SOVEREIGN: Can ASE framework maintain consistent accuracy scores while scaling inference budget across diverse legal doma

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

Large language models (LLMs) have shown remarkable skills across various activities, including text generation and code synthesis. Their widespread applicability, however, raises substantial concerns about security, privacy, and possibly misuse. Of recent legislative efforts, the most notable is the proposed EU AI Act, which classifies specific AI applications as high-risk. For detailed regulatory guidance, also refer to the GDPR and HIPAA privacy rules. This paper introduces SecureLLM, a novel framework that integrates lightweight cryptographic protocols, decentralized fine-tuning strategies,

1 Introduction

Analysis of: SecureLLM: A Unified Framework for Privacy-Focused Large Language Models. Research goal: Can ASE framework maintain consistent accuracy scores while scaling inference budget across diverse legal domain benchmarks compared to baseline adversarial training approaches?.

2 Methodology

Multi-query arXiv search (1 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

3 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 5.7/10 $\rightarrow$ REVISE (revision_round=1). Policy: ESCALATE_TO_OWNER.

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.

References

- https://doi.org/10.1145/3546577
- https://doi.org/10.3390/a18070443
- https://doi.org/10.3390/app15084180