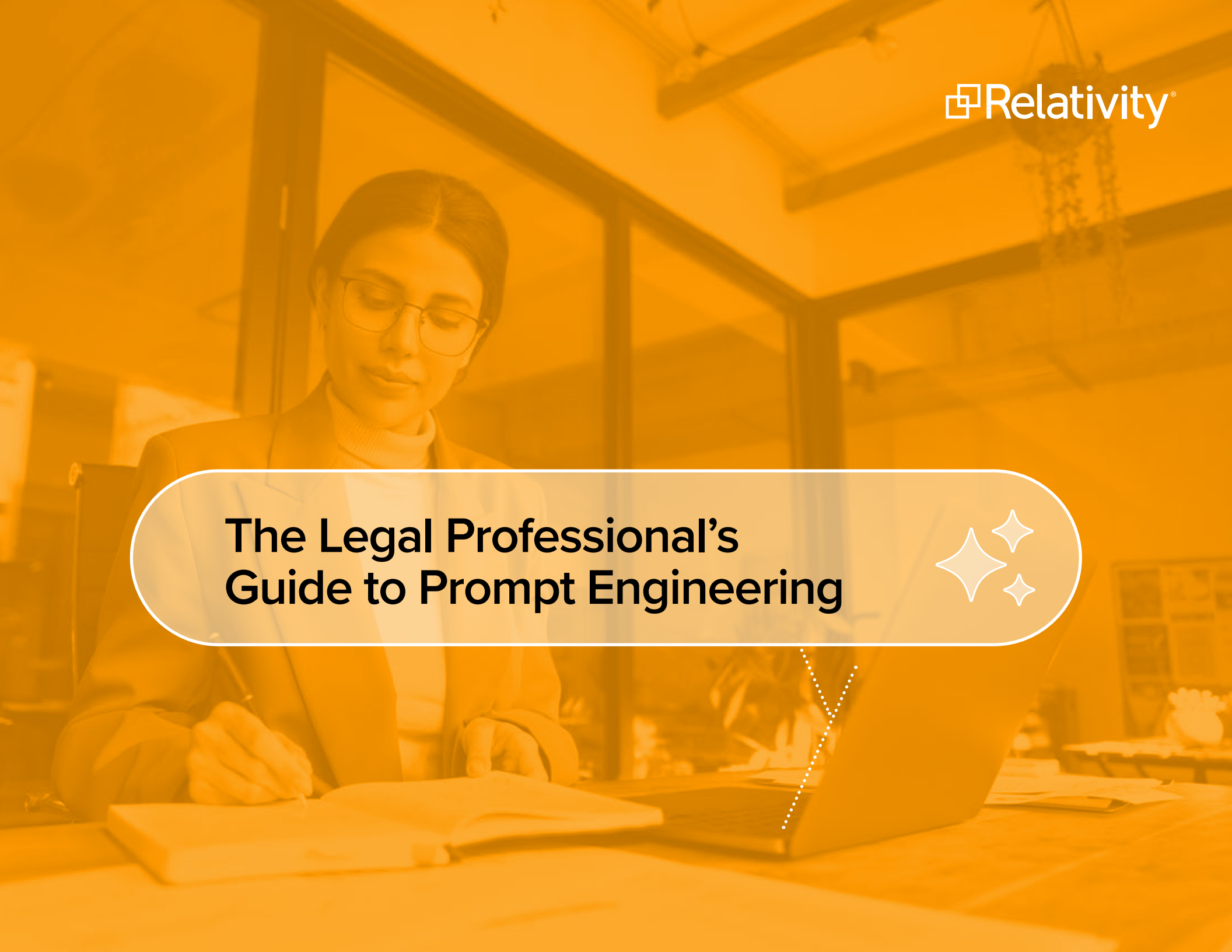
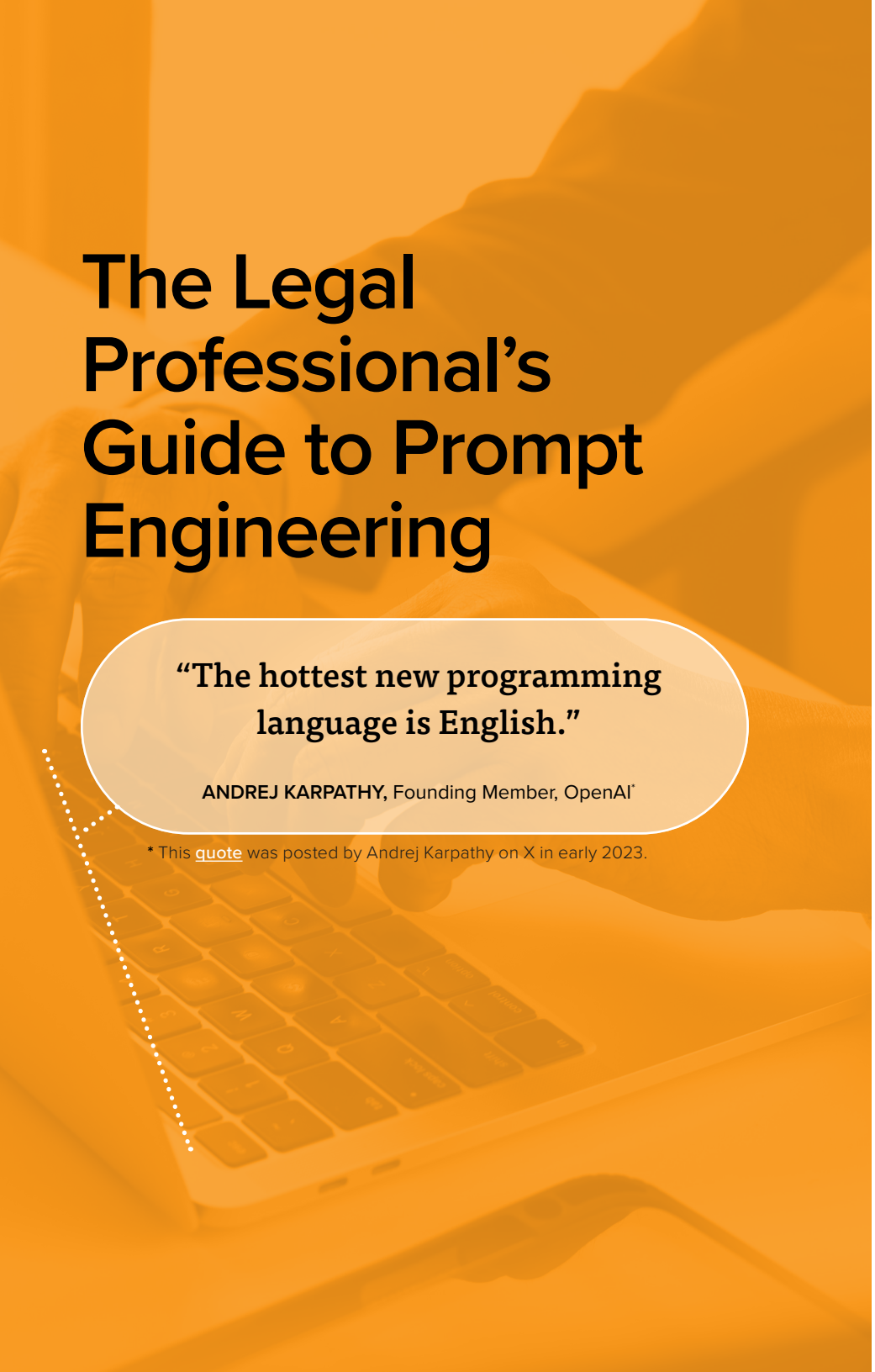


Relativity

The Legal Professional's Guide to Prompt Engineering

The Legal Professional's Guide to Prompt Engineering



> **"The hottest new programming language is English."**
>
> **ANDREJ KARPATHY,** Founding Member, OpenAI*

* This quote was posted by Andrej Karpathy on X in early 2023.

Table of Contents

Prompting vs Prompt Engineering 4

Cut Out the Noise, Pay Attention to Ground Truths 6

Field Notes: Some Best Practices for Those Just Getting Started 7

Finding Your Style as Your Prompting Skills Mature 9

Should Prompts Be Protected by Work-Product Doctrine? 10

What Happens When You Let AI Draft Prompts for You 11

The Time to Lean in Is Now 13

Around 14 months before OpenAI launched its debut model of ChatGPT[1], an AI startup hired its first prompt engineer. Anna Bernstein joined Copy.ai in September 2021 when the world was just coming out of COVID lockdowns, and the term large language model (LLMs) was in extremely rare use—except by AI researchers working in quiet obscurity.

Bernstein is *probably*[2] the world's first prompt engineer. In a *Time* article titled "How to Get a Six-Figure Job as an AI Prompt Engineer," Bernstein—an English graduate and former copywriter—stated that she had no background in computer science. In fact, she saw her humanities background as an advantage in doing her own job. "It's a really strange intersection of my literary background and analytical thinking," she told *Business Insider* in an interview.

Like Bernstein, Albert Phelps was hired by Accenture as a prompt engineer in 2021—long before the emergence of generative AI grabbed international headlines and opened the floodgates to a tidal wave of investment capital[3]. And like Bernstein, Phelps did not see programming skills as a barrier to entry. "Being clear and economical in the way that you write is important. English, history, and philosophy people rejoice because you can interact with these very advanced models just with words, which is awesome," Phelps told the *World Economic Forum* in an interview.

Fast forward about 2.5 years since the launch of ChatGPT. "Prompt engineering" is no longer just a trending neologism but an integral part of the lexicon, particularly in the legal industry. Today, more than half of attorneys at American law firms use generative AI in their work, according to a survey by *Law360*. The trend is mirrored globally: 66 percent of lawyers are already using generative AI solutions in Singapore and Malaysia; 41 percent of lawyers in the UK report using generative AI at work; and 50 percent of legal professionals in Australia and New Zealand have used generative AI solutions to perform day-to-day tasks.

For good reason.

Respondents to a *Thomson Reuters* survey of professionals spanning the legal, tax, and risk and compliance fields predicted that AI could save them 12 hours of rote work per week by 2029. And when it comes to document review—one of the most expensive and time-consuming tasks in which a legal team can partake—adopters of generative AI are getting it done as much as 80 percent faster.

Indeed, lawyers are prompting AI models to assist them on a range of legal tasks: from drafting legal briefs and reviewing contracts to summarizing documents and conducting M&A due diligence.

Like Bernstein and Phelps, the vast majority (over 93 percent) of law graduates do not have STEM degrees, according to a *Reuters* study in 2023.

If anything, lawyers have many transferable skills—such as linguistic precision, analytical reasoning, and iterative drafting—that give them an edge in prompting AI models.

1 By "debut model", we mean the GPT 3.5 model that is called simply "GPT" in common parlance. To be clear, OpenAI launched GPT-3 in private beta access.

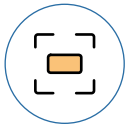
2 In this instance, "probably" is the operative word. Based on our research, Bernstein was one of the first "prompt engineers" to be interviewed by publications such as *Time*.

3 *TechCrunch* reports that generative AI companies raised $56 billion in 2024 alone.

Indeed, the parallels are significant:

Mastery of Precise Language

Lawyers are trained to use language with exceptional care, drafting contracts, pleadings, and legal opinions that leave little room for ambiguity. This same precision is useful in drafting prompts, where the phrasing of a query directly influences the quality and reliability of an AI-generated response. Indeed, the same skill that makes lawyers good at writing persuasively can be applied to draft prompts in which words are used in a measured and unambiguous way to elicit the right output from an AI model.



Strength in Framing Questions

At the heart of legal practice lies the ability to frame issues clearly for a judge, arbitrator, or opposing counsel. Similarly, prompting involves framing a task or inquiry in a way that guides the AI toward a relevant and defensible output. The clarity of a prompt, like the clarity of a legal brief, often determines its success. Attaining this clarity requires iterating and experimenting with different ways of prompting that over time gives legal professionals an experiential understanding of how AI models work, and how they react to prompts.

Understanding of Nuance & Context

Legal professionals excel at interpreting facts in the context of precedent, policy, and circumstance. Prompt engineering likewise requires an appreciation for context—what the AI model needs to "know" to return a relevant and accurate result. Providing the right background, constraints, and examples is useful in both domains.

To close the loop on Anna Bernstein's and Albert Phelps' stories, here's where they are in their career journey:

Anna Bernstein is now the head of prompt engineering at Copy.ai. Albert Phelps has since left Accenture to pursue his entrepreneurial dreams. He went on to cofound Tomoro, a European AI startup that recently raised €4 million in funding.

Prompting vs Prompt Engineering

Generative AI, built upon large language models, takes natural-language questions and instructions from its users, then creates and serves responses to them. The large language model is what enables the AI model to understand the prompt and "think" using human linguistics; the "generative" means that its response to that linguistic input is to create and serve some sort of unique data output.

OpenAI defines prompt engineering as the process of crafting prompts to extract the right output from a model. But given the fact that prompt engineering is a relatively recent coinage that came into usage at a pace that can only be described as viral, the terms calls for a clearer articulation of what it means and entails. Another reason why the term still seems nebulous is because the terms "prompting" and "prompt engineering" are often used interchangeably.

But they are, in fact, different:

Prompting is the act of giving a command, question, or input to a large language model (LLM) like ChatGPT or Claude to get a desired output.

Prompt engineering is a more deliberate, structured, and optimized process of crafting prompts to achieve precise, repeatable, and high-quality outputs from an AI model.

Indeed, prompt engineering suggests a much more iterative, trial-and-error-based approach to generating the desired outputs from an AI model. Practitioners rely on a layered set of linguistic techniques and best practices to refine results. It's worth noting that, despite the use of the term "engineering," these methods are not technical in the traditional sense. They are rooted in language and communication, not code.

Here are some of them:

Zero-Shot Prompting

In zero-shot mode, you give the model only an instruction—no demonstrations, no prior examples.

Example: "In one paragraph, distill the ratio decidendi of R. v. Jordan (2016 SCC 27)."

Chain-of-Thought (CoT) Prompting

Chain-of-thought prompting explicitly asks the model to surface its intermediate reasoning steps ("think step-by-step").

Example: "Analyze whether the following non-compete clause is enforceable under New York law. Delineate your reasoning in IRAC format before giving a conclusion."

Role-Based Prompting

Assigning the model a professional identity steers tone, vocabulary, and evaluative stance. By telling the model *who it is* ("You are a senior appellate court clerk"), you implicitly load an entire frame—genre conventions, argumentative style, domain assumptions—before the substantive question appears.

Example: "You are a veteran law-review articles editor. Suggest three titles that capture the policy implications of the EU AI Act for a general counsel of a multinational company."

Contextual Prompting

Contextual prompting embeds the substantive materials the model must rely on—contracts, statutes, pleadings, fact patterns—inside the prompt itself. This type of prompting embeds the context into the prompt and keeps the AI model from hallucinating.

Example: Drop the contract into the prompt and write: "Highlight any provisions that could trigger a change-of-control penalty under Delaware law."

System Prompting

A system prompt is the conversation's primordial directive—usually hidden from the end user—setting global boundaries.

Example: "You are a neutral legal assistant; cite primary sources; refuse political speculation."

Cut Out the Noise, Pay Attention to Ground Truths

In 2023, analysts from Goldman Sachs published a report titled "The Potentially Large Effects of Artificial Intelligence on Economic Growth." The report, which was released in the early months of the generative AI frenzy, caught many off guard as it placed high odds on the likelihood of large-scale automation across broad swathes of knowledge work. It particularly exercised the legal industry with its bold prediction that 44 percent of work done by legal professionals would be automated.

The report was famously featured by Law.com in their March 2023 cover story, becoming a fraught reference point at panel discussions, workshops, and CLE programs across the legal community.

Looking back now, many of the core predictions in the study have since been challenged by leading economists such as Nobel laureate Dr Daron Acemoglu. An Institute Professor at MIT, Dr. Acemoglu predicts that no more than 5 percent of the job market across all industries and professions would either be taken over or be heavily aided by AI in the next 10 years. In "The Simple Macroeconomics of AI," his widely acclaimed paper, Acemoglu made a strong case against AI's potential for causing job displacement in white collar professions such as the law, citing AI's incapacity for "human judgment."

See how the landscape is constantly shifting under the quick stream of new research, opinions, and technological breakthroughs that have characterized the last two years?

By the same token, prompt engineering—the focus of our e-book—needs to be understood in the backdrop of a rapidly evolving technology and the constant churn of research findings that question our assumptions of how LLMs work, what they are capable of doing, and how to best work with them.

In its 2024 AI Jobs Barometer study, PwC found that US-based lawyers with AI skills like prompt engineering commanded a 49 percent wage premium over other lawyers who did not possess similar skills.

Indeed, law firms, in-house legal teams, and legal service providers are now instituting new learning programs and AI transformation committees to upskill employees in skills such as prompt engineering just as Vanderbilt University is now offering a dedicated course on prompt engineering for legal professionals.

In the midst of these trends, two machine learning engineers at VMware made a discovery with potentially profound implications on the future of prompting and its attendant best practices. In their study titled "The Unreasonable Effectiveness of Eccentric Automatic Prompts," the paper's coauthors found prompts optimized by AI had a higher success rate than human-generated prompts created through a longer and more laborious process of trial-and-error. The prompts generated by AI seemed "eccentric". For instance, one of the auto-created prompts was an extended Star Trek reference guiding the LLM to solve a grade school math problem. The paper's co-authors summed it best: Algorithmically optimized prompts fare better because ultimately LLMs too are algorithms.

Regardless of these developments, industry leaders maintain that the ability to effectively guide AI models—to ask the right questions and extract meaningful output—will remain a valuable and enduring skill.

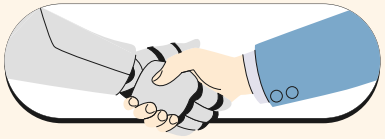
Instead of gazing into the crystal ball, we looked *deeply* into what is happening *today*.

In developing this e-book, we got down to first principles. We engaged directly with the people shaping and scrutinizing this transformation: technologists, AI researchers, legal data intelligence (LDI) practitioners, law firm partners, general counsel, data scientists, and lawyers across domains.

Our inquiry was grounded in real-world use. We gathered perspectives from those actively creating and iterating prompts to query AI models for a range of legal tasks. We also spoke to legal observers monitoring how courts, regulators, and professional bodies are responding to the use of generative AI in practice, and the impact of agentic AI on their legal workflows.

We interviewed seasoned lawyers—many without technical backgrounds—who described their journey toward becoming AI-fluent. We demystify the core concepts of prompt engineering, offer a practical overview of the technology architecture supporting generative AI, and examine case studies where these tools have delivered value.

What follows is not speculative. It is a grounded, multidimensional portrait of where the legal profession stands today in its evolving relationship with prompt engineering and AI.



Field Notes: Some Best Practices for Those Just Getting Started

In the context of generative AI for legal applications, there are countless tools out there to explore. For our purposes, we'll dig into Relativity aiR for Review, which is used for a range of purposes such as relevance review, issues review (locating material related to legal issues important for case strategy), finding hot documents across a voluminous data set, and all range of Legal Data Intelligence use cases.

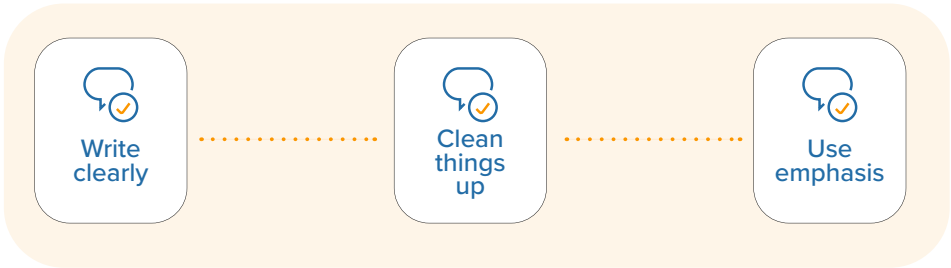
According to Jeff Gilles, a senior generative AI solutions engineer at Relativity, aiR users should know that,"90 percent of writing clear prompt criteria is simply writing clearly."

"90 percent of writing clear prompt criteria is simply writing clearly."

Prompt criteria are a set of inputs that give aiR for Review the context it needs to understand the matter and evaluate each document. In other words, writing the prompt criteria is a way of training your "reviewer," similar to training a human reviewer.

The system limits prompts to 15,000 characters (although Jeff suggests aiming for significantly fewer) and Jeff, along with other Relativity experts, encourages users to keep "the human paradigm" in mind. In other words, every time you write a prompt, ask yourself: would a human reader find this confusing?

During a crash-course session on prompting best practices, Relativity experts encouraged users to follow these guidelines:



1. **Write clearly.** This means being concise, using the active voice, avoiding double negatives, and avoiding "lawyer speak." Plain language is the key.
2. **Clean things up.** Checking for good grammar—and, though less critical, correct spelling—can help aid in clearly articulating criteria for the AI. So can intentional formatting.
3. **Use emphasis.** You won't hurt the AI's feelings by weighing important notes with all-capital letters or exclamation points. In fact, these tactics can encourage the AI to focus on particular considerations (and if you're familiar with Boolean operators, they can be helpful in this context, too).

Check out aiR for Review's best practices for a longer list of tips on writing effective prompt criteria.

With these tips in mind, users should aim to translate their formal request for production into a simple, straightforward review protocol for the AI to use as the basis for its analysis.

Armando Nardo, a senior managing director at Teneo and a 2025 AI Visionary, told us that "a review protocol includes detailed background, definitions, and guidelines, while AI prompts need to quickly communicate the key points. A well-written prompt makes sure the AI understands important factors like privilege, responsiveness, or issue coding without unnecessary details."

He noted, for instance, that "small changes, like removing vague or overly broad legal terms" can have a big impact. Specifically, "words such as 'ideally,' 'reasonable,' 'substantial,' or 'material' can be interpreted too loosely by AI."

During the Relativity Fest Sydney session earlier this year, Jill Ragan, a Legal Data Intelligence education architect at Relativity, shared an example.

Note how the Request for Production (RFP) rings of typical legal language, whereas the resulting prompt is more natural language—drawn out to illustrate specifics and emphasize key points.

Example RFP:

Any and all internal written correspondence between January 1, 2015 and the present date that relates to the design of snow shovels, including but not limited to:

a. Tensile strength

b. Handle design

Prompt criteria:

A document which relates to the design of snow shovels is relevant. A document describing tensile strength or handle design in snow shovels is very relevant. A document which mentions snow shovels but doesn't discuss or relate to their design MUST be marked non-responsive and irrelevant.

In this case, prompt language is simply a few sentences about overall themes of responsiveness in the document population. But this isn't the only structure one might use.

Prompts may also look like a list of relevant materials and topics, as in this example:

Documents are responsive if they relate to:

- the RFP from the City of Atlantis
- BigThorium's process for bidding on contracts
- discussions between BigThorium employees about giving gifts as part of the bid process
- discussions about bribing government officials
- discussions about BigThorium winning the contract from the City of Atlantis
- discussions about the purchase or delivery of luxury goods by/to BigThorium employees
- communications between BigThorium employees and officials from the City of Atlantis that show a personal relationship

Again, notice how concise and simple these guidelines are—shorn of extraneous language and verbosity, so that the AI can drill into what matters with clear and specific instructions.

Once the AI is unleashed after receiving the prompt, it's instructive to review its results and evaluate how well it's performing. One could fine-tune the specificity of their instructions, add exclusions, or pivot as new facts are discovered in the documents returned.

"A key difference between manual review instructions and prompting AI is the iterative nature of AI prompt writing. Unlike traditional reviewer protocols, which are typically finalized in one phase, AI prompts require testing and refinement to optimize effectiveness," Starling Underwood, e-discovery of counsel at Kilpatrick Townsend & Stockton, explained in an interview for this e-book.

Case teams conduct much of the refinement by testing prompts on smaller subsets of data and evaluating aiR's results before running the tool on much larger document populations.

Making this iteration as effective as possible requires assigning true experts to that craft.

"Effective prompt creation requires a collaborative approach, with input from multiple stakeholders, but a single drafter best handles the actual drafting process," Starling advised. "This ensures consistency in style, clarity, and structure throughout the refinement process, which is critical for optimizing aiR's performance."

Finding Your Style as Your Prompting Skills Mature

Notice how each of these generative AI pros had a particular style in their approach to prompting aiR for Review? They involved linguistic and even emotional nuances that have paid dividends in their projects.

For example, Underwood pointed to the value of simple grammatical changes in tone to help drive the importance of certain prompt criteria home. The AI works best if spoken to in a direct and unselfconscious way.

> **"Shorter, simpler prompts help reduce ambiguity and improve accuracy, while strategic use of capitalization highlights critical elements for added clarity"**

“Shorter, simpler prompts help reduce ambiguity and improve accuracy, while strategic use of capitalization highlights critical elements for added clarity,” he said. “For instance, when identifying documents regarding a critical issue was essential, I refine relevance prompts to explicitly label documents with the issue as ‘VERY RELEVANT’ instead of merely ‘relevant.’ This small but impactful change, combined with capitalizing key terms, significantly improves aiR’s ability to prioritize documents.”

AI won’t think it’s being shouted at; it will merely understand the level of importance it’s lending to those components of the prompt and adapt accordingly.

Benjamin Sexton, senior vice president of innovation and strategy at JND Legal Administration and a 2025 AI Visionary, emphasized placing a certain amount of trust in the built-in expertise of the large language models driving aiR.

“When instructing an AI, consider that certain instructions you may provide to a human review team may actually be limiting for the LLM. For instance, while you absolutely should provide clear criteria for responsiveness and/or issues coding, the AI is already an expert in most types of public-domain knowledge, so providing legal concepts and definitions may actually handcuff the LLM’s ability to shine,” he explained. “For this reason, we’ve found that most prompts, while detailed, wind up shorter and more direct than what we see in eyes-on review protocols.”

Armando agreed: “Don’t feel the need to explain complex concepts—it should be able to work things out.”

There’s a common notion amongst legal professionals that one should think of generative AI as a junior associate: an educated team member with a good amount of textbook knowledge, but needing plenty of context and guidance to succeed in practice.

But Bennett Borden, founder and CEO of Clarion AI Partners and a 2024 AI Visionary, told us during a recent interview that he follows a different track, particularly when working with generative AI in its chatbot forms.

“Prompting is just asking really good questions. My advice to everyone is: look at what comes out of off-the-shelf generative AI models as if it were an opposing witness. Imagine you’re deposing a hostile witness,” he said. “Make it prove its answer. If you can think in that way, then you’ll get much better information.”

Should Prompts Be Protected by Work-Product Doctrine?

Another experiential insight we’re hearing from our community may add quite a bit of complexity to inter-party conversations around AI. Best to be prepared for that conversation, should you encounter it.

To that end, Cristin Traylor, senior director of AI transformation and law firm strategy at Relativity, noted in a popular session that the question of producibility regarding case teams’ prompting strategies is erupting into quite a hot topic.

“Requests for production are written in legalese, and therefore do not yield good results in aiR,” she explained to a standing-room-only crowd during the session, entitled “Negotiating ESI Protocols for the Use of Generative AI in e-Discovery.”

“They need to be ‘translated’ into clearer, more natural language for aiR, just as they do for human reviewers. An emerging question is whether this translation—these AI prompts—are covered by the work product doctrine, or if they may be appropriate to exchange between parties, as in the case of search terms,” Cristin continued.

The session featured a panel of in-house and law firm attorneys. They discussed the merits of the issue and played out a lively sketch showcasing how these debates may occur between opposing parties during litigation.

While most of the panelists seemed to agree that prompts should be protected by the work product doctrine and not exchanged during the discovery phase, not everyone did.

Overall, the entire group *did* agree that following prompting best practices and showing a willingness to collaborate and civilly discuss such questions are both essential strategies for proportionate, effective, and responsible discovery.

As of this writing, there is no settled case law on this issue, and opinion does appear to be divided across the industry. Seth M. Cohen, partner at Alston & Bird, in his piece on *Bloomberg Law* compared AI prompts to some analogous parts of current legal workflows.

For instance, just as drafts of legal briefs and lawyers' notes are typically protected under the work product doctrine, the prompts that lawyers input into generative AI tools to assist in brief writing are likely to be privileged and protected from disclosure.

On the other hand, AI prompts used by lawyers to determine the relevance of certain documents in an action may not be subject to such privilege, just as keyword search terms are not privileged and are subject to disclosure, argued Cohen in his Bloomberg Law piece.

Only time will tell how this will shake out.

What Happens When You Let AI Draft Prompts for You

In 2024, researchers Rick Battle and Teja Gollapudi at VMware tested 60 different prompt setups across several open-source LLMs. The results showed that "autotuned" prompts—crafted by models themselves—often surpassed the best human-designed prompts. Their paper, titled "The Unreasonable Effectiveness of Eccentric Automatic Prompts," kickstarted a contentious discussion on the relevance of prompt engineering itself. At the root of the discussion was the hypothesis that if AI models can self-optimize, the role of the prompt engineer may become obsolete or, at any rate, dramatically evolve.

One of the researchers, Battle, even went so far as to tell *Business Insider* that one should never have to labor over writing the perfect prompt ever again: "You should never handwrite a prompt again. Just write basic instructions, and then let the model optimize the prompt for you."

Correspondingly, Relativity recently launched a prompt kickstarter feature that autogenerates prompt criteria from existing case documents, such as requests for production, review protocols, complaints, or case memos. After you upload up to five documents, aiR for Review analyzes them to create a first draft of the prompt criteria.

"I found the new feature very useful for prompt creation. What would have taken thirty minutes to create can now be done in less than a minute. It's incredible!"

DANNY CHAN
Senior eDiscovery Case Manager & LDI Architect, Miller Thomson

Early reviews of the feature are in and the thesis seems to have proven itself out: indeed, prompts generated using the prompt kickstarter were as good as—and, in many cases, even better than—those created manually. There is now a growing cohort of early adopters who are turning to the prompt kickstarter capability to instantaneously generate prompts that would otherwise have taken hours of work.

The Time to Lean in Is Now

Jared Spataro, the CMO of AI at Work at Microsoft, told the *Wall Street Journal* that as large language models evolved to be more iterative, conversational, and aware of context, the role played by prompt engineers was becoming less important than it used to be. "You don't have to have the perfect prompt."

The changing sentiment has had a corresponding effect on jobseekers' expectations as well. User searches on Indeed for the role surged from two searches per million total searches in the U.S. in January 2023, months after ChatGPT's debut, to 144 per million in April 2023. By April 2025, the search volume had flatlined at about 20 to 30 searches per million.

But despite the alleged waning enthusiasm for prompt engineering roles and skills, industry leaders believe that the ability to prompt an AI model to get the right information will continue to be highly relevant in the future.

"The skills relevant in the future are actually going to be really different than what lawyers currently have, and it will take us time to get there as an industry. But what I'll be looking for in my legal team are lawyers who are comfortable querying some central database, getting relevant information out of it and then taking action on that information," said Rob Beard, chief legal and global affairs officer at Coherent.

Prompt engineering is not the endgame. It is a gateway—a practical, adaptable skill that opens the door to a deeper understanding of how legal professionals can engage with AI. As we've seen through the voices in this e-book, legal professionals best prepared for this new chapter aren't necessarily those with the most technical knowledge, but those who commit to continuous learning; those who remain intellectually curious and engage with the communities shaping this transformation.

The legal profession has always evolved through conversation—through precedent, peer review, and principled debate. That same tradition now extends into the AI era.

We hope this e-book serves as both a primer and an invitation: to think more critically, question more deeply, and explore new frontiers with curiosity and confidence. We hope it encourages you to join the early adopters, the tinkerers, and bold leaders who not only see the potential in AI to transform legal work but take action towards unlocking it. Below is a list of other resources to help you on your journey to discovering and learning more about the exciting developments at the intersection of AI and the law.

The Relativity Blog

(Primer) From Beginning to Breakthrough: Navigating Document Review's AI Evolution

(Webinar) AI Advantage: Aiming for Prompt Perfection? Level up with Relativity aiR for Review

(Webinar) Proof Points: How to Sell Generative AI to Internal and External Stakeholders

(Webinar) Navigating the Ethical Maze: Generative AI in eDiscovery

IDC Research Study: Generative AI in Legal Study

231 South LaSalle Street | 20th Floor | Chicago, Illinois 60604
+1 (312) 263-1177 | relativity.com