in prompts.

So, the next day, someone queries ChatGPT with "what is the secret recipe for Dr. Pepper" – which returns the list of the 23 flavors in Dr. Pepper, along with the precise quantities and mixing instructions for each.

Prior to that attorney's actions, no such information was publicly disclosed. The highly confidential document was improperly disclosed in a way that definitely violates the protective order.

Now the real question becomes "how can attorneys leverage Generative AI without violating protective orders?"

The answer is straightforward. OpenAI, which owns most LLMs in the space, provides "private" versions of its LLMs (such as Microsoft Azure versions of GPT4), where the model does not actively learn from the inputs it leverages. With safeguards in place to securely "forget" any and all inputs, applications leveraging this type of model will generally jump through the hoops of a protective order without issues.

One other question that comes to mind is whether we've seen this in other areas of the eDiscovery world.

The quick and easy prior application of this would be with regards to translations. Copying and pasting entire documents, or portions of documents, covered by a protective order into a free translation tool is a pretty clear violation of protective orders. Legitimate eDiscovery platforms leverage machine translation applications that are private and paid, such as Azure Cognitive Services, in order to securely translate documents.

The major takeaway is to be careful which Generative AI applications you're using to ensure that the LLMs in the background will not leverage the data inputs to further train the LLM. However, as long as you perform standard due diligence to ensure your AI software is setup to maintain confidentiality, which OpenAI is fully capable of doing, there should be little reason for concern regarding the risks of violating a protective order due to the use of Generative AI tools.

Jonathan Moody

Vice President - Sales, 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.