

## AI-Powered Lawyering — A Summary

---

A summary of Schwarcz, Manning, Prescott, Barry, Cleveland & Rich, AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice, *Journal of Law and Empirical Analysis* (2026), [DOI 10.1177/2755323X261427048](https://doi.org/10.1177/2755323X261427048). Draft posted on SSRN in March 2025: [ssrn.com/abstract=5162111](https://ssrn.com/abstract=5162111).

---

### What the paper does

This is the first randomized controlled trial on whether recent-generation AI tools make legal work better. The authors recruited 137 upper-level law students from Minnesota and Michigan and assigned them to complete legal tasks under one of three conditions: using a RAG-powered legal tool (Vincent AI, mid-2024 version), an AI reasoning model (OpenAI’s o1-preview), or no AI at all. Six tasks: drafting a persuasive letter, drafting a contract, analyzing a complaint, two research memos, and a related exercise. Standardized grading. Standardized hallucination counts.

That combination — RCT, real legal tasks, real law students, standardized grading — is rare in this literature. Most of what we know about whether AI helps legal work comes from vendor benchmarks (easy to game), anecdote (hard to generalize), or classroom observation (hard to control). Schwarcz and co-authors built something closer to a clean study.

### What they found

Two headline results.

**Productivity gains were large.** Both AI tools produced statistically significant productivity gains of 50–130% across five of the six tested legal tasks. The largest effects were on complex tasks like drafting persuasive letters and analyzing complaints, where AI assistance roughly cut task time in half or better.

**Quality gains were real.** Access to both Vincent AI and o1-preview considerably improved the quality of legal work product in at least four of the six tasks — a marked departure from earlier studies of GPT-4, which showed minimal quality gains. The o1-preview condition improved *analytical depth* specifically — the kind of structured reasoning a careful reader notices.

**Hallucinations split by tool.** Vincent-assisted participants produced roughly the same number of hallucinations as the no-AI control — RAG grounding did what it was supposed to do. o1-preview-assisted participants produced more hallucinations than the control — reasoning models without domain grounding still invent facts.

The authors’ synthesis: the combination of domain-specific RAG *plus* reasoning models would plausibly yield better results than either alone.

## Why this is load-bearing for the course

Three reasons this paper matters more than the many recent “AI is good at legal tasks” pieces.

**It’s an RCT.** The literature needed one. Selection-effect problems dog observational studies of legal AI: firms that adopt AI tend to be firms that are already good at workflow, measurement, and change management — so their post-adoption outcomes are confounded. An RCT on a reasonably clean population (upper-level law students on standardized tasks) reduces that noise.

**It tests the right generation of tools.** o1-preview was the first widely-available reasoning model; Vincent AI was one of the first legal RAG products that actually worked. The paper doesn’t test GPT-3.5. It tests the stack that matters now.

**It shows the productivity and quality dimensions separately.** Most vendor claims collapse speed and quality into a single “productivity” number. The RCT distinguishes them, which matters for how law schools think about training — if AI substantially speeds analytical work while raising quality more modestly, the training implication is that judgment and review become the bottleneck. That maps onto the course’s thesis.

## What to keep in mind

The paper tested models from mid-to-late 2024 — the mid-2024 version of Vincent AI and OpenAI’s o1-preview (released September 2024), with data collection in fall 2024. The field has moved through at least two model generations since: GPT-4 → o1 → o3 / GPT-5-series, Claude 3.5 → Opus 4 → Opus 4.5, and the legal-specific tools have iterated. Extrapolating the specific numbers forward is risky. The direction is almost certainly right; the specific percentages aren’t worth the paper they’re printed on by mid-2026.

Also: the subjects were upper-level law students, not practicing attorneys. Practicing attorneys should in principle benefit less (they’re already good at the underlying tasks) but benefit more in relative terms on tasks where AI removes drudgery they’d otherwise do by hand. The RCT can’t tell us which.

## One line worth remembering

*Vincent-assisted participants hallucinated at the same rate as the control; o1-preview-assisted participants hallucinated more.* The grounding matters. Reasoning models without grounding are powerful and dangerous. Reasoning models with grounding are the direction.

---

*Full citation: Daniel Schwarcz, Sam Manning, J.J. Prescott, Patrick Barry, David R. Cleveland & Beverly Rich, AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice, J. L. & Empirical Analysis (2026), [DOI 10.1177/2755323X261427048](https://doi.org/10.1177/2755323X261427048). SSRN preprint: [ssrn.com/abstract=5162111](https://ssrn.com/abstract=5162111).*