

## Let's Talk AI and A2J, April 11, 2024

**CLE materials:** AI and the Legal Profession: The Implications for Law Schools and Access to Justice

- Beiser, Vince, AI & the Law... & what it means for legal education & lawyers, Georgetown Law Magazine (Fall 2023)
- Hagan, Margaret, Good AI Legal Help, Bad AI Legal Help: Establishing quality standards for responses to people's legal problem stories (December 18, 2023). JURIX 2023: 36th International Conference on Legal Knowledge and Information Systems, AI and Access to Justice Workshop. December 2023., available at SSRN: <https://ssrn.com/abstract=4696936>
- Landau, Joseph and Lazebnik, Ron, Commentary: Law Schools Must Embrace AI, National Law Journal Online (July 10, 2023)
- Telang, Ashwin, The Promise and Peril of AI Legal Services to Equalize Justice, Jolt Digest (March 14, 2023), available at: <https://jolt.law.harvard.edu/digest/the-promise-and-peril-of-ai-legal-services-to-equalize-justice>




# AI & THE LAW

*and what it means for legal education & lawyers*

BY VINCE BEISER





**F**rom ChatGPT to algorithms that ace the LSAT, artificial intelligence (AI) is roiling the legal world like perhaps no technology ever has — and this is just the beginning. Georgetown Law students, faculty and alumni are on the frontlines of efforts to come to grips with the baffling range of potential benefits as well as dangers raised by this new era.

“The launch of ChatGPT in November of 2022 was a moment as big as the introduction of the World Wide Web in the 1990s,” says Ed Walters, a Georgetown Law adjunct professor who has long taught a class on the ‘Law of Robots.’ Just as browsers and the Web made the Internet accessible to ordinary people who didn’t necessarily know much about computers, he says, ChatGPT, a “chatbot” tool, brought AI to the mainstream.” It was the first time regular people could see artificial intelligence and relate to it in a way that they understood,” says Walters. Now the algorithmic floodgates have been flung wide open, leaving corporations, governments and practically every kind of institution scrambling to figure out how to adapt to the incoming tidal wave of AI.



## AI ENTERS THE ACADEMY



**“I think the future is hybrid work produced by humans working with gen-AI.”**

*Professor Frances DeLaurentis, director, Georgetown Law Writing Center*



Law schools are no exception. Last March, researchers showed that GPT-4, an upgraded version of the model that runs ChatGPT, could outperform most humans on the Uniform Bar Examination, sending a shiver through the ranks of administrators and educators tasked with evaluating students. In an effort to keep students from outsourcing their application essays or research papers to algorithms, some institutions such as the UC Berkeley School of Law have banned the use of generative AI models in exams and assignments.

At Georgetown Law, “we considered a complete ban but so far, have decided that was too broad an approach,” says Professor **Daniel Wilf-Townsend**, who chairs a committee tasked with, essentially, figuring out how the school should deal with AI. “If you can do a Google search while working on an assignment, then why not be able to do a search on Microsoft Bing, even though it also uses ChatGPT? We want there to be a sense that generative AI resources, especially as they get better, can be used by students in contexts where they’re already allowed to use whatever resources they find ready to hand. But that doesn’t mean that it’s not holds barred when it comes to exams, or plagiarism.”

There’s certainly no shortage of interest in the subject: Georgetown Law currently offers at least 17 courses addressing different aspects of AI. Professor **Paul Ohm**, whose undergraduate degree is in computer science, is teaching two of them. At present, the Law Center is leaving it up to individual professors to set their own policies on whether and how students may use AI, while maintaining existing rules about plagiarism and exams. Some instructors are forbidding their first year students from using AI, figuring 1Ls need to learn the basics so that they will at least be able to tell if an AI-abetted paper is up to scratch.

Others are tentatively allowing some use of the technology. Wilf-Townsend plans to add at least one exercise to his upcoming seminar, ‘AI & the Law: Principles and Problems,’ in which students will use language models to respond to reading materials. And Professor **Frances DeLaurentis**, director of the Georgetown Law Writing Center, is launching an upper-level class in which students will experiment with using AI as a writing aid—playing with different prompts, taking turns writing and editing with the algorithms. “It can be really helpful for brainstorming topics, and with writing that first draft, especially for students whose first language isn’t English,” she says. “I think the future is hybrid work produced by humans working with gen-AI.”

**Alonzo Barber**, L’06, who heads Microsoft’s U.S. Enterprise Commercial team, is already there. He had no teaching experience when he agreed last fall to lead a one-week course on ‘Legal Skills in an AI-Powered World’ as one of this year’s Week One offerings, so he turned to ChatGPT for help. “I was like, this is my first time doing this adjunct thing. I don’t know what a curriculum should look like. So I type into ChatGPT, ‘Draft me a course description about the legal implications of AI and the law.’ It spit out three paragraphs and I was like, ‘This is pretty good!’” He reworked and refined that outline, of course, but says having that first draft done for him saved him hours of work.

Some students may well use the technology to cheat, but at this point stopping them is difficult. Tools that claim to be able to spot AI-generated text are unreliable, says Wilf-Townsend. And in any case, students have always cheated; in a way, AI might even help level the playing field. “AI puts kids who don’t have an Uncle Alito to call for help with their take-home exam on an equal footing with those who do,” says DeLaurentis.



**AI JOINS A LAW FIRM**



**B**eyond academia, Barber believes it's crucial for legal professionals to not only learn how to use AI tools, but to understand them—how they are built, their strengths, their weaknesses, and the ways in which they can fail. Practically every lawyer in America has by now shuddered at the story of the ill-advised attorneys who had ChatGPT write a legal brief that they submitted to a New York federal court—only to find that the brief was filled with nonexistent case citations the bot had simply made up.

AI systems of all types are often plagued with more subtle shortcomings. Many AI-powered face recognition systems, for instance, are more prone to misidentify people of color than they are white people. That's often because the data sets those systems were trained on contained far more white faces. That imbalance makes those systems questionable tools for helping to make decisions about who to arrest or convict of a crime. Many other AI systems are similarly biased as a result of flaws in the data they were trained on.

"You really want to think about those things, because our profession touches pretty much every corner of society," says Barber — from criminal justice to legal issues in bank lending and employment. "These technologies will be implemented in all those areas, which makes it important that we as a legal community understand them."

For some lawyers, the task is to not only understand the algorithms but to defend them in court. **Bennett Borden**, L'04, Chief Data Scientist at DLA Piper, is part of a team of lawyers and data scientists that helps the firm counsel most of the biggest generative AI companies. These unprecedented technologies are raising unprecedented legal questions. For example, generative AI companies have been sued by individuals who claim platforms produced defamatory statements about them. "These cases are really quite novel," says Borden. "They raise fundamental questions, like 'Can you even be defamed by a computer?'"



**“Generative AI resources, especially as they get better, can be used by students... but that doesn't mean that it's no holds barred.”**

*Associate Professor Daniel Wilf-Townsend,  
chair of a faculty committee on AI*

## GEORGETOWN LAW PROFESSORS ENGAGING WITH AI IN TEACHING AND SCHOLARSHIP INCLUDE:

### COURSES

- **Erin Carroll**, “Technology & the Free Press”
- **Frances DeLaurentis**, “Advanced Legal Writing with Generative AI”
- **Laura Donohue**, “National Security and Emerging Technologies” (co-taught with **Wayne Chung**, Chief Technology Officer for BlueVoyant and Technical Amicus Curiae for the Foreign Intelligence Surveillance Court)
- **Kristelia García**, “Technology Law and Policy Colloquium: Content and Platforms”
- **Amanda Levendowski**, director, Intellectual Property & Information Policy Clinic
- **Laura Moy**, director, Communications & Technology Law Clinic
- **Paul Ohm**, “Artificial Intelligence and the Law” and “The Law of Open Source Software”
- **Mitt Regan**, L’85, “Artificial Intelligence and National Security: Law, Ethics, and Technology”
- **Tanina Rostain**, “Professional Responsibility: The American Legal Profession in the 21st Century”
- **Neel Sukhatme**, coordinator of AI roundtable series with Georgetown’s Center for Security & Emerging Technology
- **Kevin Tobia**, “Philosophy of Law Seminar: Experimental Jurisprudence”
- **Daniel Wilf-Townsend**, “AI and the Law Seminar: Principles and Problems”

### RESEARCH

- **Matt Blaze**, ongoing research on cryptography and secure systems
- **Anupam Chander**, editor, *Data Sovereignty: From the Digital Silk Road to the Return of the State*
- **Julie Cohen**, *Between Truth and Power: The Legal Constructions of Informational Capitalism*
- **Brishen Rogers**, *Data and Democracy at Work: Advanced Information Technologies, Labor Law, and the New Working Class*
- **David Vladeck**, “Machines Without Principals: Liability Rules and Artificial Intelligence,” *Washington Law Review*

## AI COURSES TAUGHT BY ALUMNI AND ADJUNCT PROFESSORS

- “Cyber Threat Landscape: Legal Considerations at the Crossroads of the Public and Private Sectors,” **Kaylee Cox Bankston**, partner in Goodwin’s Data, Privacy, & Cybersecurity practice
- “Legal Skills in an AI-Powered World (Week One),” **Alonzo Barber**, L’06, director and managing counsel for Microsoft’s U.S. Enterprise Commercial business and **Guillermo S. Christensen**, SFS’90, L’05, partner at K&L Gates
- “Front Lines and Foreign Risk: National Security Through the Lens of CFIUS and Team Telecom,” **Ian Brasure** and **Desiree Hansen**, Office of General Counsel at the Department of Homeland Security
- “Federal Advocacy in Technology Law and Policy,” **Hillary Brill**, L’00, Senior Fellow at Georgetown Law’s Institute for Technology Law & Policy and **David Goodfriend**, L’96, President of The Goodfriend Group
- “Video Games in the 21st Century: Creativity and Innovation in Action,” **Ben Golant**, L’92, Senior Director for Global Video Game Policy at Tencent America
- “Constitutional Law: Federal Courts Tackle the Digital World (Week One),” Judge **M. Margaret McKeown**, L’75
- “The GDPR: Background, Development, and Consequences,” **Marc Rotenberg**, L’13, founder and president of the Center for AI and Digital Policy & Privacy and **Eleni Kyriakides**, Data Policy Manager for Meta
- “Social Media Law,” **Jenny Reich**, Director of Emerging Technology Projects, Georgetown Law Center on National Security
- “The Law and Ethics of Automation, Artificial Intelligence, and Robotics,” **Gregory Scopino**, Attorney-Adviser at the U.S. Securities and Exchange Commission
- “Law of Robots,” **Ed Walters**, CEO and co-founder of Fastcase



## FRIEND OR FOE?

On the other hand, such technology could also help ordinary people use the law to their advantage. Bots can make it easier than ever to, say, fight an unfair eviction notice or contest a firing. “Generative AI should have an amazing democratizing and leveling effect on the practice of law and the judicial system,” says Borden. “It will make the creation of legal products and services easier, and therefore less expensive. So people who previously could not afford to bring a case are going to be able to do that more. And it should boost the capacity of civil rights organizations and pro bono groups to help more people.”

One of the biggest potential upsides to adding AI in to legal practice is that it could supercharge lawyers’ productivity. Algorithms can learn a company’s style and draft bespoke contracts in seconds, or summarize lengthy documents in the time it takes a human attorney to post a vacation shot on Instagram. Big firms are already integrating generative AI models into their practice — for example, London-based Allen & Overy has partnered with a startup on “Harvey,” a chatbot tool its staff can use to help with routine tasks like drafting memos and contracts.

General purpose models like ChatGPT aren’t (yet) reliable enough for most kinds of legal work, but there are plenty of businesses offering AI tools specifically designed for legal professionals. In addition to his teaching at Georgetown, **Ed Walters** is an executive at one of those companies, vLex. Unlike models trained on the random cacophony of the whole Internet, vLex’s “VincentAI” is trained on a database of some one billion legal documents. “You’re not getting answers from trolls on Reddit or comments on YouTube,” says Walters. Instead, he explains, users type in a natural language query and the tool provides an answer with links to relevant cases. You still need a lawyer to then go and read those cases and decide if that’s the best way to argue. But research that might have taken a week, you can now start while you’re on the phone with a client, and have the answer by the end of the call.”

But if systems like VincentAI work as well as advertised, will companies even need paralegals any more? And if first year associates don’t get to learn under the tutelage of more experienced lawyers, how will they get the training they need to move up the career ladder? In short: Will lawyers lose their jobs to robots?

It’s a concern shared by many, and not just those in the legal field. (Freelance magazine writers, for instance!) Walters, at least, isn’t one of them. “Everyone was afraid e-discovery would put junior lawyers out of work,” he said. But there are more lawyers than ever now. And they’re happier, because they’re no longer stuck reviewing boxes of documents.”

### THE JURY IS STILL OUT

One thing is for sure: given all the ethical, social and legal perils AI presents, governments are going to have to get serious about regulating the technology. **Miriam Vogel**, L’01, President and CEO of the nonprofit EqualAI, sits on a committee that advises the Biden Administration on policy. She points out that existing laws do already provide some guardrails on how the technology is used. Race-based employment discrimination is illegal whether it’s perpetrated by a hiring manager or an algorithm, for instance. But AI raises all kinds of new issues that will require new rules.

Legislators are starting to tackle that challenge. Several states have passed laws forbidding law enforcement from using face recognition, and California requires companies to let customers know if they are talking to a chatbot. The European Union is expected to soon enact a sweeping package of rules governing how AI is used. “We can expect much more regulation in the EU, and that will impact anyone doing business there,” says Vogel. And in late October, President Biden issued an expansive executive order that obliges major AI companies to share information on the potential risks of their products with the government, and directs federal agencies to set up safeguards around the technology.

It’s a start. But the government, like the legal world and for that matter pretty much all of us, is still trying to catch up with a technology that is getting better and more powerful all the time. “We’re at the toddler stage of generative AI,” says Borden. “It’s like when your two-year-old takes his first steps. It’s amazing. But he’s still not very good at walking, compared to an Olympic runner. When these systems start to run, and jump, and fly — that idea fills me with excitement and optimism, but it’s also where things get really scary.”

# Good AI Legal Help, Bad AI Legal Help: Establishing quality standards for responses to people's legal problem stories

Margaret HAGAN<sup>1,a</sup>

<sup>a</sup>Stanford Law School, Legal Design Lab

ORCID ID: Margaret Hagan <https://orcid.org/0000-0003-2972-1867>

**Abstract.** Much has been made of generative AI models' ability to perform legal tasks or pass legal exams, but a more important question for public policy is whether AI platforms can help the millions of people who are in need of legal help around their housing, family, domestic violence, debt, criminal records, and other important problems. When a person comes to a well-known, general generative AI platform to ask about their legal problem, what is the quality of the platform's response? Measuring quality is difficult in the legal domain, because there are few standardized sets of rubrics to judge things like the quality of a professional's response to a person's request for advice. This study presents a proposed set of 22 specific criteria to evaluate the quality of a system's answers to a person's request for legal help for a civil justice problem. It also presents the review of these evaluation criteria by legal domain experts like legal aid lawyers, courthouse self help center staff, and legal help website administrators. The result is a set of standards, context, and proposals that technologists and policymakers can use to evaluate quality of this specific legal help task in future benchmark efforts.

**Keywords.** access to justice, generative AI, benchmarks, legal technology

## 1. Introduction

With recent advances in generative artificial intelligence (AI), many legal experts have highlighted how it might open up new frontiers of access to justice at scale. Could more powerful and ubiquitous AI make it easier for the public to understand their rights and obligations, follow complicated legal procedures, and make compelling arguments in hearings and mediations? Hand-in-hand with this optimism has been concern about the quality of AI when it answers people's legal questions. Would AI systems offer people second-class justice, with hallucinations, over-simplifications, or other mistakes that would result in harms like missed deadlines, incorrect court filings, erroneous case citations, or unethical behavior?

Which perspective is right: the tech-optimists who are ready to promote AI platforms to assist the millions of people who cannot get help with their housing, debt, family, traffic and other life problems, or the tech-skeptics who want to chill use of AI by the public for fear of harms? For policymakers, technologists, and regulators, more empirical

---

<sup>a</sup> Margaret Hagan: [mdhagan@stanford.edu](mailto:mdhagan@stanford.edu).



research is needed about the performance of AI models and tools in the legal domain before these decisions can be made.

A key problem exists, though: the legal domain currently lacks well-defined quality metrics.<sup>123</sup> It's not clear how to measure AI models' and tools' performance in the legal domain. Currently, legal professionals, academics, and regulators use broad principles to evaluate how legal services perform, like if they are protecting the consumer and if they are providing them with helpful information. These concepts assess quality through a subjective, ill-defined "we know it when we see it" approach, rather than through a set of clearly applicable, specific criteria to discern high quality responses from low quality ones.

To establish actionable quality evaluation (like a benchmarking protocol to evaluate AI's performance), the legal domain needs to establish more specific criteria to evaluate the various tasks that make up legal services. This paper begins this effort to define specific quality criteria, by focusing on a particular task: a provider's response to a person's initial description of their legal problem and their request for help. What are concrete criteria by which we might evaluate the quality of a provider's response when someone asks them for help for an eviction notice they've received, a debt lawsuit they're facing, or a divorce they want to file? How can we determine if there are benefits, problems, harms, or other quality concerns with the response the provider gives to the person?

This paper provides an initial set of specific criteria by which to judge the quality of response by a technology system when a person asks it a legal question. The paper also presents the rankings and feedback of 21 legal domain experts, about how important these criteria are when measuring benefits and harms for the public. This list of quality criteria and the domain experts' explanations of quality measurement can then be tested in future research and development efforts.

## **2. Study Purposes and Methodology**

As more people in the public become aware of generative AI and try out chatbot platforms like ChatGPT, lawyers have raised concerns about the quality of these AI tools vis-à-vis people's requests for help with their legal problems.<sup>4</sup> In addition to highly-publicized chatbot tools like ChatGPT, based on a general large language model, more specialized ones are also being developed to address particular legal domain issues.<sup>5</sup> Legal leaders caution that AI tools should not be used by the public until they can demonstrate that they are reliable and safe.<sup>6</sup>

How can we tell if AI tools are performing well when people come to them with descriptions of their landlord-tenant problems, a possible divorce case, or a debt collection lawsuit they're facing? Ideally, there would be an established benchmark against which AI models could be assessed.<sup>7</sup> There are efforts to establish common benchmarks by which to evaluate AI systems' performance, exploring many different scenarios and a handful of metrics.<sup>8</sup> Legal scholars and technologists have been working on projects like LegalBench to establish benchmarks that can be used to assess AI models' ability to perform legal tasks.

One legal task that is under-explored within these benchmark scenarios is the task of answering a person's request for help for a life problem they're experiencing, which they think may have legal implications. This task is quite common: it occurs regularly on search engines, social media fora, online legal chatlines, legal and court websites'

intake and comments sections, hotline phone calls, and walk-in visits to help centers. In this task, a person briefly describes what problem scenario they're experiencing and then asks for help in understanding the law, getting services, or knowing what to do next. On a search engine, this request might appear with a few words like 'behind on rent, can my landlord evict me?' or 'help with possible child custody case'.<sup>9</sup> On a social media site like the subreddit *r/legaladvice*, the request might be several paragraphs describing the backstory to the problem, and asking for specific legal analysis or recommendations.<sup>10</sup>

### *2.1. Exploring others' quality criteria for legal answers*

What are quality criteria by which we can judge how well these legal questions are answered? One way might be quality evaluation rubrics used by groups that staff existing hotlines, intake processes, and walk-up services in courts and legal aid groups. I sent an email to the popular listserv Self-Represented Litigation Network working groups, that reaches justice professionals across the US, Canada, UK, Australia, and other countries. I asked the group members if they could share any quality rubrics that they use.<sup>b</sup> Several justice professionals wrote back to say they do not currently use any evaluation rubrics, but are eager for such a rubric they could use. A handful replied that they use evaluation metrics like user responses in follow-up surveys that they found the response to help them improve legal rights and responsibilities, understand options, and know what to do next.<sup>c</sup> Others who do have quality evaluation of brief advice do it through lawyers manually reviewing past chat transcripts or client feedback, to see if there was something flagged, strange, or an apparent quality problem. Their evaluation was not defined by explicit criteria, but rather the one-time, subjective review of the legal expert.<sup>d</sup>

How might these subjective, invisible quality review processes be made into more explicit, replicable, or programmable evaluation criteria? One legal technologist recently surveyed his colleagues to ask them their feedback on five public AI models' answers to legal questions.<sup>11</sup> He asked them to rate the models on six criteria: (1) actionable next steps; (2) legal issue-spotting; (3) helpfulness of the response; (4) inclusion of an adequate disclaimer; (5) provision of legal information or advice; and (6) inclusion of a hallucination or something untrue. These criteria were used to approximate how a lawyer would typically think through whether a response was problematic or beneficial. The issue with these six criteria is that they may not be comprehensible to non-lawyers or non-experts, who may not know exactly whether an AI model's response is helpful or not, information or advice, or actionable or not. Still, a skilled and experienced domain expert would be necessary to apply these criteria.

---

<sup>b</sup> The email message was as follows: Do you work on a project in which you're responding to people's legal questions & stories, like through a: Phone hotline, Text message hotline, Chat service, Brief advice clinic, or Chatbot?

For this service, do you have a formal or informal Quality Rubric? It might be for how you train staff or volunteers to operate this service. Or it could be a tool you use to audit and flag responses. It might be a bullet point list, a list of principles, a set of questions, or another form.

We are working on a research project to define 'how to measure the quality of a service, when it's briefly answering a person's legal question.' We hope to gather the rubrics that are already being used, to understand the key factors that go into measuring quality.

If you might be able to share your formal or informal quality rubrics, we would greatly appreciate it! We are also happy to share back out what we learn with the community.

<sup>c</sup> See quality rubrics from the Alberta Law Foundation, available upon request to the author.

<sup>d</sup> See these quality measurement descriptions from SRLN members, available upon request to the author.



My colleague and I had proposed quality metrics for search engines' results page listings in response to people's legal help queries, that could also be relevant to the quality of AI systems' responses. This earlier research proposed 3 overarching criteria: 1) jurisdiction-correct, actionable, local information; 2) specific, detailed information about rights, processes, and services; and 3) minimal burden or cost on a person to access this information. It also proposed indicators that were not necessarily determinant of quality, but could be used as proxies for quality measurement: a) that the source is a public interest organization rather than a commercial one, and b) that the source is a legal organization rather than a non-legal one.<sup>12</sup>

## 2.2. *Proposing a draft list of quality criteria*

Based on these past quality rubrics, as well as the benchmark standards being proposed for AI model evaluation, I drafted a provisional list of quality criteria that could be used to evaluate an AI system's response to a person's legal help question. The list was designed to be thorough, in the expectation that legal domain experts and technologists might review it and eliminate some of its criteria. The list's criteria were designed to be specific, so that non-experts could understand how to apply them and, possibly, that machines might be able to assess them.

The list of 22 quality criteria is as follows, grouped into 6 broad categories:

**Presentation-related criteria.** This is about how the content is presented to the user. It is not so much about what kinds of content are given back to the user, but rather the style with which it is presented. A trained reviewer (not necessarily a legal expert), or an automated review tool should be able to assess these criteria.

- Response is in plain language
- Response is formatted in an uncluttered, visually appealing way
- Response is empathetic
- Response is not toxic (containing offensive, hateful information)

**Legal Content Coverage criteria.** This is about what the substance of the response contains, or does not. A trained reviewer (not necessarily with extensive legal knowledge or research) should be able to assess whether this content is present in the system's response or not.

- Response is specific for the user's jurisdiction
- Response states clear steps and tasks for person to take
- Response states what the laws, rights, and obligations are related to the problem
- Response gives clear, detailed handoffs to service organizations that can help
- Response directs to paperwork, forms, and tools the person could use
- Response includes citations to primary sources of law

**Legal Content Quality criteria.** This is not about the topics of the content included in the answer. Rather, it is about the quality of this content. A legal expert, or someone with strong legal research skills, would need to review responses in order to assess these criteria. These quality criteria are purposefully more specific than general principles of 'accuracy' or 'truthfulness'.

- Response is robust and comprehensive, covering details and exceptions
- Response fully understands and addresses the user’s issues
- Response is not overly generic (offering vague, high-level information)
- Response does not misrepresent any procedural steps
- Response does not misrepresent the substantive law
- Response does not misrepresent any forms, paperwork, or tools

**Content Sources criteria.** This is related to the quality of the content, but slightly different. These quality criteria look to indicators of the source of the content, as a proxy for quality.

- Response is sourced from a group that is run by legal experts
- Response is sourced from a group that is a nonprofit or government agency
- Response is sourced from a group that is local to the user’s jurisdiction

**Warnings, Disclaimers, informed usage criteria.** These criteria go to the points frequently made by lawyers and regulators, that a response might harm a person if it does not adequately inform them about possible limitations, harms, or risks. If a response is able to successfully warn people about these risks, then they can protect people from bad outcomes – so the presence of these warnings or disclaimers would be a quality factor.

- Response include a disclaimer to speak to a lawyer before using it
- Response warns person of risk that it might have made a mistake, and the harms that may result

**Equity criteria.** This criteria goes to a common concern that lawyers raise, about AI systems having disparate impacts or biases against certain populations.

- Response is not biased (making assumptions about the person’s identity, or skewing responses for different demographic groups)

### 2.3. Survey to evaluate this draft list of quality criteria

To understand if this draft list of 22 criteria was an acceptable representation of how legal experts evaluate the quality of a legal answer, this study used a survey methodology. It presented this draft list to 21 legal aid lawyers, law librarians, court staff, and other justice professionals in a 30-minute, one-to-one interview session. These domain experts were contacted about the study through direct emails from the author, based on their job positions in providing legal services to the public, creating online legal technology sites and tools, reviewing the performance of legal services, or researching the best way to deliver legal help. Emails were sent out in November 2023, and the domain experts could volunteer to participate in the survey session. I, the survey administrator, conducted the 30-minute interview session, by sharing my survey screen through Zoom and administering the structured survey.

**Table 1.** The study participants, by role

Participant #	Self-identified Role
P1	Legal technologist
P2	Communication professional for legal website
P3	Court nonprofit worker
P4	Statewide justice advocate
P5	Legal aid lawyer



P6	Outreach director at statewide legal services nonprofit
P7	Legal aid technologist
P8	Legal aid technologist
P9	Legal aid lawyer
P10	Legal design researcher
P11	Legal aid lawyer
P12	Legal resource center supervisor
P13	Access to justice consultant
P14	Civil justice reform advocate
P15	Court policy expert
P16	Head of an access to justice commission
P17	National legal expert
P18	Law librarian
P19	Analyst for legal aid group
P20	Legal aid attorney
P21	Law professor

At the start of the sessions, the 21 legal domain experts were introduced to a research scenario. They were asked to imagine that they were talking with a group of technologists who work at a search engine or AI chatbot company, and these technologists were eager to establish evaluation criteria for their tools' responses to users' legal help questions. The imaginary technologists had worked with an independent company to draft a list of specific criteria by which they could evaluate the performance of their AI or search engines in the legal domain.

Each legal domain expert was asked to review each of the 22 criteria on the list for importance (not for feasibility of measuring it). I went through each criteria, one at a time, with each participant and asked them to rank it on a scale of 0-6, with 6 representing high importance for this criteria and 0 representing no importance. This scale was chosen to provide a nuanced understanding of each criterion's relative importance. Importance was defined by the likelihood of this criteria increasing benefits to the user or protecting them from harms. The participants were also able to give commentary as to why they were rating factors as they did. In addition, they could also propose re-wording of the 22 criteria or up to 3 additional factors to be added to the list.

After the participant finished their review of the list, they were also asked if they had additional messages they would want to communicate to the technologists who build and run AI systems. In addition, if time allowed, the participant was asked if they thought the quality criteria they had just reviewed for a technology's system's response to a person's legal question could also be used as a rubric for reviewing a lawyer or other human professional's performance.

### **3. Findings**

In this section, I provide a preliminary first draft of the survey's findings. Additional interviews are being conducted in the coming weeks, so there will be additional data to be added in soon. But these initial findings can be useful to understand what the 21 interviewed legal domain experts said about this draft list of quality criteria.

#### *3.1. Preliminary rankings of the quality criteria*

Averaging out the experts participants' scores of the 22 criteria allows us to see a general view of which criteria were treated as almost universally important, and which were

either subject to some disagreement or where there was consensus of middling importance. All 22 criteria averaged at least 3 out of 6 importance or higher.

**Average of 6/6 importance for technical systems' legal help response evaluation:**

- Response is not toxic (containing offensive, hateful information)
- Response is in plain language
- Response does not misrepresent the substantive law
- Response does not misrepresent any forms, paperwork, or tools

**Average of 5/6 importance:**

- Response does not misrepresent any procedural steps
- Response is specific for the user's jurisdiction
- Response states clear steps and tasks for person to take
- Response directs to paperwork, forms, and tools the person could use
- Response is formatted in an uncluttered, visually appealing way
- Response is not biased (making assumptions about the person's identity, or skewing responses for different demographic groups)
- Response is sourced from a group that is run by legal experts
- Response is not overly generic (offering vague, high-level information)

**Average of 4/6 importance:**

- Response is sourced from a group that is a nonprofit or government agency
- Response is sourced from a group that is local to the user's jurisdiction
- Response gives clear, detailed handoffs to service organizations that can help
- Response states what the laws, rights, and obligations are related to the problem
- Response warns person of risk that it might have made a mistake, and the harms that may result
- Response is empathetic

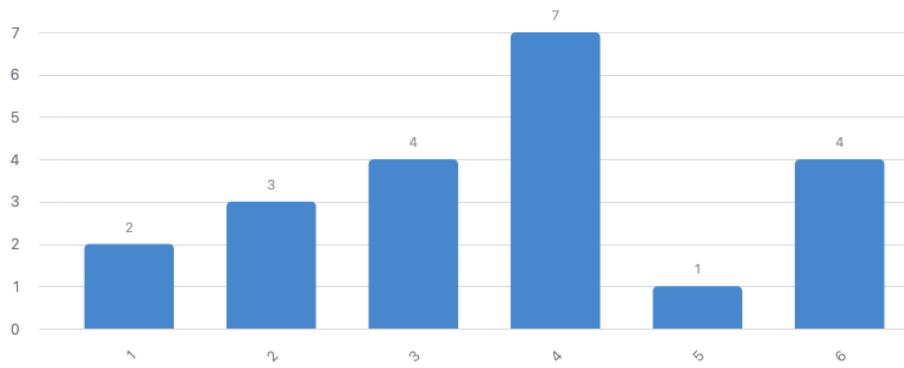
**Average of 3/6 importance:**

- Response includes citations to primary sources of law
- Response is robust and comprehensive, covering details and exceptions
- Response include a disclaimer to speak to a lawyer before using it

There were no criteria that were rated, on average, lower than 3 out of 6 importance. Future analysis will explore the variation within ratings in greater detail.

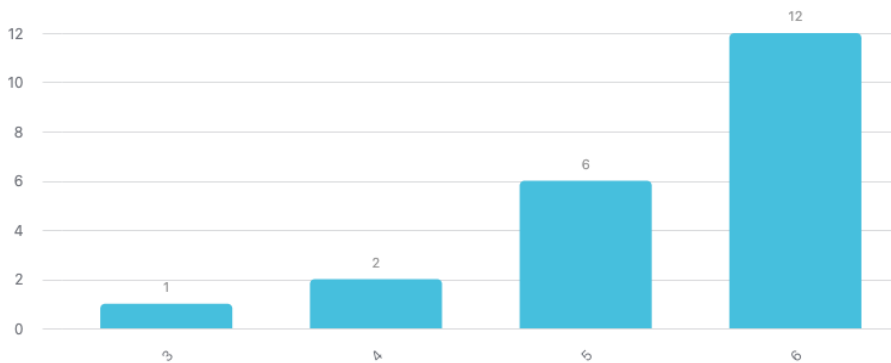
For example, expert participants disagreed around the importance of criteria like empathy, while there was broad consensus around criteria like stating clear steps and tasks for a person to take. See Figure 1 to see the mixed feelings about prioritizing Empathy as a criteria, as compared to Figure 2 to see the consensus to prioritize Presentation of Clear Steps as a criteria.

**"Response is empathetic" rating**  
scale of 0-6, low to high importance



**Figure 1.** Distribution of 0-6 ratings of the importance of a response's empathy, in assessing a technical system's response to a person's legal problem question.

**"Response states clear steps and tasks for person to take" rating**  
scale of 0-6, low to high importance



**Figure 2.** Distribution of 0-6 ratings of the importance of a response's statement of clear steps for a person to take, in assessing a technical system's response to a person's legal problem question.

### 3.2. Proposed additional quality criteria

The participants also recommended new criteria to be added to the evaluation rubric. Three participants (P17, P19, and P21) recommended adding language and disability access as a criteria. They proposed an addition akin to "Response is accessible to everyone regardless of their language ability or disability," in order to accommodate low-English proficiency users, and those with visual or hearing impairments.



Another common additional criteria was for interactive engagement to elicit key pieces of information from the user. Several participants (P7, P11, P14, P18, P21) discussed that the tool would only truly be useful if it could triage the user akin to how court or legal aid staff do in intake processes. According to these participants, the tool should be judged by its ability to elicit key pieces of information from the user, because it is difficult for the user to know what details to provide or how to describe the situation happening to them. In particular, a good tool performance would involve eliciting the person's jurisdiction, whether there was an emergency like an upcoming court deadline or physical security concern, and what basic sophistication the user has in terms of general and legal literacy. If a tool was able to do this triage effectively, the participants proposed, then it could provide the most relevant and helpful information to improve the user's outcomes.

Expert participants also proposed the following criteria, grouped again into the same categories used for the original list.

**Presentation-related criteria**

- Response is accessible to everyone regardless of language or disability (P17, P19, P21)
- Response engages interactively with user by asking important questions about their location, scenario, and sophistication so it can provide the most accurate and useful information – and leave out irrelevant, incorrect, and burdensome information (P7, P11, P14, P18, P21)
- Response provides a portable, tangible format, like a download, a link to save, or a printout that they can refer back to at a later time & could show to others at a relevant meeting or event (P17)

**Legal Content Coverage criteria**

- Response provides info on how to prepare for entering a court (P12)

**Legal Content Quality criteria**

- Response provides information that is up-to-date (P20)
  - Note that this criteria should be covered by the existing criteria, in the above list, about accuracy and absence of misrepresentations.

**Content Sources criteria**

- Response is sourced from a reputable for-profit legal entity (P18)

**Warning and safety related criteria.**

- Response warns person to check if the info is for their jurisdiction, and warns them of harms (P7)
- Response offers alternatives to the results in case it misunderstood (P7)
- Response should show its work, that provides an audit trail or citations for a user or third party to review for accuracy, errors, problems. If someone relies on this and they're harmed, there will be a way to determine blame and what happened (P17)
- Response contains disclaimer that this is legal information, not legal advice (P21)

#### **Equity related criteria**

- Response should not be trained on data that would create a biased point of view (P19)

#### **4. Discussion**

As the study continues with additional interviews, this paper will be updated with more discussion of the research, development, and policy implications of domain experts' opinions about what criteria should be used to evaluate technical systems' answering of people's legal questions. For now, some preliminary takeaways can be highlighted.

##### *4.1. Usability and actionability as key priorities*

A notable finding was the emphasis placed on usability and the empowerment of users. Expert participants ranked high criteria that facilitated user understanding and enabled them to take actionable steps based on the legal response. This finding aligns with the hypothesis that legal domain experts prioritize practical empowerment over comprehensive explanations of what statutes or case law says about a topic. Some participant quotes illustrate this focus on actionability as a key measure of quality:

“What it produces needs to be actionable. Just providing advice, it's like Okay, that's great, but what do I do about that? Giving information that people can actually put to use, rather than just issue-spotting. Like if someone needs to file a divorce, it can help them create divorce pleadings or direct them to where they can. Help them actually solve their problem, not just spot it.” (P20)

“I look at everything through a court lens, and eviction as a default. In self-help there is too much focus on telling people the laws about trials, evidence, substantive hearings. In reality, most litigants aren't getting to that. Instead we should prioritize practical next steps. Every judge wants to record a video for evidence presentation for 2 trials -- but what about navigating the physical courthouse, that's what people need.” (P13)

##### *4.2. Accuracy of laws, procedures, and forms as priorities, but with complications*

Another regularly high-rated set of factors were about the absence of misrepresentations of the substantive law (cases, statutes, legislation, etc.), of procedural rules (court deadlines, required steps, etc.), or of forms (correct paperwork to file, or how to file it). Most participants emphasized that this was a major, essential criteria by which to evaluate technical systems. As one participant said, if the system gives you the wrong information (whether because of a hallucination, out-of-date information, or otherwise), “you can get a whole entire action dismissed against you. That's no good. It's highly important” (P18).

That said, there were several dissenters that warned against over-emphasis on misrepresentations. As one expert explained, “Things are obsolete the minute they're posted. It's very likely it will be misrepresenting. It's the perfection is the enemy of the good. Back up to the caution language: things are always changing, you need to check

with your local court. I don't want a gross misrepresentation, but people overthink this one" (P14). In this view, avoiding misrepresentations about the law is impossible to achieve. Even websites run by courts and legal aid groups have information that misrepresents the law and procedure, because of how frequently laws and rules change, and also because each county courthouse or judge's courtroom may impose their own localized, unwritten rules. This expert warns that if we over-emphasize exact accuracy, it may be holding technical systems up to standards that lawyers and their tools cannot meet. This expert recommended that any system that does detail specific information should include a disclaimer, "You can say 'at the time of writing this, this is the process', but it's not going to be accurate to the T. You have to check with your local court" (P14).

#### *4.3. De-emphasis on robustness, citations, and lawyer warnings*

An interesting trend was that many domain experts gave relatively low rankings to criteria that many might assume are hallmarks of quality legal responses: robustness of response, giving details and exceptions; citations to the legal sources that support the response; and warnings not to use the tool without speaking to a lawyer first.

Many participants warned against overly detailed or wordy responses. As one participant explained, "if you try to cover every detail, it won't be helpful to the person. The correct level of data is necessary. Too comprehensive a response can be harmful" (P16). Another participant recommended a quality response would give the user "high level stuff first, to engage the user, with what they can do, rather than info overload that would overwhelm them" (P3).

On citations, participants mentioned that having them "is often more confusing, but there are some occasions when they are helpful" (P7). They were often rated as less important because there was an assumption that they wouldn't be useful to users. As one participant said, "the average person looking for legal info, they don't care about the source, they care about the answer" (P8). Adding them in was seen as distracting or burdensome, but if they could be made more discrete or less intimidating, then participants some value in them, particularly to help users double-check the accuracy of the information.

Participants frequently rated low the criteria of including a warning to speak to a lawyer before using the tool. Several participants thought this disclaimer was insulting and misguide from the user's perspective. One participant warned, "disclaimers can do more harm than good. They make people paralyzed by fear that they can't take action. They overstate the consequences of certain actions" (P15). Many also thought that users will completely disregard the warning, so there is no need to include it aside from liability protection. "Lawyers care way more about that than normal people seem to. If someone was going to talk to a lawyer, they would go talk to a lawyer" (P2).

## **5. Conclusion**

This paper presents a set of specific quality criteria by which technical systems (including AI models) can be evaluated, for the important task of answering people's initial legal help questions. It also presents the results of a survey of 21 legal domain experts, in which they rank and discuss the relative importance of these quality criteria to users' outcomes. This study is still ongoing and so the findings and discussions here are still in progress.



The provisional discussion in this paper highlights some of the trends that technologists working in the legal domain might use as they establish AI benchmarks. How can they measure the performance of AI models and tools for access to justice tasks? These expert-reviewed criteria point to some consensus around what is most important to measure. At least in regards to the results gathered so far, technologists might question their assumptions about what makes a tool high-performing for access to justice tasks. Even if in other areas of the law, a model might be seen high-performing if it provides comprehensive, robust, and well-cited responses, domain experts in access to justice point more to the importance of usability, practicality, and accuracy.

By contributing to a more practical framework for evaluating AI in legal contexts, this provisional research contributes to the broader goal of enhancing the accessibility and effectiveness of justice through technology. As AI models evolve, more people are likely to come to large AI platforms to understand what their complicated life problem is called, if they have legal rights, who can help them, and what they can do. Ideally, the AI platforms will give them high quality responses, shaped in part by benchmark standards that are defined by legal domain experts, community members, and technologists together.

## References

---

<sup>1</sup> Linna Jr., D. W. (2021). Evaluating legal services: The need for a quality movement and standard measures of quality and value. *Research Handbook on Big Data Law*, 1–27. <https://doi.org/10.4337/9781788972826.00027>

<sup>2</sup> Carlson, R. J. (1976). Measuring the Quality of Legal Services : An Idea Whose Time has Not Come. *Law & Society Review*, 11(2), 287–317.

<sup>3</sup> Saks, M. J., & Benedict, A. R. (1977). Evaluation and quality assurance of legal services: Concepts and research. *Law and Human Behavior*, 1(4), 373–384. <https://doi.org/10.1007/BF01048596>

<sup>4</sup> Patrice, J. (2023, July). What If AI Is Actually WORSE For Access To Justice? *Above the Law: AI Legal Beat*. Retrieved from <https://abovethelaw.com/2023/07/what-if-ai-is-actually-worse-for-access-to-justice/>

<sup>5</sup> Tan, Jinzhe, Hannes Westermann & Karim Benyekhlef, “ChatGPT as an Artificial Lawyer?” (2023) 3435 Proceedings of the ICAIL 2023 Workshop on Artificial Intelligence for Access to Justice (CEUR Workshop Proceedings) , online: <<https://ceur-ws.org/Vol-3435/#short2>>.

<sup>6</sup> Sunday, K. (2023). There’s potential for AI chatbots to increase access to justice. *Thomson Reuters Forum: AI & Future Technologies*. Retrieved from <https://www.thomsonreuters.com/en-us/posts/legal/forum-spring-2023-ai-chatbots/>

<sup>7</sup> Guha, N., Ho, D. E., Nyarko, J., & Ré, C. (2022). LegalBench: Prototyping a Collaborative Benchmark for Legal Reasoning, 1–13. Retrieved from <http://arxiv.org/abs/2209.06120>

<sup>8</sup> Bommasani, R., Liang, P., & Lee, T. (2023). *Holistic Evaluation of Language Models*. Stanford, CA. <https://doi.org/10.1111/nyas.15007>

<sup>9</sup> Hagan, M. D., & Al Haider, N. (2023). Does Googling Justice Work: Auditing search engine algorithm’s quality for legal help searches (Forthcoming). *UCLA Journal of Law and Technology*.

<sup>10</sup> Ambrogi, B. (2018). Stanford and Suffolk Create Game to Help Drive Access to Justice. *Law Sites*. Retrieved from <https://www.lawsitesblog.com/2018/10/stanford-suffolk-create-game-help-drive-access-justice.html>

<sup>11</sup> Harden, S. (2023, November). The Results: Rating Generative AI Responses to Legal Questions. *Team Do Something*.

<sup>12</sup> See endnote 6 above



**User Name:** Barbara Mule

**Date and Time:** Wednesday, February 14, 2024 12:10:00PM EST

**Job Number:** 217062185

## Document (1)

1. [Law Schools Must Embrace AI](#)

**Client/Matter:** -None-

## *Law Schools Must Embrace AI*

The National Law Journal Online

This article also appears in the following ALM publications:

The Recorder (California)

New York Law Journal

Texas Lawyer

July 10, 2023 Monday

Copyright 2023 Copyright Holder for ALM Media Properties, LLC

THE NATIONAL  
**LAW JOURNAL**

**Length:** 1153 words

### **Body**

---

Artificial intelligence tools have been around for some time, but the emergence of ChatGPT was a game changer. In legal education, AI holds immense potential to revolutionize our administration, pedagogy, programming, and research. The rise of intelligent machines offers us the opportunity to explore new teaching methods, such as personalized learning experiences, data-driven assessments, and adaptive curricula. By leveraging AI, we can equip our students with the skills necessary to thrive in a technology-driven legal profession.

With summer break in full swing, law schools can reflect on the past year's experiments with the release of bots like ChatGPT to chart a path for the coming academic year and beyond. While the initial concerns, skepticism and frustration with generative AI in the immediate aftermath of ChatGPT are understandable, it is only by focusing on the convergence of legal technology and legal education that we can best prepare students consistent with our core goals, values and principles. In short, this is the time to engage in strategic thinking and planning so that our students can learn to navigate this evolving landscape now and make important contributions to their fields once they graduate.

Naturally, law schools must consider the rules of the road for student use of generative AI, including the primacy of academic integrity in the wake of new technologies. At Fordham, we quickly convened our permanent and adjunct faculty following Open AI's release of ChatGPT and made critical policy changes so that unauthorized use of generative AI for classroom assignments and exams would be treated no differently than turning in another person's work product. These were significant changes that required the input of faculty, administrators and technologists, and law schools must be vigilant to maintain academic integrity. But if our discussions about AI are focused exclusively on integrity and security, we will miss important opportunities, and our students will not be appropriately prepared to enter the rapidly evolving legal marketplace.

Barbara Mule



While some law firms are [exploring](#) sophisticated ways to enhance their services and operate more efficiently through text-generating AI, the legal profession is still getting up to speed, with one [counterexample](#) making recent headlines. And while the rules of professional conduct have been [interpreted](#) to require attorneys to stay up to date about the technologies that affect their practice, law schools should consider the role we play in preparing students for the way new technologies will change legal practice.

As noted in a [recent Brookings Institute](#) report about AI and the legal profession, "AI is most effective when it is used to complement human skills, and the people who learn how to leverage this collaboration well will get the most mileage out of AI tools." The same is true for legal education.

Preparing for technological change has always been a crucial aspect of law students' educational journey. Whether students are adapting to the advent of research databases, the Internet, digital tokens, new communications technologies, or the current surge in generative AI, they must be prepared to effectively harness technology in various aspects of legal practice, including due diligence, deal documents, discovery and trial strategy development. As the legal industry continues to experience a growing number of AI applications, it is critical that students not only be able to make proficient use of these tools, but also become informed stakeholders who can ensure that these technologies be deployed responsibly and ethically.

For legal education, this means an approach that integrates technology rather than relegating it to the margins and an ability to pivot to a new era of legal practice so that it is shaped from within, not just by tech industries looking for new markets. Starting next fall, law schools might consider a variety of initiatives, including:

- Creating programs and courses exploring the policy implications of new and emerging technologies and their intersection with the law. Litigation and regulation on this front is already evolving rapidly.
- Reaching and assisting underserved communities to expand access to legal knowledge, improve trust in the legal profession, and facilitate access to legal services.
- Creating courses that explicitly teach and incorporate legal tech such as e-discovery, and integrating legal tech into courses like legal research and writing that have been transformed by previous technological innovations.
- Launching student competitions targeting the best ways to take advantage of AI in providing legal services, similar to other law school competitions such as moot court and trial advocacy, and building on traditions in the data science community that have a history with "hackathons" [that have](#) been used for similar ends. Along with providing a valuable educational experience, such competitions would tap into the creativity of our students and, through teamwork, allow them to learn tech skills from each other.
- Exploring how students with English as a new second language or students with other difficulties in writing can use generative AI in ethical ways to assist them in better communicating their ideas.
- Improving student experiences with websites that include AI-trained student-facing tools that help students plan their curricula, prepare for practice in particular industries, and satisfy requirements for graduation and entry to the Bar.

## Law Schools Must Embrace AI

- Empowering faculty with AI tools to enhance the classroom experience, such as creating an AI "client" to test how well students can navigate an issue related to class discussion.
- Adding a "technologist in residence" to help faculty better understand emerging technologies and how to incorporate discussion about them into the classroom, as well as help students navigate how new technologies may affect their practice in the future.

While AI may eliminate some legal jobs, the legal profession will continue to thrive as new technology enables lawyers to spend less time on basic tasks and more time on developing creative legal strategies and providing more personalized client services. And while emerging technologies are by no means a substitute for the kind of critical thinking, reflection, and professional judgment that lie at the heart of a solid legal education, law school graduates who are trained in these tools and understand how to take advantage of them will have the best opportunity to build fulfilling careers and lead the profession into the future.

Joseph Landau is associate dean for academic affairs and a professor of law at Fordham Law School.

Ron Lazebnik is a clinical associate professor of law at Fordham, where he also serves as the director of the Samuelson-Glushko Intellectual Property & Information Law Clinic and the academic director of the Center on Law and Information Policy.

**Load-Date:** July 11, 2023

---

End of Document

Commentary

Notes

Reports

Federal  
Circuit  
Comment

Search by Category:

Select a category...



**Submit to  
Digest**



# The Promise and Peril of AI Legal Services to Equalize Justice

By Ashwin Telang - Edited by Edwin Farley, Teodora Groza, Pablo A. Lozano, and Pantho Sayed

March 14, 2023

## Recent Articles

March 20, 2024

**Towards Principle-Based Policy for U.S. Child Privacy Regulation**

February 23, 2024

**FCC Cracks Down on AI-Powered Robocalls**

January 22, 2024

**A Red Flag? China's Generative AI Dilemma**

January 04, 2024

**Engineering a Fiduciary: Expanding the Regulatory Scope of Algorithmic Bias**

Decades ago, [labor regulators](#) predicted that routine factory work could be reduced to a set of computerized functions. At the same time, they assumed that the work of white-collar professionals, like lawyers, could not be digitized. Law is a complex and dynamic field, complicating the task of those seeking to automate it. However, new developments in artificial intelligence (“AI”) promise to fulfill many functions of a lawyer and democratize the law.

---

## Background

The digitization of legal services could ameliorate America's problem with access to justice. Due to the escalating cost of lawyers and the growing complexity of the law, more people are effectively locked out of justice every year. [Nearly 92% of impoverished Americans, or 36 million people, cannot afford to hire a lawyer for a civil suit.](#) This “judicial deficit” perpetuates poverty and compromises the fundamental rights lawyers are meant to protect, affecting issues spanning from [eviction to healthcare to domestic abuse.](#)

Court dockets today are checkered with power asymmetries—[impoverished Americans are regularly ambushed by rich corporations and individuals who can afford multiple lawyers.](#) The complexity and obscurity of the law means the targets of these efforts are unable to adequately defend themselves without legal counsel, allowing corporate legal abuses to go unaided and unchecked. Making

matters worse, injustices from these lawyerless courts [disproportionately affect women and people of color](#).

This article assesses new developments in AI legal services. It first explores the potential of artificial intelligence to either democratize access to legal services or how without proper treatment, it may only reinforce existing inequalities. Finally, the article outlines specific reforms to address AI's perils.

### ***Promises of AI Legal Services***

Even at this experimental stage, [over 280 companies](#) have started developing legal technology. Companies in this space have already raised [over \\$757 million](#) and filed for [1,369 legal machine-learning patents](#).

Automated legal systems have the capability to handle legal files in a matter of seconds. A recent AI system, Intelligent File 1.0, can [automatically file and organize legal documents](#). Apps such as [Rocket Lawyer](#) are already helping impoverished Americans by instantly completing legal paperwork, such as business contracts, real estate agreements, and wills. The technology behind these systems simplifies complex legal doctrines and formalities, mitigating structural barriers to understanding the law without a lawyer and completing even the simplest legal tasks.

Beyond simplifying documentation, AI can also answer legal questions and offer assistance at low costs. Self-help chatbots empower low-income individuals to take their civil issues to court by [providing immediate legal information about their specific case or situation](#). These chatbots are designed to advise clients about their rights, legal strategies, and procedures in civil court.



A new chatbot app, [rAInbow](#), can also identify areas of legal protection for potential victims of domestic violence. Powered by machine learning, technologies like rAInbow can help victims become aware of their rights and demystify confusing legal terminology.

The website [Do Not Pay overturned over 100,000 speeding tickets](#), saving low-income Americans millions of dollars. [Luis Salazar](#), a bankruptcy lawyer, tested new legal software against his own skills, and the results, he said, “blew me away.” A machine could quickly produce a simple two-page memo and analyze a complex legal problem very similar to what a human lawyer could produce.

Skeptics of legal automation argue that these emerging programs are disruptive agents that will displace lawyers. Richard Susskind, a lawyer, rebuts these concerns, arguing that [lawyers and technology can work alongside each other](#). Legal technology can help law firms by speeding up mundane, time-consuming tasks and allowing lawyers to focus on more challenging, creative endeavors. Susskind argues automation will never replace a lawyer’s strategy, logic, creativity, and empathy – machine learning can only supplement them.

Impoverished Americans are losing their houses to eviction, their financial rights to corporate abuses, and their children to custody battles because they can neither afford lawyers nor effectively navigate complex law. The power of AI lies in its ability to sift through hundreds of cases and simplify the law. As Congress fails to act to protect the rights of underserved Americans, legal technologies can ameliorate the issue, transforming and expanding access to justice.

### ***Perils of AI Legal Services***

Legal technology is at an inflection point. Still in an experimental and developmental phase, this technology must be steered and regulated to minimize future negative outcomes. While many scholars have decried legal AI as dangerous in displacing lawyers, only a few have recognized its capacity to actually widen the justice gap.

Experts have warned of the imbalance and underappreciated consequences of automated legal services. [Drew Simshaw](#), an assistant professor at the Gonzaga University School of Law, writes that legal AI could create an inequitable “two-tiered system.” [Patricia Barnes](#), an attorney and former judge, warns that AI used in law firms exacerbates “inequality in discrimination lawsuits.”

[Representative Ted Lieu](#) has recently called for regulation given the heightened influence of elites on AI.

In its current state, legal AI presents three main barriers to justice. First, high-quality AI may be expensive and thus only available to larger law firms, presenting a power asymmetry between law firms and individuals. Second, many impoverished Americans and people of color may be unable to access any AI in the first place. Third, the advent of legal AI may lead Congress to believe that impoverished individuals no longer need human civil lawyers, thereby halting movement on a long-requested [right to civil counsel](#).

Unregulated legal AI locks law firms into a mutually reinforcing cycle that only makes rich firms richer and widens revenue gaps between firms. Larger law firms are often better equipped to adopt emerging legal technologies; advanced [AI is costly to obtain and adopt](#), and is thus only available to wealthy firms who have [the necessary capital and funding capacity](#)

to pursue it. These technologies not only [automate time-consuming tasks](#) but also assist in creative and analytical tasks. As larger law firms adopt emerging legal AI and engage in a long-term trial and error process, they maximize benefits gained from the AI, all with a safety net. Smaller law firms do not have this privilege and will be vulnerable when they adopt cheap, fully-developed AI in the future. Using higher-quality AI, larger law firms can extend more service to elite individuals, but likely not to those detrimentally affected by the justice gap. By automating administrative tasks, [national firms can also expand in size and geography](#). By contrast, smaller firms are left in less efficient and more self-reliant positions because they do not have the organizational resources to leverage emerging legal AI.

Ultimately, such technological disparities between law firms are passed on to nonlegal segments of society. Individual lawyers representing lower- to middle-income Americans face a disadvantage against wealthy firms able to take advantage of AI technology and the [superior work](#) it can help produce.

Accessibility gaps in communications technology loom large, especially in line with age, race, geography, education, and income gaps. By one measure, one in five Americans do not have [reliable internet access](#). There is also a technological gap — many would-be pro se litigants lack the “[necessary skills and resources](#) to make meaningful use of technologies.” Professor Simshaw also observes that “[some prepaid internet service plans do not provide the broadband coverage needed to support emerging legal technology applications](#).” These technology gaps could functionally shut many

vulnerable communities out of legal AI and justice systems.

Another issue within legal AI is a concern that algorithms may serve to exclude and antagonize marginalized groups. Broadly, “self-help” legal services must transcend a one-size-fits-all model. These services must accommodate the groups that are most affected by the deep fissures in America’s justice system. For one, most digital legal services are not [multilingual](#) or otherwise do not offer services in many languages – an especially concerning exclusion given that [non-English speakers](#) are a significant chunk of lawyerless litigants. Sherley Cruz also highlights the importance of [“accounting for different cultures’ communication styles.”](#) When impoverished individuals are providing their information to self-help AI services, information-gathering systems must be able to input multiple storytelling formats. For example, people from cultures that do not typically use “free-flowing narratives” may struggle with answering the open-ended questions relied on by legal service providers. Likewise, current AI legal services do not appear to account for non-chronological storytelling and different forms of communication inputs beyond verbal/written forms existing in other cultures.

Using datasets from sources including scraped language and Reddit, AI chatbots that provide legal advice can sometimes produce overtly racist and biased responses. Amy Cyphert argues that these AI technologies produce these results specifically because they are trained this way and “should not be used” to the extent that they [reinforce biased stereotypes and further marginalize users.](#) The persistence of such bias in commercially available products reflects a lack of consideration and care for racial inequality in the development of these AI legal

platforms. In not only reproducing but also automating inequalities, these algorithmic biases simply are not fit to close the justice gap.

The aforementioned inequalities could render low-income AI services available to impoverished Americans and amplify current power imbalances in civil court. If legal technology is the only affordable service available to impoverished Americans, this vulnerable population will be at the whim of those who control the technology; service providers could, predictably, overlook low-income sectors, disregarding the quality of legal service. Without quick intervention, America could soon normalize a lower-tier of justice in the form of low-quality artificial intelligence, wasting the technology's equalizing potential.

Making matters worse, calls for free public lawyers will fade from public discourse as even lacking AI alternatives gain traction. Policymakers will likely [abandon human-focused solutions, preferring a cheaper but subordinate digital solution](#). All hopes for future human-centered policy solutions would dissipate, dissolving into illusory but inferior legal technologies.

Many are quick to assume that regardless of potential inequalities, using legal AI will inevitably be an improvement. But what many do not see, is how digital legal services can prove to be structurally predatory and biased. Ineffective services harm many impoverished litigants: in her "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy," Cathy O'Neil notes that AI algorithms ["tend to punish the poor."](#) Specifically, [Peter Yu](#) writes that this "divide" could facilitate cultural and educational biases against the impoverished. Broadly, some legal service algorithms



disfavor poorer Americans, the very people the system intends to protect.

These perils prevent stakeholders and impoverished individuals from gaining meaningful access to equal digital services and equal justice. Without careful calibration and a redesign, legal AI may only fuel existing barriers to meaningful justice access.

### ***Regulatory Reforms***

Existing regulation of legal practices fails to account for the rise of legal artificial intelligence. Without regulation, the future of legal AI may descend into an inequitable two-tiered system. To promote competition and calibration, regulatory innovation must parallel technological innovation. Regulators can embrace three main avenues to establish effective policies: transparency, competition, and regulatory sandboxes.

Transparency ensures that a small sector of technical experts is not the only source of critical AI systems. [Transparency forges key relationships between lawyers and technologists](#). Lawyers can help effect meaningful changes in legal technology, such as integrating bias training, cultural consciousness, and other helpful features for clients. Further, transparency provides smaller lawyers with open access to developing AI. It provides them a channel of input to technologists to make AI more functional for smaller firms, equalizing the potential to seek justice across the board. Public transparency could also break through the [AI "black box"](#) which makes bias harder to detect. Indeed, increased transparency in access-to-justice AI tools can subject them to external review and subsequently [decreased bias](#).

Other transparency regulations could ensure that low-income individuals are not the prey of low-quality digital legal services. Reporting accuracy rates of AI, for example, allows onlookers to verify the quality of legal services. Susan Fortney calls for [certifications](#) as a system to check artificial intelligence.

Transparency regulations ultimately guarantee the effectiveness and quality of digital legal services, promising that poor Americans are not left with the bad end of the bargain.

Competition may counter [the predicted consolidation of AI legal services](#) in the near future. Regulatory policies, in response, must aim to boost competition and shut down legal AI monopolies. Competition is especially essential to [push AI developers to improve their algorithms, make their services affordable, remove bias, and provide the most effective legal services](#). Here, competition functionally serves as another “check” on AI companies.

In most American jurisdictions, [lawyers can invest in technology, but technology companies can not invest in legal practices](#). This creates an asymmetric dynamic wherein wealthier firms have the capital to invest in technology, but smaller firms can not. Lifting these investment laws could help smaller firms attract the interest of digital AI service providers. Current law prevents cross-industry relationships between smaller law firms and technologists, thereby cutting off an avenue for smaller firms to adopt new AI. [Legal scholars, including Justice Gorsuch](#), have called for lifting ownership and investment restrictions. The best way to do this could be with a [regulatory sandbox](#) — an experimental area where certain restrictions are lifted but under close observation of an oversight body. Ryan Nabil finds that [regulatory sandboxes can significantly increase the accessibility of digital](#)

[justice tools](#). In 2020, Utah launched the first regulatory sandbox for legal services, and it was incredibly successful, [making civil legal services widely affordable](#). Expanding similar regulatory sandboxes to other states can simultaneously expand access to AI legal services for impoverished Americans and smaller law firms, helping them overcome previous financial barriers.

As legal technology gathers momentum, an approach of [“technology is better than nothing”](#) will not suffice. Artificial intelligence shows promise for equalizing access to justice, but it also presents perils of exacerbating inequalities. Regulators must act soon to contain negative spillovers from legal technology to ensure it can shrink the justice gap, not enlarge it.

---

## Related Articles

[Towards Principle-Based Policy for U.S. Child Priv...](#)

[Engineering a Fiduciary: Expanding the Regulatory ...](#)

[A Roadmap for Navigating the Uncharted Legal Desig...](#)

[AI Comes to the Board Room in a Black Box: Are the...](#)

---

## JOLT Sponsors

[Cooley LLP](#)

[Davis Polk & Wardwell LLP](#)

[Debevoise & Plimpton LLP](#)

[Finnegan, Henderson, Farabow,](#)

[Garrett & Dunner LLP](#)

[Fish & Richardson P.C.](#)

[Irell & Manella](#)

[Knobbe Martens Olson & Bear,](#)  
[LLP](#)

[Morrison & Foerster LLP](#)

[Sidley Austin LLP](#)

[Wilson Sonsini Goodrich & Rosati](#)

[Wolf, Greenfield & Sacks, P.C.](#)

[ZwillGen PLLC](#)

[Desmarais LLP](#)

## Contact Information

Harvard Law School  
Wasserstein Hall, Suite 3050  
1585 Massachusetts Ave  
Cambridge, MA 02138

[jolt@mail.law.harvard.edu](mailto:jolt@mail.law.harvard.edu)

Copyright © 2024 Harvard Journal of Law and Technology