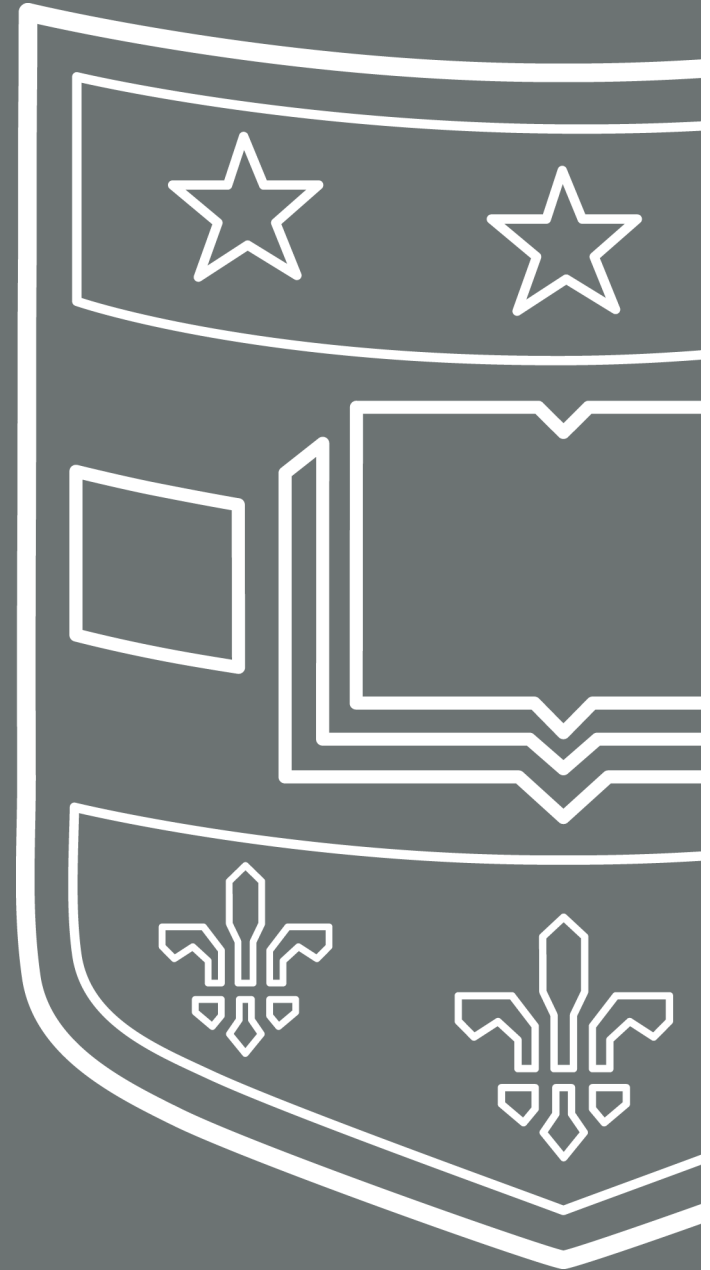


Enhancing Paralegal Workflows with AI: Ethics, Applications, and the Future

James Daily





What is Artificial Intelligence (AI)?

"Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."

– Nils Nilsson

"AI is whatever hasn't been done yet."

– Larry Tesler

What is Machine Learning (ML)?



CLASSIFICATION

Is this email relevant to this e-discovery request?

Is this contract clause market standard?

PREDICTION

How is the judge likely to rule on this motion?

How long will it take?

SEARCH

What cases can I cite for this proposition?

What other authorities might be relevant to this brief?

QUESTION ANSWERING

What law governs this contract?

When are non-compete agreements enforceable in California?

TEXT GENERATION

What are some questions I should ask when deposing this witness?

Write a summary of this legislation.

An Example: Technology Assisted Review (TAR)



2005

Anne Kershaw establishes that automated review can exceed human performance.

2006

TREC establishes its Legal Track.

2011

Maura Grossman and Gordon Cormack definitively demonstrate that TAR is both better and faster than human review.

2012

Da Silva Moore v. Publicis Groupe is the first case to endorse TAR.
Global Aerospace v. Landow Aviation is the first case to endorse TAR over an opponent's objection.

2017

64% of the 318 corporate counsel polled by Norton Rose Fulbright report using TAR.

12 Years from "What Hasn't Been Done Yet" to Ordinary,
6 Years from Permissible to Widespread



Why Did Technology Assisted Review Succeed?

1. E-discovery is increasingly unavoidable.
2. Manual review is time-consuming, expensive, does not scale well, and is rarely where cases are won or lost.
3. Objective tests conclusively demonstrated that TAR was better. (Kershaw 2005; Grossman & Cormack 2011).
4. Humans stayed in the loop, but at much reduced time and cost.

2023: GPT-4 Passes the Bar Exam

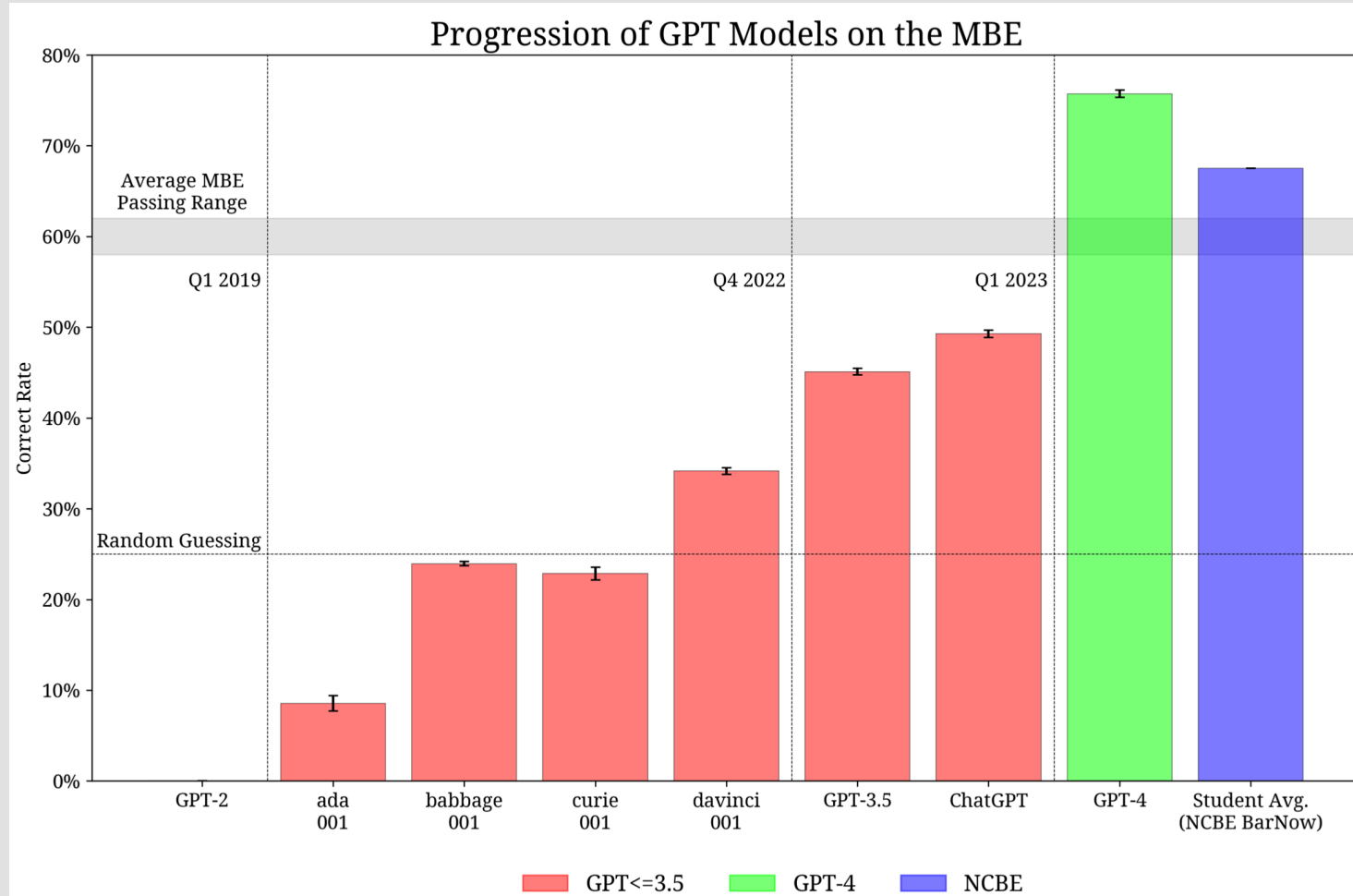


Figure 1. Progression of Recent GPT Models on the Multistate Bar Exam (MBE)

Katz et al. (2023)



2025: Or Does it?

- GPT-4's "90th percentile" ranking was based on the *February* bar, which includes many repeat test takers who failed the July exam
- Based on July exam scores, GPT-4's performance falls to the 68th percentile
- Compared to first-time exam takers, GPT-4's performance falls to the 48th percentile
- Compared to those who successfully passed the bar, GPT-4 falls to the 45th percentile

Martínez (2025)



The Takeaway

- GPT-4 has since been replaced by much more capable models that can easily pass the bar exam
- But nonetheless:
 - Question the hype
 - Question the methodology

So what do the legal tech companies say about this?

RELX (LexisNexis) Responsible AI Principles



#3 WE CAN EXPLAIN HOW OUR SOLUTIONS WORK

An appropriate level of transparency creates trustworthiness for users and regulatory bodies.

This principle doesn't prescribe what models to build and use, and it doesn't prohibit the use of "closed box" models. The aim is to have an appropriate level of transparency for each application and use-case to ensure that different users can understand and trust the output.

Different contexts and audiences require different explanations. As part of the design process, we consider what elements of the solution will need explaining, to whom they might need to be explained, and how to go about explaining them. We also evaluate a solution's reliability and are explicit about its intended use.

Thomson Reuters AI Ethics Principles



7 Transparency

- We implement practices to make the use of data and AI in our products and services understandable.



Benchmarks? What Benchmarks?

- Foundation model companies (e.g. OpenAI, Anthropic, Google, and Meta) all use industry standard benchmarks and publicly disclose failure modes
- Most legal tech companies do not publish benchmarks or discuss failure models, despite transparency pledges. Harvey is the notable standout in this area, but even it only recently created open benchmarks.
- When asked, they (including Westlaw, Lexis, and Spellbook) have explicitly declined to provide more information on the subject



What Independent Researchers Have Found

Table 5: Comparison of STARA with other AI systems on the federal criminal offenses task

System	# Found	Precision	Recall
STARA	1,983	0.98	0.998
Gemini Deep Research	282	0.890	0.144
OpenAI Deep Research	84	0.881	0.044
Westlaw AI Jurisdictional Survey	113	0.415	0.072

Surani et al. (2025)

The Deep Research systems achieve reasonable precision (0.88–0.89), but Westlaw AI Jurisdictional Surveys exhibited both poor recall and concerning reliability issues. Despite being a specialized legal research tool with access to a legal database, Westlaw achieved only 41.5% precision—not only mis-classifying non-criminal provisions, but also hallucinating 121 statutory provisions in the U.S. Code that do not actually exist.



What Independent Researchers Have Found

Recently, however, legal technology providers such as LexisNexis and Thomson Reuters (parent company of Westlaw) have claimed to mitigate, if not entirely solve, hallucination risk (Casetext [2023](#); LexisNexis [2023b](#); Thomson Reuters [2023](#), *inter alia*). They say their use of sophisticated techniques such as retrieval-augmented generation (RAG) largely prevents hallucination in legal research tasks.¹ (We provide details on RAG systems in Section [3.1](#) below.)

However, none of these bold proclamations have been accompanied by empirical evidence. Moreover, the term “hallucination” itself is often left undefined in marketing materials, leading to confusion about which risks these tools genuinely mitigate. This study seeks to address these gaps by evaluating the performance of AI-driven legal research tools offered by LexisNexis (Lexis+ AI) and Thomson Reuters (Westlaw AI-Assisted Research and Ask Practical Law AI) and, for comparison, GPT-4.

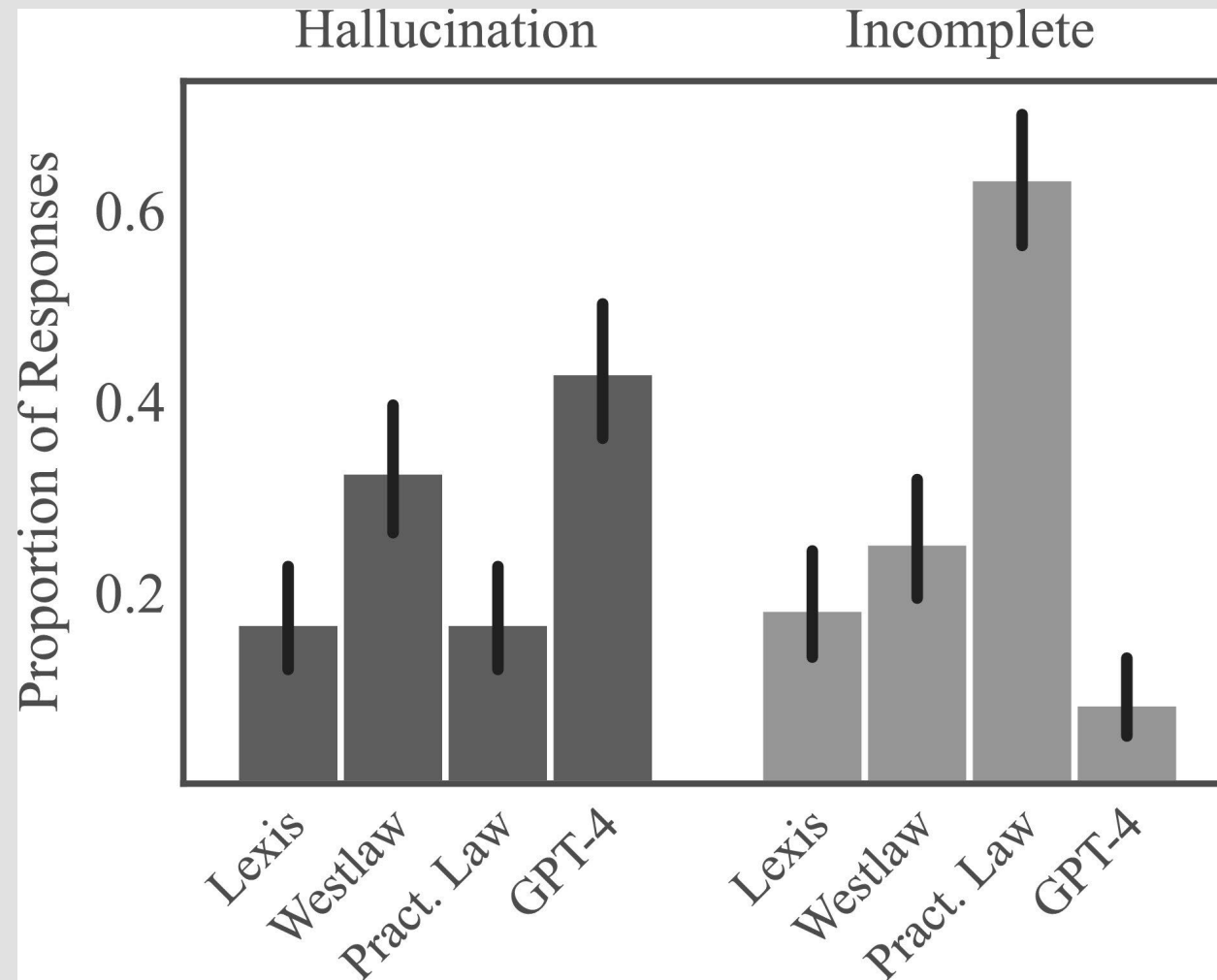
Magesh et al. (2025)

What Independent Researchers Have Found



Our findings, summarized in Figure 1, reveal a more nuanced reality than the one presented by these providers: while RAG appears to improve the performance of language models in answering legal queries, the hallucination problem persists at significant levels. To offer one simple example, shown in the top left panel of Figure 2, the Westlaw system claims that a paragraph in the Federal Rules of Bankruptcy Procedure (FRBP) states that deadlines are jurisdictional. But no such paragraph exists, and the underlying claim is itself unlikely to be true in light of the Supreme Court's holding in *Kontrick v. Ryan*, 540 U.S. 443, 447–48 & 448 n.3 (2004), which held that FRBP deadlines under a related provision were not jurisdictional. ²

What Independent Researchers Have Found



What Independent Researchers Have Found



Input Query

Are the deadlines established by the bankruptcy rules for objecting to discharge jurisdictional?

Westlaw AIAR Response

[A] **paragraph from the Federal Rules of Bankruptcy Procedure**, Rule 4007 states that the deadlines set by bankruptcy rules governing the filing of dischargeability complaints **are jurisdictional**. [...]



Input Query

What standard of review applies to abortion regulations under the U.S. Constitution?

Lexis+ AI Response

Currently, the “undue burden” test, as established in *Planned Parenthood of Southeastern Pa. v. Casey*, is the governing standard. Planned Parenthood of the Heartland, Inc. v. Reynolds, [...]. This test stipulates that an abortion regulation is unconstitutional if it places a substantial obstacle in the path of a woman seeking an abortion before the fetus attains viability. [...]



Input Query

Why did Justice Ginsburg dissent in Obergefell?

Practical Law Response

Justice Ginsburg dissented in Obergefell because she believed that the court's interpretation undermined Congress's intent under Section 601(a) (1) to protect copyright owners against unauthorized importation of low-priced, foreign-made copies of copyrighted works. She argued that the text of the Copyright Act clearly applies to works governed by the statute. [...]



Input Query

Do any special requirements apply to online dating services in Connecticut?

GPT-4 Response

Yes, Connecticut has specific legal requirements that apply to online dating services. Under Connecticut law, particularly **Conn. Gen. Stat. § 42-290**, online dating services that charge a fee and have members who reside in Connecticut must provide their users with specific safety awareness notifications. These notifications must include: [...]





What Real Benchmarks Look Like

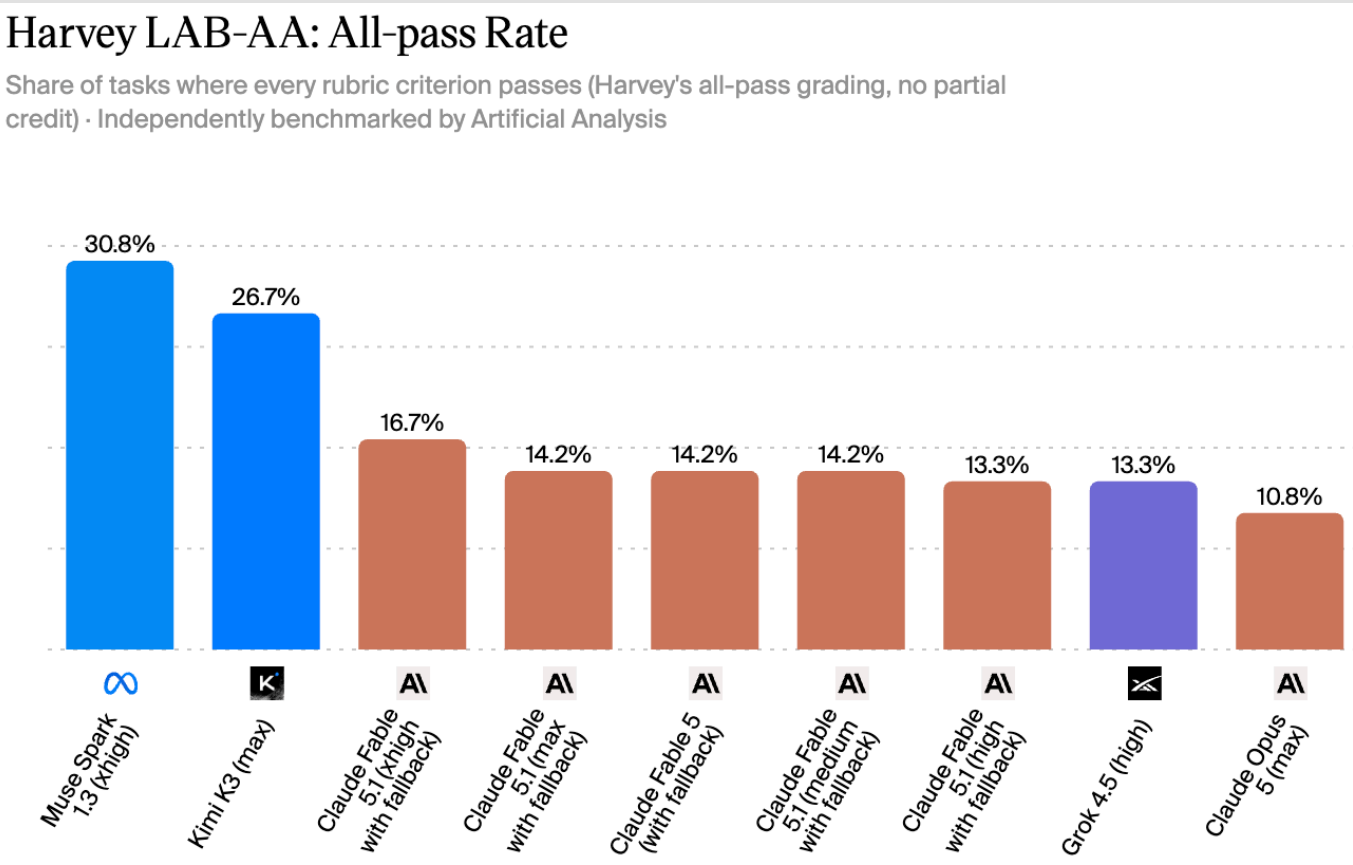
Harvey released the open source Legal Agent Benchmark in May 2026.

“The first version of LAB includes more than 1,200 agent tasks across 24 legal practice areas, and is evaluated by over 75,000 expert-written rubric criteria.”



What Real Benchmarks Look Like

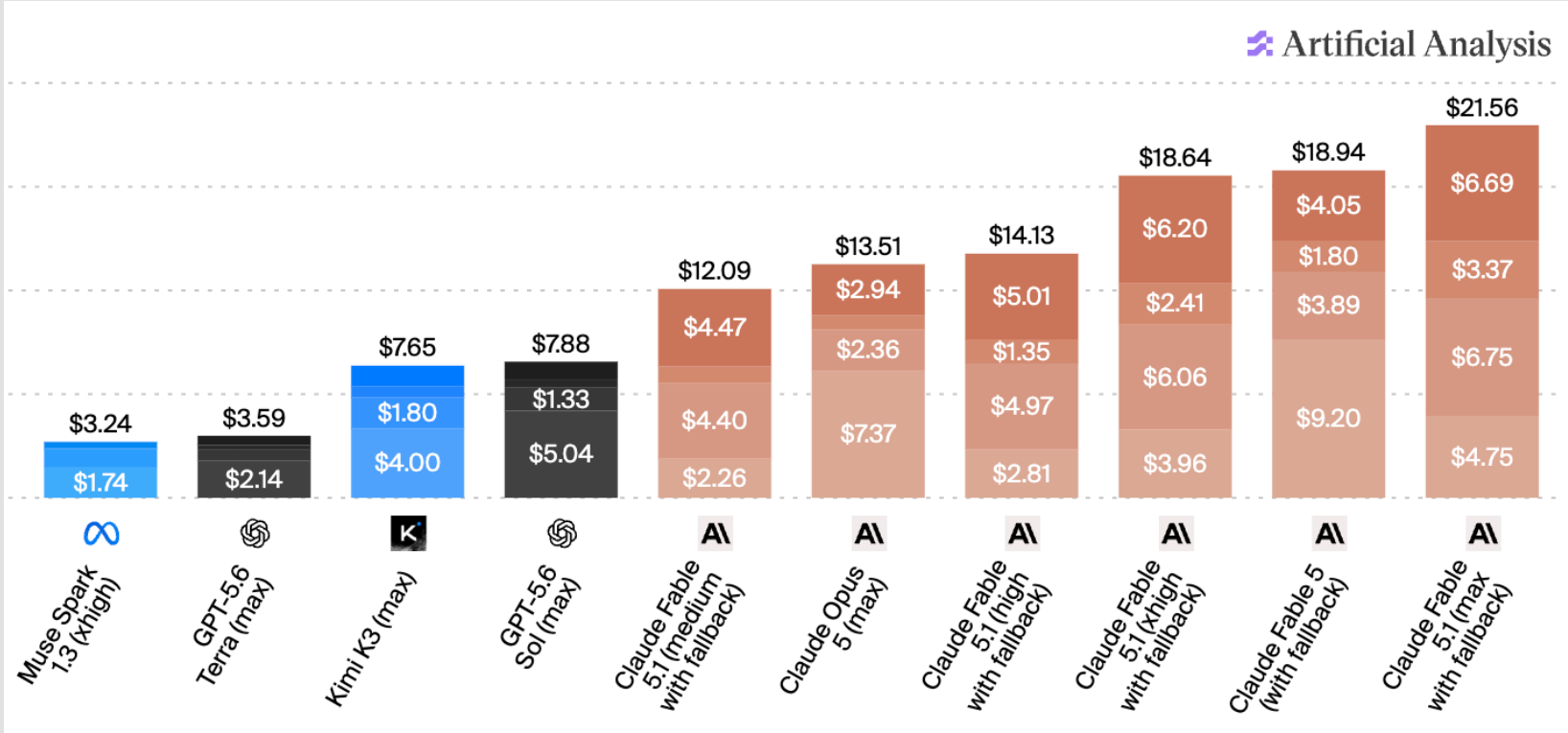
By September 2026 this is what the leaderboard looks like





What Real Benchmarks Look Like

And the cost per task



Note that the best performer was also very cheap!



The Implications of Non-Transparency

- How can we make informed decisions without objective data? (There are ethical implications here, too!)
- What does it suggest that these companies (allegedly) have rigorous internal benchmarks but do not publish them?
- What does it suggest that the companies know the products are imperfect but will not tell us in which ways?

Westlaw CoCounsel in March 2023



What is the current minimum wage in Missouri?

Answer

According to the legal authorities provided, the current minimum wage in Missouri is \$11.15 per hour. This rate is the result of Proposition B, which was passed by Missouri voters in 2018 and increased the state's minimum wage in 2019. Proposition B also provides for annual increases until the minimum wage reaches \$12.00 per hour in 2023. After that, the minimum wage may increase or decrease each

Thanks for waiting for me. At this current time, CoCounsel is under the assumption that it's 2020 and doesn't understand what the current year is. We're working on fixing that soon!

CoCounsel Deep Research in October 2025



What is the minimum wage in Missouri?

Summary

The minimum wage in Missouri is \$13.75 per hour, effective January 1, 2025 ( [V.A.M.S. 290.502](#))[[1](#)]. This rate will increase to \$15.00 per hour on January 1, 2026 ( [V.A.M.S. 290.502](#))[[1](#)]. Missouri law requires employers to pay whichever is higher between the state minimum wage or the federal minimum wage rate ( [V.A.M.S. 290.502](#))[[1](#)].

More Like CoCoun-Saul, Am I Right?



James – 05:20 PM

You are an attorney for Walter White, owner of a chemical manufacturing business in Albuquerque, New Mexico. Mr. White is considering a business partnership with Mr. Fring, owner of a fried chicken franchise and a laundry facility as well as his own chemical manufacturing company. Please draft a letter to Mr. Fring offering to discuss a partnership, subject to a mutual NDA.



✦⁺ Draft

Download

October 19, 2025

Mr. Gustavo Fring

[Title]

[Company Name(s): Los Pollos Hermanos; [Laundry Facility Name]; [Chemical Manufacturing Company Name]]



HANK! HAAANK!

**THE UNDERLYING LLMS
MIGHT MAKE UNSTATED ASSUMPTIONS
ABOUT YOUR INPUT, HANK!**

CoCounsel is not actually your co-counsel

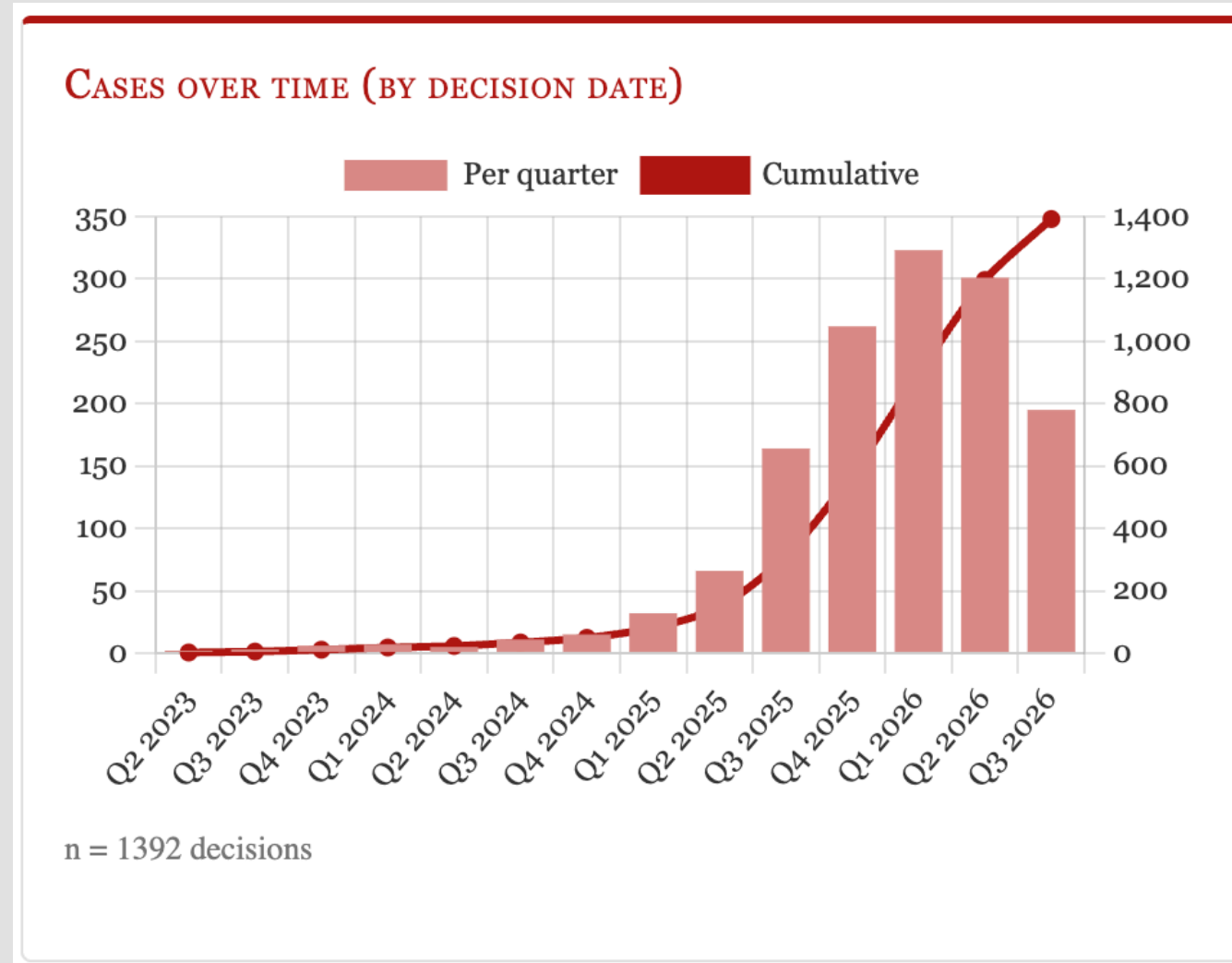


CoCounsel's Draft skill uses a combination of the underlying large language model (LLM), your provided materials, and, when explicitly cited, Thomson Reuters proprietary content to generate legal and non-legal documents. Any citations, including hyperlinked citations to Thomson Reuters content, should be independently verified. As a legal professional, you are ultimately responsible for ensuring the final work is accurate and does not misrepresent the law.

What Happens If You Don't Check Everything?



Opposing
counsel or the
court will call you
on it!

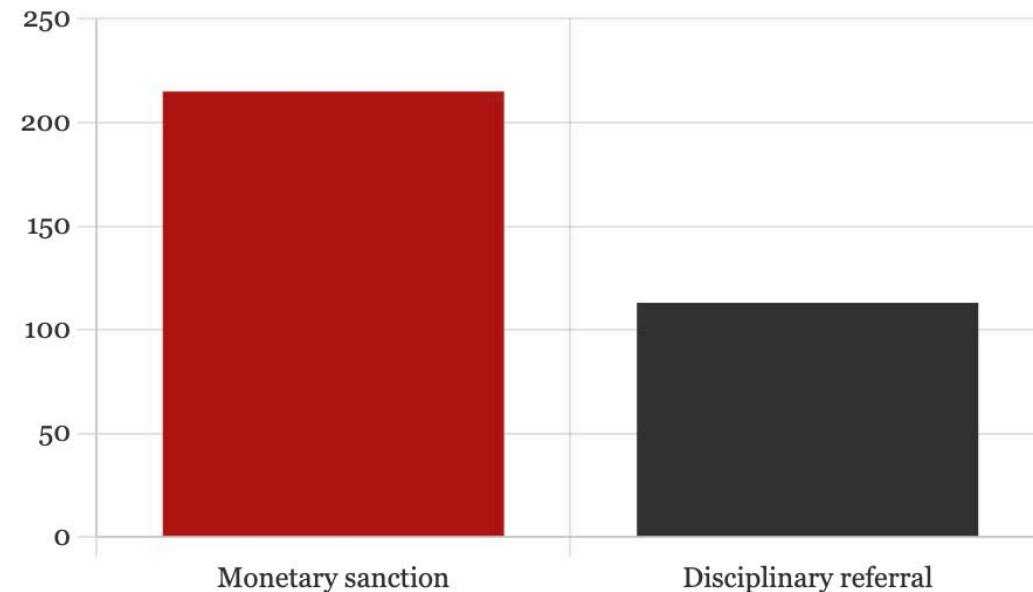


Credit: <https://www.damiencharlotin.com/hallucinations/>

What Happens If You Don't Check Everything?

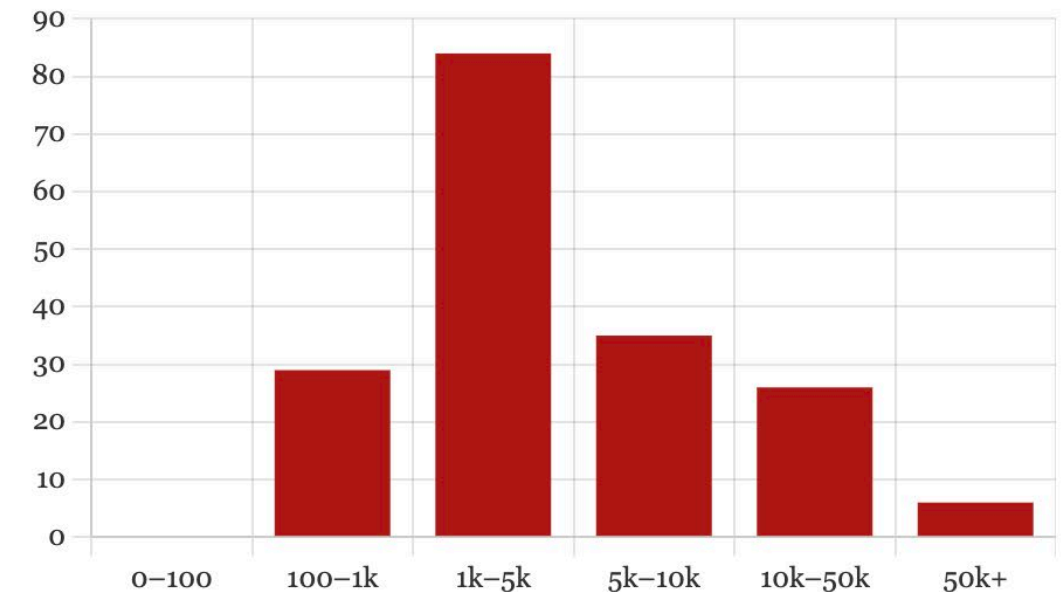


OUTCOME SEVERITY



n = 1386 cases (excludes 6 alleged-only)

MONETARY PENALTY DISTRIBUTION (USD) ①



n = 180 USD penalties · 7 in other currencies (not shown) · 28 with an unknown amount (excluded)

This includes one case in E.D. Mo in which a party was ordered to pay costs & a \$1000 fine. One attorney was ordered to withdraw.

Credit: <https://www.damiencharlotin.com/hallucinations/>



Where Is This Going?

The complexity of tasks** that an LLM can reliably*** complete autonomously has been doubling every 4 to 6 months since 2019*

* As measured by the number of hours it takes humans with 5+ years of experience in the field to do it

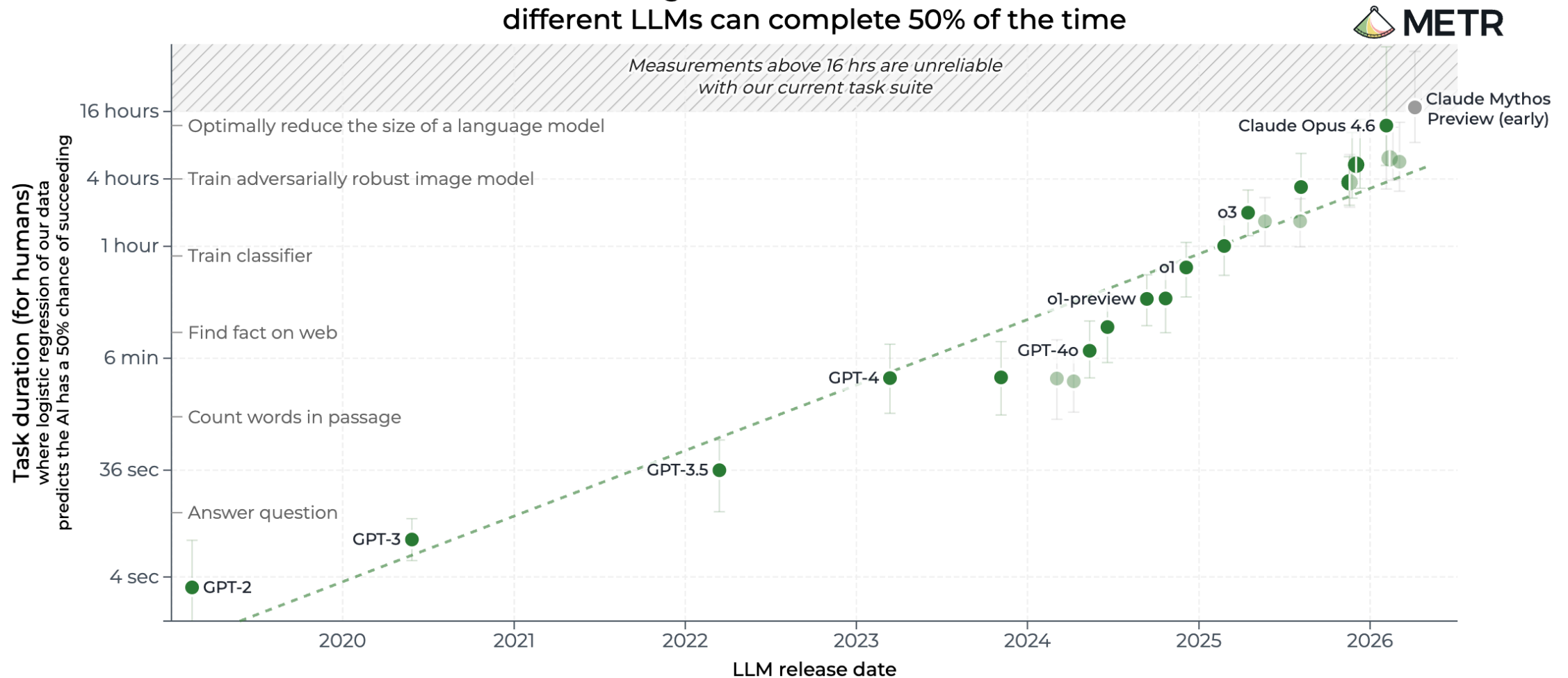
** Primarily in the software development context

*** >50% of the time

Where Is This Going?



Length of software tasks that different LLMs can complete 50% of the time





Are the Machines Coming For Our Jobs?

Consider Comment 6 to Rule 4-1.1:

“To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology”

and its connection with Rule 4-5.3 regarding supervising nonlawyer assistants.

Lawyers are ultimately responsible for their own competence, but having technologically competent assistants can be very helpful!



Are the Machines Coming For Our Jobs?

A few possible scenarios:

1. Legal AI replaces paralegals
2. Paralegals leverage legal AI
3. Paralegals leverage legal AI to replace (some) associates

Which scenario you are in probably depends more on your firm than anything, but a good rule of thumb is “If you’re not the one doing the automated, you will be automated”

Questions?

jedaily@wustl.edu

