SOVEREIGN: How well do SOTA legal reasoning models support abductive reasoning?

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

We examine how well the state-of-the-art (SOTA) models used in legal reasoning support abductive reasoning tasks. Abductive reasoning is a form of logical inference in which a hypothesis is formulated from a set of observations, and that hypothesis is used to explain the observations. The ability to formulate such hypotheses is important for lawyers and legal scholars as it helps them articulate logical arguments, interpret laws, and develop legal theories. Our motivation is to consider the belief that deep learning models, especially large language models (LLMs), will soon replace lawyers bec

1 Introduction

Analysis of: How well do SOTA legal reasoning models support abductive reasoning?. Research goal: How does the robustness of o1-preview and DeepSeek-R1 to adversarial token-level perturbations (e.g., synonym substitution) vary with model size and inference budget, measured by accuracy drop and inference throughput on legal reasoning benchmarks?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

10 papers retrieved. 5 claims extracted, 5 verified. Tribunal: 7.3/10 $\rightarrow$ REVISE (revision_round=1). Policy: SOFT_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
SOTA legal reasoning models fall short in supporting abductive reasoning tasks.	✓	0.35
A logic-augmented dataset for abductive reasoning with 498,697 samples was built.	✓	0.26
Abductive reasoning is a form of logical inference in which a hypothesis is formulated from a set of observations, and t	✓	0.39
The ability to formulate hypotheses is important for lawyers and legal scholars as it helps them articulate logical argu	✓	0.36
Deep learning models, especially large language models (LLMs), will soon replace lawyers because they perform well on ta	✓	0.42

References

- http://arxiv.org/abs/2310.05276v1
- http://arxiv.org/abs/2304.06912v2
- http://arxiv.org/abs/2505.03970v1