

Law Firm AI Dreams Rest On Clean Data Practices

By **Steven Lerner**

Law360 (August 2, 2023, 3:57 PM EDT) -- Law firms cannot fully realize the benefits of artificial intelligence without first ensuring that they have clean and accessible data, according to leaders at BigLaw firms.

Data cleaning is considered to be the crucial step that law firms must take before implementing AI tools, especially as more firms start using these tools and a few lawyers publicly stumble with AI due, in part, to bad data.

Training data is the foundation of AI. Models, which are programs or algorithms, use it to learn, make predictions and create content. By training AI models on clean data, users can expect outcomes that are more accurate. But if the data sets used to train AI are not clean, then the output might be riddled with mistakes.

In other words: garbage in, garbage out.

"Everything begins and ends with data quality," Thomas I. Barnett, a principal and chief data officer for Jackson Lewis PC, told Law360 Pulse. "If you don't have data quality, you can't really rely on the result."

That reliability in data gives Jackson Lewis more confidence in the results that it is getting from its AI projects, according to Barnett. And Jackson Lewis is not the only BigLaw firm to feel this way.

Data is also considered a valuable resource for Reed Smith LLP. The firm is test-driving the new generative AI platform from LexisNexis Legal & Professional, and is running pilot programs with commercial generative AI, specifically the generative pre-trained transformer or GPT systems developed by OpenAI.

"Any firm can buy GPT capabilities and gain value but to be disruptive it will take smart investments not only in data but also in design, process engineering, strategic pricing, and master's-level AI model expertise," David Cunningham, chief innovation officer for Reed Smith, told Law360 Pulse.

Most top legal data for a law firm comes from internal sources, such as records and court cases.

Baker McKenzie is another BigLaw firm that is testing out the new AI platform from LexisNexis.

"AI, we view it as a tool. A tool can be used well, or it can be used really badly and create a lot of damage."



THOMAS I. BARNETT
principal and chief data officer
for Jackson Lewis PC

Danielle Benecke, founder and global head of Baker McKenzie's machine learning practice, told Law360 Pulse that the quality of data is a key variable in the performance of AI models, and that building on these models requires domain and task-specific data. This refers to data that is trained specifically on legal matters.

"In a specialized, regulated sector like legal, that means combining — in a controlled environment — the models, enterprise plus external data, and human expertise," Benecke wrote. "To us, the big area of value creation is less about using, say, GPT 'out of the box' to automate lower-rung tasks and more about developing combinations of these things to augment and create new services."

Experts agree that an AI project can be more effective if good data is in place before it's launched.

How to Clean Data

Data cleaning allows an organization to remove bad data, including information created as a result of human errors, such as spelling mistakes, and helps preserve the quality of data for AI models. It also means removing duplicate data and ensuring that data is consistently formatted throughout the file, including dates, capitalized words and names.

A file with human resources-related information needs to have entries of employee data that are in the same format across the document, for example, with the last name of employees before the first name or vice versa, according to Barnett.

"If you don't have confidence in your data and how it's structured, even down to formatting ... you can't really rely on the results," Barnett said.

Experts say that firms should keep track of the source of the data to ensure accuracy.

For a law firm, that might mean going through document management systems and law firm management systems to review the potential data to clean and analyze. Larger firms might have greater challenges with acquiring and standardizing data from systems across departments and geographies, according to Cunningham.

It is obviously easier for firms to clean data if they have a dedicated team for that task. For firms that lack these in-house teams, experts suggest seeking outside help to properly handle data and AI.

Jackson Lewis' staff cleans the data and reviews the quality of data every day to ensure it is in the proper form to be analyzed, according to Barnett. The firm has developed its own homegrown tools for database normalization, which refers to the management of properly structuring data in the same format.

Barnett describes Jackson Lewis as a "data-centric" firm and said its practices around data cleaning have allowed it to get a head start on AI, long before other firms started using it.

"We wouldn't be able to do what we do without thinking about the quality of our data and how we clean it," Barnett said.

What Data Can Do for AI

To get an idea of how a law firm can make the most of its AI projects, let's examine Jackson Lewis. The law firm has been using AI-based tools for a few years now, according to Barnett, long before AI became a trendy

tool in law firms with the release of OpenAI's ChatGPT in November.

Jackson Lewis has a dedicated cross-functional team that experiments with different uses for AI, including legal research, case management, document analysis and operations. Barnett said that the firm has used AI in the past for textual analysis in discovery so that teams could answer questions about data rather than just searching for different terms in documents.

The release of OpenAI's GPT technology opened the floodgates for some law firms to experiment with it. But the problem with using the public-facing version of ChatGPT is that the data may not be secure. It logs all conversations and the data could be shared on the web, which creates a conflict of interest for law firms that must protect client confidential information.

To harness this technology while being cognizant of data privacy concerns, Jackson Lewis has its own private version of GPT that it received through Microsoft. This allows Jackson Lewis' team to think carefully about the types of data the firm is accessing from its long history of operations. And because the firm is using its own, cleaned data, it knows that the results are more reliable and trustworthy.

Barnett said the firm has experimented with its own version of GPT to use in different practice areas and to use as a foundation for document creation and analysis.

"AI, we view it as a tool," Barnett said. "A tool can be used well, or it can be used really badly and create a lot of damage."

That damage is already on display in the legal sector. Two lawyers from a New York personal injury law firm were sanctioned after they submitted a brief in federal court in May that contained nonexistent cases and opinions produced by ChatGPT. In June, a talk-radio host filed a defamation lawsuit against OpenAI after ChatGPT allegedly created a bogus legal complaint.

These examples fall under the category of hallucinations, which is when a generative AI tool makes incorrect claims that may sound plausible.

One reason why generative AI tools such as ChatGPT hallucinate is because of insufficient training data. This is why law firms need to have clean data before training AI models.

With all of these recent hallucinations, experts agree that it is also important to have human input to ensure accurate results.

For firms that are just getting started with AI, Barnett suggests that teams understand how the technology works and approach potential AI tools with healthy skepticism.

"I think this is going to transform the way law is practiced over the years, but you need to make sure you're proceeding with caution so you don't become a casualty of the technology."

And proceeding with caution begins with cleaning data.

--Additional reporting by Ryan Boysen, Matt Perez and Dorothy Atkins. Editing by Nicole Bleier.

Law360 is owned by LexisNexis Legal & Professional, a RELX Group company.
All Content © 2003-2023, Portfolio Media, Inc.