

Q&A: Large language models and the law

May 2 2023, by Pablo Arredondo



Credit: Unsplash/CC0 Public Domain

CodeX—The Stanford Center for Legal Informatics and the legal technology company Casetext recently announced what they called "a watershed moment." Research collaborators had deployed GPT-4, the latest generation Large Language Model (LLM), to take—and pass—the Uniform Bar Exam (UBE). GPT-4 didn't just squeak by. It passed the

multiple-choice portion of the exam and both components of the written portion, exceeding not only all prior LLM's scores, but also the average score of real-life bar exam takers, scoring in the 90th percentile.

Casetext's Chief Innovation Officer and co-founder Pablo Arredondo, JD '05, who is a Codex fellow, collaborated with CodeX-affiliated faculty Daniel Katz and Michael Bommarito to study GPT-4's performance on the UBE. In earlier work, Katz and Bommarito found that a LLM released in late 2022 was unable to pass the multiple-choice portion of the UBE. "GPT-4 Passes the Bar Exam," their working paper recently released on the *SSRN Electronic Journal*, quickly caught the national attention. Even The Late Show with Steven Colbert had a bit of comedic fun with the notion of robo-lawyers running late-night TV ads looking for slip-and-fall clients.

However for Arredondo and his collaborators, this is serious business. While GPT-4 alone isn't sufficient for professional use by lawyers, he says, it is the first large language model "smart enough" to power professional-grade AI products.

Here Arredondo discusses what this breakthrough in AI means for the legal profession and for the evolution of products like the ones Casetext is developing.

What technological strides account for the huge leap forward from GPT-3 to GPT-4 with regard to its ability to interpret text and its facility with the bar exam?

If you take a broad view, the technological strides behind this new generation of AI began 80 years ago when the first computational models of neurons were created (McCulloch-Pitts Neuron). Recent

advances—including GPT-4—have been powered by neural nets, a type of AI that is loosely based on neurons and includes [natural language processing](#). I would be remiss not to point you to the fantastic article by Stanford Professor Chris Manning, director of the Stanford Artificial Intelligence Laboratory. The first few pages provide a fantastic history leading up to the [current models](#).

You say that computational technologies have struggled with natural language processing and complex or domain-specific tasks like those in the law, but with advancing capabilities of large language models—and GPT-4—you sought to demonstrate the potential in law. Can you talk about language models and how they have improved, specifically for law? If it's a learning model, does that mean that the more this technology is used in the legal profession (or the more it takes the bar exam) the better it becomes/more useful it is to the legal profession?

Large language models are advancing at a breathtaking rate. One vivid illustration is the result of the study I worked on with law professors and Stanford CodeX fellows Dan Katz and Michael Bommarito. We found that while GPT-3.5 failed the bar, scoring roughly in the bottom 10th percentile, GPT-4 not only passed but approached 90th percentile. These gains are driven by the scale of the underlying models more than any fine-tuning for law. That is, our experience has been that GPT-4 outperforms smaller models that have been fine-tuned on law. It is also critical from a security standpoint that the general model doesn't retain, much less learn from, the activity and information of attorneys.

What technologies are next and how will they impact the practice of law?

The rate of progress in this area is remarkable. Every day I see or hear about a new version or application. One of the most exciting areas is something called Agentic AI, where the LLMs (large language models) are set up so that they can "themselves" strategize about how to carry out a task, and then execute on that strategy, evaluating things along the way. For example, you could ask an Agent to arrange transportation for a conference and, without any specific prompting or engineering, it would handle getting a flight (checking multiple airlines if need be) and renting a car. You can imagine applying this to substantive legal tasks (i.e., first I will gather supporting testimony from a deposition, then look through the discovery responses to find further support, etc).

Another area of growth is "mutli-modal," where you go beyond text and fold in things like vision. This should enable things like an AI that can comprehend/describe patent figures or compare written testimony with video evidence.

Big law firms have certain advantages and I expect that they would want to maintain those advantages with this sort of evolutionary/learning technology. Do you expect AI to level the field?

Technology like this will definitely level the playing field; indeed, it already is. I expect this technology to at once level and elevate the profession.

So, AI-powered technology such as LLMs can help to close the access to justice gap?

Absolutely. In fact, this might be the most important thing LLMs do in the field of law. The first rule of the Federal Rules of Civil Procedure exhorts the "just, speedy and inexpensive" resolution of matters. But if you asked most people what three words come to mind when they think about the legal system, "speedy" and "inexpensive" are unlikely to be the most common responses. By making attorneys much more efficient, LLMs can help attorneys increase access to justice by empowering them to serve more clients.

We've read about AI's double-edged sword. Do you have any big concerns? Are we getting close to a "Robocop" moment?

My view, and the view of Casetext, is that this technology, as powerful as it is, still requires attorney oversight. It is not a robot lawyer, but rather a very powerful tool that enables lawyers to better represent their clients. I think it is important to distinguish between the near term and the long term questions in debates about AI.

The most dramatic commentary you hear (e.g., AI will lead to utopia, AI will lead to human extinction) is about "artificial general intelligence" ("AGI"), which most believe to be decades away and not achievable simply by scaling up existing methods. The near term discussion, about how to use the current technology responsibly, is generally more measured and where I think the [legal profession](#) should be focused right now.

At a recent workshop we held at CodeX's FutureLaw conference, Professor Larry Lessig raised several near-term concerns around issues like control and access. Law firm managing partners have asked us what this means for associate training; how do you shape the next generation of attorneys in a world where a lot of attorney work can be delegated to

AI? These kinds of questions, more than the apocalyptic prophecies, are what occupy my thinking. That said, I am glad we have some folks focused on the longer term implications.

More information: Daniel Martin Katz et al, GPT-4 Passes the Bar Exam, *SSRN Electronic Journal* (2023). [DOI: 10.2139/ssrn.4389233](https://doi.org/10.2139/ssrn.4389233)

Provided by Stanford University

Citation: Q&A: Large language models and the law (2023, May 2) retrieved 13 May 2024 from <https://techxplore.com/news/2023-05-qa-large-language-law.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.