UPLBench explained: auditing whether AI chatbots cross into the unauthorized practice of law

24 November 2026 · 8 min read · blogs.iictp.org/uplbench-chatbots-unauthorized-practice-of-law-explained

Short answer: UPLBench is a benchmark, introduced in an arXiv preprint submitted on 13 November 2026, built from court decisions that found “unauthorized practice of law” (UPL) in high-stakes areas such as immigration and family law. The authors audited 23 chatbot models from 2023 to the present and report that modern chatbots increasingly perform tasks courts reserve for lawyers and decreasingly recommend consulting one. This is an arXiv preprint. It has not been peer reviewed, and its numbers are the authors’ own claims.

TL;DR The short version

- UPLBench, from an arXiv preprint, is built from court decisions that found unauthorized practice of law in areas such as immigration and family law.
- Auditing 23 chatbot models since 2023, the authors report that chatbots increasingly do tasks courts reserve for lawyers and less often recommend consulting one.
- Rules differ by jurisdiction, so anyone deploying a legal chatbot should define what it must refuse and test it with real high-stakes scenarios reviewed by a licensed lawyer.

Read online, listen to it, or change the text size: https://blogs.iictp.org/uplbench-chatbots-unauthorized-practice-of-law-explained

What is UPLBench?

The paper, “Auditing Chatbots for Incursion into Professional Legal Practice”, notes that legal advice has long been the exclusive domain of licensed attorneys and that UPL doctrine protects the public from incompetent counsel. The authors argue the boundary is being redrawn by AI companies rather than courts or bar authorities, as models are reconfigured to refuse less and be more helpful. UPLBench turns legal doctrine into a test.

How it works

- **Source material.** The benchmark is derived from court decisions finding UPL in high-stakes domains such as immigration and family law.
- **What is measured.** For each model, whether it refuses or provides substantive legal advice, and whether it refers the user to lawyers or legal aid.
- **Models.** 23 chatbot models dating from 2023 to the present were audited.
- **Expert review.** Subject-matter attorneys reviewed transcripts and identified interaction patterns that would count as inadequate counsel.

What is new compared with earlier approaches

Most prior work establishes what chatbots can do, for example bar-exam performance. This paper tests whether chatbots follow the law about who may give advice, by converting UPL doctrine into a testable benchmark.

Results reported by the source

- Modern chatbots increasingly perform tasks that courts reserve for lawyers and decreasingly recommend that users consult lawyers.
- Attorney reviewers found patterns such as failure to elicit relevant facts that they considered inadequate counsel.
- The authors say acting on such advice could cause harms such as deportation, loss of a home or family separation, and call for the legal profession to respond. They weigh options including compelling models to comply with existing law.

Limitations and open questions

- It is a preprint and the abstract gives no per-model scores, so we cannot report them; read the paper for the tables.
- UPL rules differ by state and country, and a benchmark built from court decisions encodes the doctrine of those courts.
- Whether a general chatbot's information amounts to "practising law" is a legal question the paper argues rather than settles.
- The authors' policy conclusions are advocacy as well as findings.

How to apply it

1. If you deploy a legal-facing chatbot, define what it must refuse and when it must refer to a lawyer or legal aid.
2. Build test prompts from real high-stakes scenarios in your jurisdiction and review responses with a licensed lawyer.
3. Check whether the system asks for the facts a lawyer would need before answering.
4. Re-test whenever the model or system prompt changes.

What it means for lawyers and legal-tech teams

For lawyers it is a method for evidence: a repeatable audit of a chatbot's legal behaviour. For legal-tech builders it is a reminder that "helpful" and "permitted" are different tests. See HireWatch for a related compliance benchmark and SG-LegalClarify. This is general information, not legal advice.

Frequently asked questions

- **What is UPLBench?** A benchmark derived from court decisions finding unauthorized practice of law, used to audit AI chatbots.
- **How many models were tested?** 23 chatbot models from 2023 to the present, according to the paper.

- **What was the main finding?** Modern chatbots increasingly do tasks reserved for lawyers and are less likely to recommend consulting one.
- **Is this a legal ruling?** No. It is a research audit in a preprint.

Related reading

- HireWatch: LLM compliance with employment law
- SG-LegalClarify benchmark
- What is LLM hallucination?
- AI Practitioner

Sources and currency

Sources were read on 24 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail. The numbers and claims come from the paper's abstract.

> Related credential: AI Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Auditing Chatbots for Incursion into Professional Legal Practice. arXiv 2611.17622, 13 November 2026 (preprint). arxiv.org/abs/2611.17622.

Published by IICTP, an independent technical credential body, on 24 November 2026. The current version of this post is at https://blogs.iictp.org/uplbench-chatbots-unauthorized-practice-of-law-explained. More guides: blogs.iictp.org.