Law Society of Ontario | Barreau de l'Ontario

TAB 1

AI for Litigators

AI in Legal Research: How AI is Providing Everyone Access to Information and Leveling the Playing Field for Firms of All Sizes

Jake Heller, CEO
Casetext

Laura Safdie, COO & General Counsel
Casetext

Pablo Arredondo, Co-Founder and Chief Product Officer
Casetext

April 16, 2021

cpd

Chapter 7

AI in Legal Research:

How AI is Providing Everyone Access to Information and Leveling the Playing Field for Firms of All Sizes

By Jake Heller, CEO, Casetext
Laura Safdie, COO & General Counsel, Casetext
Pablo Arredondo, Co-Founder and Chief Product Officer, Casetext

Legal research is critical to the practice of law, and artificial intelligence is increasingly critical to legal research. As co-founders of Casetext, a company pushing the boundaries of what artificial intelligence can do in law, we have had a front-row seat to the evolution of this technology and its impact on legal practice. We've watched as our scientists and engineers, as well as our competitors, evolved this technology to where it is today, and we see what's developing "in the lab" that will debut commercially in months or years. In this chapter, we'll share what we've seen so far from this technology and what we expect in the coming years.

But first, let's establish some basics, starting with: what is legal research and why care about it? Fundamentally, legal research is an exercise in determining *what the law is*—information that is then used to advise a client, craft contract language, or persuade a court that it should rule in favor of one party rather than another. Research can make or break a multi-billion dollar legal dispute, determine whether someone accused of a crime goes free or to prison, or jeopardize a critical business relationship. Because of its

centrality to the practice, lawyers spend nearly one out of every five hours at work researching.[1]

In many countries, including the United States, the task of research is made especially difficult because of the common law system. Under the common law, courts decide disputes on a case-by-case basis, issuing written opinions explaining their rationale. These written opinions become precedent that binds courts going forward. Determining what the law is is therefore more complicated than just "looking it up" in a set of rules or guidelines—it requires locating and understanding a rich tapestry of precedents, piecing together based on how courts ruled in the past how they will likely decide a particular issue in the future. The problem gets necessarily more complex as time goes on and more precedents are created—both because there are then many more precedents to locate and understand, but also because there are many more precedents that are *not* relevant to a particular situation that may get in the way of finding the answer. There are now well over ten million court decisions available in most legal research databases comprising hundreds of millions of pages of text. The task of researching within this quantity of information is daunting. Faced with this difficult and growing challenge, legal research has become for many attorneys not only a big drain on their time, but something painful they'd rather avoid. And because creating and curating these databases, including critical secondary information, is difficult and costly, legal research tools have traditionally been expensive—often prohibitively expensive to attorneys who represent less monied people and businesses.

[1] American Bar Association, Legal Technology Survey 2019—Online Research (2019) (explaining that lawyers spend on average 17% of their time researching); Steve Lastres, LLRX Report: Rebooting Legal Research in a Digital Age (2013) (finding that younger attorneys spend more than 30% of their time conducting legal research)

Enter artificial intelligence. Artificial intelligence, broadly speaking, is the ability for machines to mimic aspects of human intelligence. The aspect of intelligence particularly applicable to the law is *language*—understanding it (a field called natural language processing) and creating it (natural language generation). Outside of law, we have natural language processing and generation technologies to thank for Google finding the right website based on a simple search, automatic translation of a text from one language to another, and IBM's Watson besting Jeopardy's champions. This field has been studied and worked on to some degree since at least the 1950s. But very recently, machine understanding and generation of text has made exponential leaps, described later on in this chapter, that opens new and exciting possibilities nobody was even dreaming of just a few years ago. These include advancements that have already made research faster, more precise, easier, and less expensive—and therefore accessible to more.

We will look at three ways in which artificial intelligence has influenced legal research: the core search functionality, the creation and curation of the database of information to be searched, and the generation of legal documents—the latter of which we believe will change what "legal research" looks like fundamentally. Finally, we'll explore how the jobs of lawyers and the business of law has already changed in response to these advances in technology.

Alt: Can AI Save Lives? The Stakes of Legal Research

Before we dive into the details of the technology and business of artificial intelligence in legal research, it's important to remember what's at stake.

Soon after releasing our first major artificial intelligence product, we got a note from Crawford, a Florida criminal defense attorney. It began: "Casetext literally saved my client's life." He wasn't using the term "literally" hyperbolically. Crawford explained that his client was on trial for a mandatory life sentence. Just as he was gearing up to trial, his partner left the firm, leaving him to try the case on his own against a team of two prosecutors who would be backed by even more attorneys, legal assistants, investigators, and interns. Crawford explained that without the advantage of artificial intelligence research technology, it would have been difficult or impossible to keep apace with the prosecution as they filed motions critical to the outcome of the trial—determining which witnesses may testify, what evidence is admissible, and which charges should be dismissed. Decisions would have been made in the case detrimental to his client—not because his client was in the wrong, but because the government had more resources. Using advanced technology helped Crawford level the playing field. After a four day trial, his client was found not guilty.

Still today, years later, we think about Crawford a lot because it's a reminder of what's at stake. The ability for an attorney to research better and faster is not an abstract concept with theoretical benefits. It directly implicates people's livelihood, business, freedom, and, sometimes, lives.

The Dark Times (Before Artificial Intelligence)

Legal research as it exists today began in earnest when legal texts were originally digitized and made into searchable databases, efforts that began nearly fifty

years ago. In those early days, the ability to search these databases was crude at best. At first you would need to know the precise citation for a case or statute to retrieve it—and entering a citation and pulling legal texts was the only form of “legal research” available.

Later, you could search by simple keywords. For example, in a case about whether a car manufacturer should be held liable for a defective air conditioning unit, you could search for documents that contained the words “products liability” and “car” and “air conditioning.” But these methods were far too crude. Simple keyword search would bring back thousands of results that include all those words that may have nothing to do with what you’re looking for. You might also miss cases where the court, in its decision, used language different from what you searched for (“automobile” or even “truck” or “semi” instead of “car” may still be relevant).

Research systems then introduced increasingly complex but still insufficient methods to filter out the junk. The creation of “terms and connectors” or “Boolean” searches allowed attorneys to do more complex searches. For example, an attorney could enter (car OR automobile) w/100 “products liability” w/p “air conditioning” — requiring the words car or automobile to be within 100 words of products liability, which must be in the same paragraph as air conditioning. This approach came with serious costs. It was difficult to wrangle creating these queries, and even well trained attorneys would inadvertently design queries that brought in irrelevant material and, worse, unknowingly exclude material that was relevant. And there was still the problem of picking a word for the search that may be different than the words that appear in the precedent most relevant to your research issue.

Things evolved from there, but still in crude ways. Search engines started to enable hand-crafted thesauruses, so “FMLA” would be understood to also search for the “Family Medical Leave Act.” And algorithms got better at surfacing more relevant cases, preferencing more recent cases, or cases decided by courts at the appellate level, or cases where the words searched for appear more times more closely together. But the fundamental problems persisted, and, to this day, those using search engines like these often miss critical information while spending most of their time reviewing irrelevant results. No wonder research takes so long!

Search wasn’t the only problem. To create a functional database for legal research, it is necessary to include much more than just the text of cases, statutes, and regulations. Critical information like “was this case subsequently overturned?” is something that must be added to the database. It is now near malpractice to research using legal databases without these “citators” that show the links and relationships between cases. Moreover, a lot of helpful information like a summary of the case or the legal issues (often called “headnotes”) the case covers have become standards in legal research. These too must be added to the database.

Without technology to help, these databases of information were created and curated entirely manually—which is to say, thousands upon thousands of people worked on it. Although it may be apocryphal, it was once said that Thomson Reuters’s Westlaw and Reed Elsevier’s LexisNexis were the largest employers of attorneys in the country—attorneys who would spend all their time adding this information into their respective databases manually.

It’s no secret that attorneys are expensive. This practice drove up the costs associated with running a legal research system like Westlaw or LexisNexis, and these costs were passed along to attorneys who have struggled with the high cost of legal research for decades. To this day, there are still thousands of attorneys, most of whom represent less-monied people and organizations, who must go to the public law library to conduct legal research because they simply cannot afford the high costs associated with legal research.

So the time before artificial intelligence made its mark on legal research *sucked*—and today, for the attorneys that use legal research databases that have not fully utilized artificial intelligence, legal research *still sucks*. You pay too much for technology that all too often wastes your time by surfacing irrelevant information while excluding what you’re really searching for.

AI in Legal Research Today

Artificial intelligence has made the situation dramatically better. The core problem with older search technologies can be broken down into two deficiencies: *context* and *concepts*. First, the search engines did not have context about what the attorney was working on, a critical piece in retrieving relevant results and omitting irrelevant noise. Second, legal search engines before artificial intelligence searched by keywords instead of *concepts*, which is why it was so common to miss important cases—a court decision may say the same thing you’re searching for, but the judge said it using different language. Advancements in contextual, and conceptual, search are made possible because of exciting breakthroughs in natural language processing technology.

The first of these inputs, contextual search, takes a bit of explanation. When doing legal research, what an attorney is working on—their context—*really matters*. Finding precedents relevant to your case means, in practice, finding times when a court has evaluated a situation like yours. Understanding the particularities of your case, and reflecting those details in your legal research, means that the information most relevant to your unique situation can be surfaced to the top. The context of a case is almost necessarily more intricate than what you can fit in a few keywords. That context includes at least the sequence of events that led up to the legal dispute, the parties involved and their organizational or personal particularities, the jurisdiction, and the legal charges levied. Usually, this information spans pages, not keywords, and is contained in documents like a complaint that is filed at the initial stages of the litigation or the various legal briefs that come later.

Until recently, the only way attorneys could communicate their context to a legal research engine was by constructing lengthy keyword queries that invariably led to results that were at once over-inclusive (bringing in irrelevant decisions that happen to include the term) and under-inclusive (missing relevant decisions that happened to use other words to describe the relevant concept). Casetext pioneered a breakthrough in legal research by turning entire legal documents like briefs and pleadings into a form of mega-query. Casetext's CARA, the first brief-as-query tool brought to market, enabled attorneys to simply drag-and-drop a brief or complaint and effortlessly discover case law relevant to their context.

One of our favorite briefs to demonstrate CARA with was a summary judgment motion filed in a widely followed litigation concerning the employment status of Uber

drivers. A group of Uber drivers had filed a class action lawsuit alleging that they were improperly categorized as independent contractors, and were entitled to the benefits that come with employee status. Uploading this motion into CARA instantly returned a decision from the same court where a judge had denied summary judgment to Lyft on exactly the same claims! Other results centered on other cab drivers, bus drivers and even FedEx drivers bringing similar claims. Finding all of these cases using traditional tools would have taken a substantial amount of time and one study showed that attorneys missed some of these decisions entirely.

Because adding context to the search experience was so powerful, a lot of legal technology companies have pursued this approach—including Judicata's Clerk, Ross Inteligence's Eva, vLex's Vincent, Westlaw QuickCheck, Lexis Brief Analyzer, and Bloomberg Brief Analyze.

Contextual research is just one of the major advancements in legal search technology. The second, *conceptual searching*, is equally important. Conceptual searching means that the search program "understands" what you are looking for and finds relevant material, even if the language you use to search is dramatically different from the language in the search result. This means that an attorney can search using their own language without fear that they will miss out on a relevant precedent that happens to articulate the same concept differently. For example, a search about "not earning a diploma" might return results regarding "failing to graduate" or "not completing academic training." It also means that the system is less likely to bring back irrelevant results because it won't be fooled by instances where the same words were used but to an entirely different meaning. For example, a query for "patent AND DNA" might return

an opinion where a judge wrote “It is a patent falsehood that prison officials had a warrant to collect a sample of Mr. Smith’s DNA.” Finally, this form of searching is much easier for the attorney. Instead of using “Boolean” searching logic, the attorney need only write a sentence in their own phrasing (for example, a sentence for which they would like legal support in a brief), and the system will understand the meaning behind the sentence and use that understanding to find relevant material.

How do machines “understand” sentences? The key breakthrough was Google’s 2018 release of a natural language processing technique called BERT (short for “Bidirectional Encoder Representations from Transformers”). Anything resembling a deep dive into BERT is beyond the scope of this chapter, but essentially it overcame some of the limitations that stunted the development of language models (as opposed to other areas of AI like image-recognition). For example, the BERT approach allowed models to be initially trained on enormous volumes of data without the need for resource-intensive human-labelling. The human labelling is now only required for specific fine tuning tasks that require much less data to work.

Legal texts have a number of domain-specific idiosyncrasies including vocabulary, semantic meaning, and sentence formatting. Considerable work must be done to get these language models to work in a legal context. Even more work, and a few large technical breakthroughs, must be done to make this technology accessible as part of a search application usable by end users instead of in the labs. Given how new this technology is and the difficulty involved in making it work well in a legal context, only a handful of legal technology companies have brought this technology to market as of this writing. Casetext is to our knowledge the only company that has released a full-

fledged legal research engine based on the BERT approach (released as a tool called "Parallel Search"). The early response has been overwhelmingly positive, which is unsurprising given how often attorneys need to find conceptual matches to a query. LexisNexis's latest platform upgrade, Lexis+, has a question and answer feature called Lexis Answers that is advertised as being driven by BERT technology.

Contextual search and conceptual search are not mutually exclusive technologies; rather, they are most powerful when combined. Search technologies that take into account the context of the specific litigation an attorney is working on while also searching for concepts rather than keywords represent the best-in-class artificial intelligence search technologies.

Besides these technologies, which are search applications, artificial intelligence has also reduced the expense associated with creating and curating legal research databases—and ultimately has helped make legal research more affordable and accessible to more attorneys. As described above, legal research databases contain far more than just the law itself, but also information that makes legal research efficient and easy. For years, that information was curated manually by thousands of attorneys. Today, artificial intelligence technologies are doing more and more of the heavy lifting, reducing the labor requirements and associated expenses while enabling the attorneys to review and work on the most difficult and pressing tasks beyond the capabilities of artificial intelligence technologies.

Take for example the warnings legal research databases contain that indicate whether a precedent has been subsequently overturned—a feature in LexisNexis called "Shepard's," in Westlaw called "KeyCite," and in Casetext called "SmartCite." There are

hundreds of millions of instances of cases citing previous cases, and reviewing each of these relationships manually would take centuries. For example, LexisNexis's Shepard's feature was born out of a manual effort, where each case relationship was meticulously recorded by hand and published in many volumes of books, that began in the 1800s.

Natural language processing technologies have made this process considerably more achievable on a shorter timescale and with fewer people. These technologies can find language that likely indicates that one case is overruled another, and flag that for human review, confidently and at a high level of precision saying that the remaining relationships do not represent one case overturning another. One thing we found fascinating was the "features" (words/phrases) our NLP systems found predictive of a case being overruled. For example, the word "today" was suggestive as judges, perhaps in a bit of fanfare, would often announce that they "today" render an earlier holding obsolete. There were many more examples where the machine taught the lawyers about patterns we never would have guessed.

By taking centuries of rote work out of the process of creating and curating the legal research database, legal research becomes more affordable for two reasons: First, as the costs for creating legal research databases goes down, those cost savings are passed on to the attorneys that subscribe to these services. Second, creating a legal research database has become substantially more achievable by newer startup companies that are creating a truly competitive environment with the older, legacy providers that have enjoyed near monopoly status for decades. Casetext, for example,

is approximately a third the cost of the larger competitors, one key reason over 6,500 law firms have joined the platform over the last two years.

Thanks to artificial intelligence, today attorneys can find better information faster while paying less. Even more exciting is what's on the horizon.

AI will help to Write the Future

To us, the most exciting application of legal research will be coming in the next few years. These applications will be different in one substantial way: rather than just helping attorneys find information to add to a brief or other legal document, they will be helping attorneys write those documents.

Most legal research is done with the intention of deploying that information researched in a handful of ways: most commonly, to advise a client, strategize regarding whether a lawsuit can be pursued, or write a legal brief to persuade a court that your position is backed by precedent. In each of these cases, there is usually associated written work product, like a memorandum or brief. We anticipate that the most exciting applications of artificial intelligence will skip the step of research and go directly to aiding an attorney write work product that automatically and correctly identifies the right precedent for the attorney.

Today, the field of natural language generation is just starting to show real signs of promise. For example, the technology company OpenAI released a new algorithm, GPT3, that writes convincing-sounding language when given a prompt—including legal language. The technology is eerily good at writing language that appears on the surface

indistinguishable from something written by a human.[2] But technologies like GPT3 are still in the early stages of development; in the legal context, algorithms like this may *sound* like a lawyer wrote it but get the substance completely wrong.

It is unlikely that machines will ever truly do the writing for lawyers—there is too much knowledge, strategy, and persuasion built into legal writing. But they can assist substantially. In the not-to-distant future, for example, a lawyer may write out the factual circumstances of the case and a machine process will suggest the legal arguments available to the attorney, the cases and other legal authorities they can cite, and some starter language to begin writing out the legal argument. Another example would be a much more advanced "autocomplete" feature for legal writing, where an attorney begins a thought and the computer suggests what is most likely going to be said next along with corresponding legal authorities to back it up.

In these sorts of examples, what is known today as "legal research"—searching through a database of legal authorities to find relevant precedents and other materials—is turned on its head. Instead of the focus being on searching, the focus becomes on the final written work product. And the attorney does barely any searching at all in this new world. Rather, the attorney focuses on the heart of the craft—framing the facts of the case, choosing the arguments they will pursue, and writing persuasively—and the system provides the research, or at least suggestions that the attorney can validate and add if appropriate.

A future without legal research as it is currently known may seem far off to many attorneys, but the beginnings of this work is happening today. For example, Compose, a

[2] If you want to see examples, a website has compiled them: https://gpt3examples.com/

new product we have been working on, already suggests all common legal arguments and legal authorities for a growing number of legal issues that an attorney can add to the language of their brief with one click. And when Compose has not covered a legal issue, the contextual and conceptual search technologies described above enable an attorney to find legal support for any sentence they write. These technologies represent the first step towards a dramatically more efficient (and enjoyable) research and writing process for attorneys.

AI Will Continue to Help Lawyers Do Higher Value Work

With the introduction of new technologies to the legal profession, especially those that automate or make easier some aspects of legal practice, there is always some amount of understandable consternation—will these technologies take away the attorney's autonomy, the more fulfilling aspects of practice, or job?

To the contrary, these advances in technology are already changing for the better what it means to be a lawyer and law firm. As artificial intelligence becomes better at locating the best precedents and avoiding irrelevant material, lawyers are shifting their time to higher-value and more rewarding tasks, like strategizing, writing persuasively, and investigating the facts of the case. Because the work is much more efficient, law firms are finding friendlier ways to bill for their services, including charging a flat fee rather than charging by the hour. As where in the past there was a real advantage to having an army of associates research a topic—which benefited larger

firms and the well-financed clients who could afford them—modern technologies enable a single attorney to quickly find the most relevant precedents and produce high-quality work product, in effect leveling the playing field. And as legal work becomes increasingly efficient and predictable, the attorneys can represent clients who are not well-resourced, and do so at the highest level of practice.

In short, we see a future of legal practice that is more efficient, just, and fulfilling—which is why we do what we do.