SG-LegalClarify explained: can LLMs reason about Singapore law and ask when facts are missing?

23 November 2026 · 6 min read · blogs.iictp.org/sg-legalclarify-legal-reasoning-ambiguity-benchmark-explained

Short answer: SG-LegalClarify is a benchmark, posted to arXiv in November 2026, that tests whether language models can reason about Singapore law and recognise when a lay question lacks the facts needed to answer. It has 1,100 lawyer-authored questions across 10 legal domains, and in the authors' tests the best model satisfied only 69.8% of mandatory rubric items closed-book. It is a preprint, so treat the numbers as the authors' claims.

TL;DR The short version

- SG-LegalClarify is a benchmark that tests whether models can reason about Singapore law and notice when a lay question lacks key facts.
- It uses 1,100 lawyer-authored questions in 10 legal domains, most of them deliberately underspecified.
- Across 19 models the best met 69.8% of mandatory rubric items closed-book and 77.8% with web search; design legal assistants to list missing facts before concluding.

Read online, listen to it, or change the text size: https://blogs.iictp.org/sg-legalclarify-legal-reasoning-ambiguity-benchmark-explained

What is SG-LegalClarify?

The authors say Singapore combines inherited English common law with local statutes, so a model must separate local doctrine from generic common-law intuition. Lay questions are often underspecified, because the facts that decide the outcome depend on the governing rule. The benchmark tests both reasoning and ambiguity recognition.

How it works

- 1,100 lawyer-authored source questions in 10 legal domains: 340 fully specified and 760 underspecified.
- Each underspecified question has a fixed hidden-fact assignment and a fully specified counterpart, yielding 1,860 evaluation instances.
- Responses are graded with lawyer-authored IRAC rubrics (Issue, Rule, Application, Conclusion).
- A multi-turn pipeline lets the model decide whether to answer or ask for clarification, and a user simulator reveals hidden facts only when asked.

Results reported by the authors

- 19 LLMs were tested closed-book and with web search.

- Closed-book, the best model met 69.8% of mandatory rubric items (average 53.8%). With web search, 77.8% (average 63.0%).
- Models also struggled to recognise when clarification was necessary.
- The authors describe the benchmark as far from saturated.

Limitations and open questions

- This is an arXiv preprint. It has not been peer reviewed, and its numbers are the authors' own claims.
- The abstract does not name the models or give clarification-rate figures; we did not read the paper in full.
- Rubric grading by lawyers is rigorous but expensive, and the benchmark is Singapore-specific.
- A public split, model outputs and pipeline are planned for release, per the abstract.

How to use the finding, step by step

1. Design legal-assistant prompts so the model must list missing material facts before it concludes.
2. Grade answers on issue, rule, application and conclusion, not only the conclusion.
3. Test with jurisdiction-specific questions, not generic common-law ones.
4. Keep a lawyer in the loop for advice.

What it means for legal technologists

Asking a good clarifying question is a capability that general benchmarks underweight. For retrieval and grounding context, see RAG explained and how to evaluate an LLM application.

Frequently asked questions

- **What does the benchmark add?** It tests ambiguity recognition in Singapore law, with hidden facts and clarification turns.
- **How was the best model's score reported?** 69.8% of mandatory rubric items closed-book and 77.8% with web search.
- **Is it peer reviewed?** No. It is an arXiv preprint.

Related reading

- RAG explained
- How to evaluate an LLM application
- What is LLM hallucination?
- AI Practitioner

Sources and currency

Sources were read on 23 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

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

Sources and further reading

1. Ziwei Gu, Michael S. Lee, Drew Rein, Yash Maurya, Vikneswaran Kumaran. SG-LegalClarify: A Large-Scale Benchmark for Legal Reasoning and Ambiguity Recognition in Singapore Law. arXiv 2611.19643, 14 November 2026. arxiv.org/abs/2611.19643. Preprint.

Published by IICTP, an independent technical credential body, on 23 November 2026. The current version of this post is at https://blogs.iictp.org/sg-legalclarify-legal-reasoning-ambiguity-benchmark-explained. More guides: blogs.iictp.org.