

Title: **Opportunities and Challenges in Legal AI**

Date: January 2025

Authors: Aksh Garg, Dr. Megan Ma

## 1 Introduction

With a global market size of [\\$1 Trillion](#), the legal industry has received significant attention from new startup founders as an area to build in. At face value, this interest makes sense. Law is largely grounded in text-based transactions, which appear ideal for automation and processing using large language models (LLMs). These LLMs offer the opportunity to rapidly accelerate the rate at which humans can understand and process large documents – similar to those that lawyers work with--offering avenues for different startups to interject and innovate at the cusp of the generative AI (gen AI) wave. Gen AI seems particularly useful for making legal work more efficient, since, despite being some of the highest-paid professionals, lawyers spend a disproportionate amount of time on repetitive tasks such as basic research, document review, and case preparation. This insight is affirmed by the amount of funding directed towards AI companies in 2023 and 2024. For example, in [The Generative AI Legal Landscape 2024](#), Ma et al. found that “funding for LegalTech startups totaled [~700M](#) between Jan 2023 - Feb 2024.”

This interest in legal AI is further bolstered by the rapid reduction in per token [LLM cost](#), which has roughly trended along a 10x annual reduction per year as we can see in the figure below, and expanding context window sizes, which have seen GPT, Claude, and Gemini windows rise to [128K](#), [200K](#), and [1M](#) token sizes, respectively. These changes are enabling lawyers to process hundred-page briefs, the typical length of legal documents, at a [fraction of the costs](#) of actual lawyers doing the work, resulting in new levels of savings.

The scale of the legal workload further underscores the urgency of innovation: [65 million cases processed annually](#) and [1.6 million patent filings](#) handled by 450,000 U.S. law firms. As the complexity of cases grows and client expectations rise, AI offers an unprecedented opportunity to reshape legal workflows by improving efficiency and reducing costs in most areas of the legal stack. This raises an important question: given the seemingly obvious synergies between law and LLMs, why have there not been many breakthroughs in the legal AI space, with most companies still relying on antiquated workflows to achieve their daily tasks?

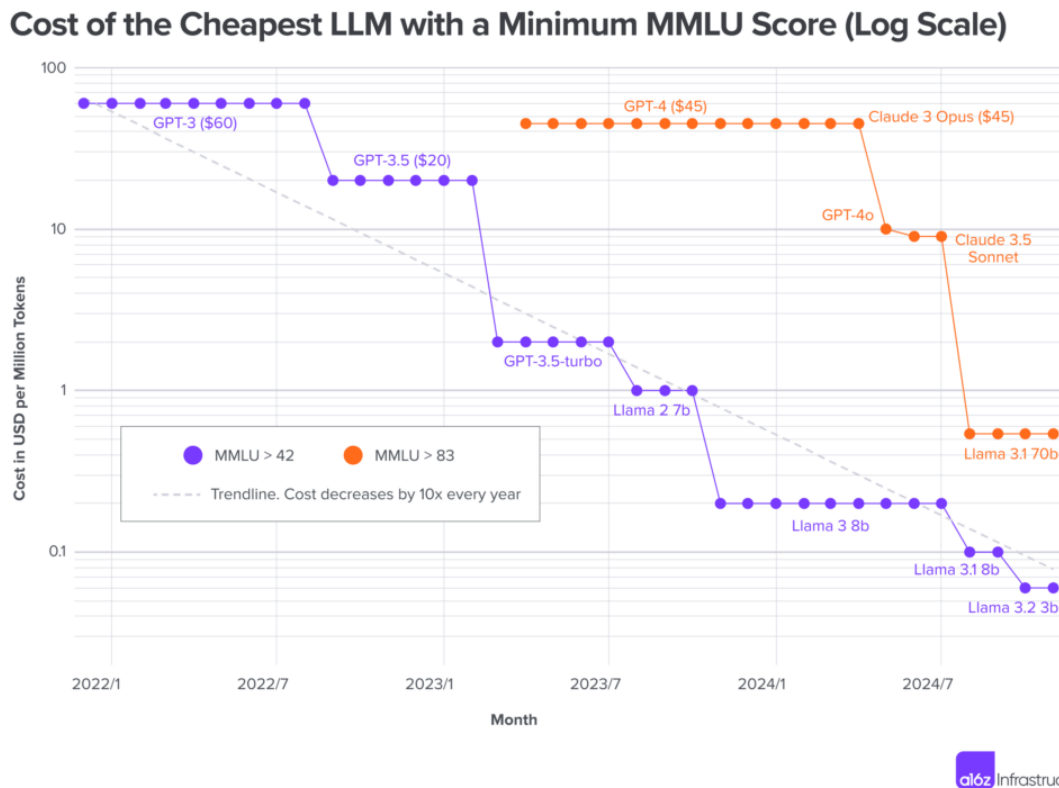


Figure 1: Guido Appenzeller, “Welcome to LLMflation – LLM inference cost is going down fast”, Andreessen Horowitz, available at: <https://a16z.com/llmflation-llm-inference-cost/>

In this white paper, we hope to provide more context to this fundamental question. First, we cover some market headwinds that make legal a difficult area for incoming startup founders to enter. Next, we discuss traps of false traction that founders should be made aware of. Finally, we end with some potentially interesting directions for future exploration.

## 2 Challenges in Legal AI Adoption

In this section, we discuss some of the challenges of building and selling within the legal-AI space, largely discovered empirically as we pursued different areas of research prototyping. At a high-level, we clustered these challenges to the difficulty of obtaining quality data, the law firm business model, the scattered subvertical, platform constraints, and cultural resistance.

### 2.1 Guarded and Proprietary Data

The quality of outputs generated by LLMs is a direct function of the quality of data they are trained on/use. Legal data is highly sensitive and closely guarded, creating a **chicken-or-egg problem**, in which AI systems require proprietary data for effective training. Yet, firms are hesitant to share data without first seeing tangible benefits, and accordingly, incumbent

Legal Tech companies view data as their highest demonstrable value. While publicly available datasets, such as patent data and/or other court data made available via [Free Law Project](#), provide a starting point, proprietary data is essential for building tools with real competitive value. This is because unlike data for standard writing like/coding tasks, for which a great deal of public data exists, legal data is tightly controlled under client-attorney privilege and within companies as their secret sauce, preventing LLMs from being geared towards legal writing. Moreover, the data that is released and published often only contains the final product, missing the intermediary stages in the pipeline, which are crucial for building reasoning capabilities within these models.

The best way founders can attempt to break this data wall is via well-structured pilots and design partners. However, even data from generous and willing firms can be hard to directly utilize. For instance, at CodeX, we partner with law firms to experiment and build legal AI tools, such as an automated redlining training platform. However, to protect client confidentiality, all shared data must necessarily be passed through redaction software, which generates PDFs with blacked out regions. Unfortunately, this requires leveraging complicated PDF parsers. There are also modality differences between taking these multi-page PDFs, mapping them into usable insights, and then applying these insights to MS Word documents. This often means that the best teams suited for these tasks have strong parsing and UI/UX backgrounds, rather than AI ones. However, these UI/UX backgrounds are quite different from the predominantly machine learning focused backgrounds of founders who are looking to penetrate legal AI.

## 2.2 The Law Firm Business Model

As advances in legal AI companies continue to push the frontier of legal services provision, the broader architecture of the law firm is placed into question. While this is not necessarily the first time the ‘traditional’ law firm model has experienced pressure to change, it is certainly one where, without change, could invoke potential implications to the future of the profession. In this section, we will briefly outline the history and context around the emergence of “[BigLaw](#),” defining key instances in which the organizational structure had inadvertently generated inertia and inefficiency not only in the development of legal professionals, but also in the integration of technology in the domain.

### **Historical Roots and Persistent Structures**

In the late 19th century, the notion of a “[large law firm](#)” was merely defined as a firm comprised of four or more lawyers. These firms responded to the demands of rapidly growing businesses that required increasingly bespoke, specialized transaction work. While there were only 15 firms recognized as a “large law firm” in 1872, these numbers exponentially soared such that, by 1924, there were over 1000 firms classified under this definition.

Nevertheless, what would come to be interpreted as the modern law firm, otherwise known as the “Cravath System,” emerged in the early twentieth century. The Cravath System was

a pyramidal organizational structure built with the intention of profit maximization. There were fundamentally only two “classes” of professionals: partners and associates. The goal was to incentivize associates to eventually be admitted as partners to the firm on a meritocratic basis. With law firms structured as partnerships, there is an implicit signal of belonging; that the duration of your tenure as an associate is conceivably a probationary period until you are joined in the ranks of your mentors. Furthermore, as partners are largely chosen within the firm, apprenticeship and culture hail from a rather insular heritage. The industry became a mirror to the fundamental “[priesthood](#)” of the legal profession.

Accordingly, following the footsteps of rainmakers and renowned attorneys was primordial to success in the business. Entering the legal sector thereby required traineeship, directly integrated into a Darwinian model of operation. In effect, work product creation in the profession and skill development and training are intimately interrelated.

In the 1960s, law firms began to develop “deep and enduring relationships with corporate clients”, pivoting and expanding the role of the legal professional from purely adversarial to advisory. These close institutional relationships became foundational to a law firm’s growth and development. Trust became the secret ingredient to the success of the business, enabling information asymmetry between corporate clients and lawyers. Put differently, reputation became a key driver behind business and operational strategy of a practice. In effect, these factors encourage a business model centered on the individual. The law firm became perceivably structured as a collective, rather than a corporation.

These structures persisted in the years that followed. Even amidst changes, including the rise of in-house legal departments and alternative legal service providers, the law firm remained focused on the individual practitioner and their proteges.

More importantly, unlike other domain specific industries, relational and reputational qualities of the legal practice, coupled with information asymmetry, replaced the need for the development of explicit quantifiable metrics to evaluate the work of legal professionals. Therefore, time and billing on time spent became a proxy for value, as value generated from the work could not be measured concretely.

The billable hour model, where revenue for firms is directly correlated with the number of hours spent on the task, is a cornerstone of the legal industry. This creates a structural disincentive for law firms to adopt efficiency-enhancing technologies like AI, as reducing hours worked translates to reduced revenue. While some firms are exploring alternatives, such as [fixed-fee pricing](#), the transition is gradual and fraught with challenges.

The legal profession has historically emphasized long hours as a marker of diligence and expertise. As such, many clients, particularly large corporations, have accepted billable hours as the standard pricing model. While some small-medium businesses and/or more agile clients have begun to push back against the inefficiencies in the pricing model, this implicit acceptance reduces the urgency for firms to innovate. As a result, while firms are

willing to entertain AI integration as conceivably important, it remains largely theoretically rather than an immediate demand for revenue generation via time and billing.

That said, this headwind is undergoing a renewed cycle of tension, potentially enabling an opportunity for founders who are able to build distributional moats. Often to reduce friction from the industry, recommendations for founders often dwell in the monotonous, low-effort tasks done by junior associates in the firm. We have observed from certain future-forward firms [Ashurst](#), [Baker McKenzie](#), and [Wilson Sonsini](#) that these areas of higher inefficiency have made the relative lift gained from using AI more significant to clients.

## 2.3 Technical Constraints

Lawyers are deeply integrated into the Microsoft Office suite, and getting lawyers to AI-first platforms presents high migration friction (i.e. complex change management). As a result, founders are forced to build on top of Microsoft's relatively clunky feature suite, which enforces bottlenecks on the design paradigms they can use. In our personal redlining work, for example, we found that line/word level edits made programmatically are really difficult to show on the UI - instead you have to replace an old paragraph with a new paragraph, making changes hard to process and identify. Furthermore, an ideal interface might resemble a legal copilot, with text autocomplete, and section insertion enabled. However, in the short-term, these platforms are blocked given difficulty creating event wrappers (programmatic functions that run in response to certain actions taken by a user instead of always running) typing in word.

Additionally, LLMs are bad pointwise editors. While lawyers can be surgical with their edits on large documents, requiring on the removal of minimal information to achieve their goals, LLMs regenerate the entire document. This becomes problematic for large 100+ page files, since this generation process enables hallucinations, as token probabilities diverge. Our work explored workarounds to this using an agentic approach, where we first identify relevant sections and only have the LLM generate new content for the section to be inserted/removed/edited. This tends to improve results; however, it reduces global reasoning accuracy. Further research on pointwise edits could be helpful for improving the efficacy of legal AI tools.

## 3 Enterprise Adoption Models: The Promise and Pitfalls

Legal AI startups often find early success through pilot programs, where firms experiment with technology in controlled environments. While these pilots offer valuable opportunities to build relationships and demonstrate feasibility, they also expose founders to early traps to consider. Navigating these challenges requires startups to focus on high-priority problems, demonstrate clear ROI, and build trust through transparency.

### 3.1 Fragmented Market:

To demonstrate deep ROI, founders often have to focus on a specific legal workflow. However, some back of the envelope math shows that these contract sizes are small. Even generously assuming 100 seats added per firm and \$100/month seats, yearly revenue per contract would only be around \$120K. Even expanding to 1000 seats would restrict revenue to \$1.2M/yr, making the upfront sales costs hard to justify. Further complicating the matter, the legal industry is extremely fragmented. While there are exceptions, generally large firms tend to handle complex, high-value cases requiring specialized solutions, while smaller firms often perform broader, resource-constrained work. Targeting the larger firms, therefore, pushes founders to develop more bespoke products that do not easily scale from firm to firm. These firm-to-firm differences, in turn, slow fast distribution as multiple product lines need to be maintained – positioning a tech startup as pseudo consultancy. Furthermore, data privacy issues mean that it is very difficult to have true network effects. The products must be returned to every partner using their internal proprietary data.

### 3.2 Long Sales Cycles:

Through experience from our own pilots with our partnered law firms, we discovered that adoption requires buy-in from multiple stakeholders—partners, IT teams, and innovation leaders—which, for startups, often translates to long sales cycle ([3 - 18 months](#)). Moreover, given the “nice-to-have” but not crucial status for most AI tools right now, it is difficult to design a monopolistic product that attracts inbound from customers. Instead, founders have to sell to many different partners, with each process being a separate pilot and long sales cycle. In many ways, this makes an ideal legal AI founder one who has strong go-to-market connections and experience rather than a purely technical background. There are certainly ways to bridge these deficits – for example, bringing on legal cofounders, advisors, or early hires, who can expedite the trust-building phase.

### 3.3 The Illusion of Early Wins:

Law firms are generally open to pilots and are well-positioned for the AI wave they anticipate. However, early successes in securing pilots may mislead founders into believing the market is ripe for exploration. Specifically, through conversations with partners at prominent BigLaw firms, we found that firms are typically testing many different external partners for adoption. First, the conversion rates between acquiring pilots to acquiring paying customers can be surprisingly low. Additionally, these pilots often involve a specific group within the firm (typically on the order of 20-50 people) testing out your product and sharing feedback on it and other products at the end of the trial period. To make a lasting impression, founders should focus on a specific problematic workflow to improve, as such a lift on a high-pain workflow can leave a lasting impression.

## 4 Interesting Directions in Legal AI

Several promising areas of innovation have emerged to address the industry's challenges and unlock AI's potential:

### 4.1 Legal Personas

Personalized AI personas - small LLMs that behave like the people they represent - are a significant leap in the realm of artificial intelligence (AI). Especially in the legal domain, these personas offer a unique opportunity to preserve and translate partner knowledge, enabling them to scale their expertise and client services to new frontiers. By virtue of being trained on highly curated information, they act as a personalized bank of knowledge that others can interact with and learn from. These personas hinge on the core thesis that we don't want to make a general all-purpose AI lawyer, but one that is specifically geared towards a partner's background and experiences. This makes a lot of sense given law's partnership and interpersonal model. All deals are obtained and managed on the basis of relationships, which makes capturing individual style and behavior important. We will be releasing a forthcoming in-depth paper detailing our research in this space, including findings with simulating legal experiences via multi-agents.

It is perhaps easier to see the utility of such personas within specific contexts. For example, consider running simulations of personas of different stakeholders responding to changes in policies or firm structure. These simulations can help us identify adverse effects of changes in advance, saving wasted time and effort. Alternatively, these personas can be used to educate junior attorneys faster by injecting additional context into redlines from their partners using partner personas. These partner personas would draw on how partners have historically redlined, drawing on similarities to previous examples to better equip associates to learn from each redline.

### 4.2 Computational Law

Startups could benefit from experimenting with robust versions of computational law – composable and verifiable building blocks that can be used to write contracts, briefs, etc. The underlying thesis here is that it is possible to express legal strategies in a formally verifiable language to reduce the risk of hallucination. Given law is otherwise very sensitive and requires accuracy, these verification mechanisms are clearly important. Unfortunately, most verification methods right now involve generating drafts and then applying a verifiability model at the end. But there can be cases where divergence happens very early in the process, and we want to detect these issues far earlier. That said, unlike agentic systems for math, which correctness may be guaranteed, it is quite difficult to have a verification system for language – which the legal domain entails. Additionally, many areas of law are subjective, where it is difficult to arrive at these explicitly correct answers.

## 4.3 Patent Automation

Another promising area for AI is within patent law, where the data is typically well-structured and publicly available. This makes the data challenges that make building in other areas of law so difficult much less of an issue, reducing entry barriers for newcomers to law. We anticipate AI tools like semantic retrieval, graph-based retrieval, and agentic retrieval will disrupt traditional workflows like prior art searches, invalidity analyses, and patentability assessments. This is especially useful when traditional keyword searches fail due to minor differences in the framing of search queries. With current alternatives involving humans explicitly sifting through these documents to identify relevant infringements/invalidation claims, automated search tools to do this faster and cheaper enable patent attorneys to focus on the more interesting aspects of drafting patents or invalidating other patents, instead of sifting through prior art. Additionally, given that many large firms explicitly ask external search firms to find prior art on their behalf, selling these products doesn't result in explicit reduction in billable hours for law firms, increasing their openness to adopting them.

However, entering the patent space has its challenges as well. The market for patent-related tools is becoming increasingly saturated, making differentiation difficult. This exposes founders in the patent space to get commoditized unless they are able to find data or distributional moats. Furthermore, while embeddings-based searches are optimized for detecting macro level similarities, they tend to perform poorly when two patents are only weakly related on the basis of a small feature. However, these similarities are important to identify as they can be basis for an infringement. Moreover, default embeddings are trained on a concept space that is quite different from the language that patents follow. This means startups have to make significant upfront investments in training their embedding modules. Additionally, there are some UI/UX processes to consider. To stand out, AI solutions must not only improve efficiency but also offer clean visualizations of claim dependencies, seamless integration with patent office databases, and help attorneys craft arguments around prior art matches.

## 5 Conclusion

Synergies between Gen AI and law can drastically increase efficiencies within the legal market. This insight coupled with law's primarily text-based modality has attracted many young founders to experiment with building companies in the space. However, the path forward requires addressing gaps like proprietary data, the billable hour model, cultural resistance, and integration hassles with current tool suites. Moreover, the ease of acquiring design partners/pilots can be a red herring for founders unless they realize that pilots do not necessarily equate to revenue and that selling to additional customers doesn't necessarily scale easily. There are some promising areas of legal AI that are still being developed and that I'm personally passionate about, such as creating legal personas, building a framework for computational law, and streamlining patent prosecution/litigation. Carefully navigating these opportunities and challenges could help make the legal industry faster, more efficient, and more accessible for all.