



AI for Legal Domain Identification and Guidance in Sri Lankan Civil Law: A Comparative Study of Open-Source vs Proprietary AI Models

Athulathmudali A. M.
(MS23475540)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2025

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Dr. Jeewanee Bamunusinghe

Approved for MSc. Research Project:



MSc in IT Programme Co-ordinator, SLIIT

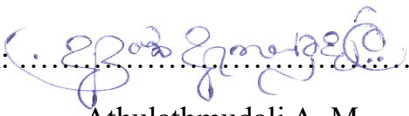
Approved for MSc:



Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: .....
Athulathmudali A. M.

Date: 03.01.2026.....

ABSTRACT

AI for Legal Domain Identification and Guidance in Sri Lankan Civil Law: A Comparative Study of Open-Source vs Proprietary AI Models

Athulathmudali A. M.

MSc. in Information Technology

Supervisor: Dr. Jeewanee Bamunusinghe

December 2025

This study investigated the application of retrieval-augmented generation (RAG) architectures powered by large language models (LLMs) to improve access to civil-law information in Sri Lanka. It addresses a key challenge in the country's justice system, the limited accessibility to affordable and reliable legal guidance. A RAG-based legal information assistant was designed, implemented, and evaluated using two back-end models: OpenAI's GPT-3.5-Turbo and the open-source Mistral-7B-v0.1. Both systems were integrated into a curated Sri Lankan civil-law corpus and compared across three metrics: accuracy, latency, and cost using a set of test queries. GPT-3.5 Turbo achieved higher accuracy (92.5%) and lower average latency (4.17s) at a lower cost (USD 0.000487 per query) than Mistral-7B-v0.1 (82.5% accuracy, 15.64s average latency, USD 0.000742 average cost). Statistical tests confirmed significant differences in latency and cost. GPT-3.5-Turbo therefore exhibited superior responsiveness and efficiency for real-time, citizen-facing legal assistance, whereas Mistral-7B offers a competitive, viable, privacy-preserving alternative for institutional or offline use. The research contributes a reproducible evaluation framework for legal-domain LLMs and a localized civil-law corpus designed for retrieval-augmented systems. More broadly, it demonstrates that responsibly designed AI can enhance access to justice in low-resource contexts. The findings establish a foundation for future multilingual, ethically aligned, and jurisdiction-aware legal-AI systems in Sri Lanka.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my family for their unwavering love, encouragement, and belief in me. Their continuous support has been my greatest source of strength, keeping me focused and motivated throughout this journey.

My deepest appreciation also goes to my supervisor, Dr. Jeewanee Bamunusinghe, for her invaluable guidance, patience, and encouragement. Her insightful feedback and expertise have played a vital role in shaping this research and helping me navigate every challenge along the way.

I would also like to extend my sincere thanks to the Sri Lanka Institute of Information Technology for providing me with the resources, facilities, and academic environment necessary to carry out this study successfully.

Finally, I dedicate this work to my future self as a reminder that with determination, perseverance, and self-belief, every goal is within reach and every effort contributes to growth.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS.....	v
List of Figures	viii
List of Equations	ix
List of Tables	x
Chapter 1 Introduction	1
1.1 Background.....	1
1.2 Research Problem	3
1.3 Objectives	4
1.4 Significance of Study	5
1.5 Proposed Solution	6
Chapter 2 Literature Review	8
2.1 Legal AI Systems.....	8
2.1.1 Types and Examples of Legal Chatbot Systems	8
2.1.2 Core Limitations and Risks.....	11
2.2 RAG Architectures: Retrieval-Augmented Generation	13
2.3 LLM Evaluation Methods: Measure Legal Accuracy.....	16
2.4 Open-source vs Proprietary LLMs: Performance Trade-offs in Specialized Domains.....	18
2.4.1 Accuracy and Task Complexity	20
2.4.2 Model Size and Scaling.....	20
2.4.3 Fine-Tuning vs. Prompting and Specialization.....	20
2.4.4 Cost and Resource Requirements.....	21
2.4.5 Privacy, Control, Legal and Ethical Risk.....	21
2.4.6 Latency, Infrastructure, and Maintenance.....	21
2.4.7 Generalization vs. Domain Accuracy	21
2.5 Sri Lankan Legal Informatics	22
2.6 Research Gap Identification.....	24
Chapter 3 Research Design and Methodology.....	26
3.1 Dataset Construction.....	26
3.1.1 Rationale for Expert-Curated Records	26
3.1.2 Design of the Expert-Curated Dataset.....	27
3.1.3 Integration with the System Architecture	28

3.1.4 Sample Size and Trade-Offs	28
3.1.5 Ethical and Legal Considerations	29
3.2 System Architecture: RAG Pipeline Design	29
3.3 Model Selection and Rationale	31
3.4 Evaluation Framework	33
3.5 Experimental Setup	34
Chapter 4 Implementation	35
4.1 Data preprocessing pipeline	35
4.1.1 Corpus Creation and Document Definition	35
4.1.2 Text Cleaning and Normalization	35
4.1.3 Segmentation and Chunking	36
4.1.4 Document Embedding Preparation	37
4.1.5 Metadata Annotation and Provenance	37
4.2 Retrieval system implementation	37
4.2.1 Similarity-Search Configuration	38
4.2.2 Retrieval-Generation Integration	38
4.2.3 Performance and Operational Factors	39
4.2.4 Traceability and Provenance	39
4.2.5 Prompt Engineering and Context Construction	39
4.3 Integration with Multiple LLM Backends	41
4.3.1 Model Selection and Rationale	42
4.3.2 Implementation of GPT-3.5-Turbo Backend	42
4.3.3 Implementation of Mistral-7B-v0.1 Backend	43
4.3.4 Evaluation Support and Logging	44
4.4 Testing Infrastructure	44
Chapter 5 Evaluation and Results	46
5.1 Comparative Analysis	46
5.1.1 Accuracy	47
5.1.2 Latency	48
5.1.3 Cost	50
5.1.4 Error Analysis and Failure Cases	53
5.1.5 Summary of Trade-Offs	54
5.1.6 Implications for Legal-AI Deployment	57
5.2 Statistical Significance Testing	58
5.3 Cost-Benefit Analysis	61
Chapter 6 Discussion	63

6.1 Implications for legal practice and access to justice	63
6.2 Limitations	66
6.3 Ethical considerations	68
Chapter 7 Conclusion.....	70
7.1 Summary of Findings.....	70
7.2 Contributions.....	72
7.3 Future Research Directions	74
Bibliography	77
Appendix.....	83
Appendix 1: Abbreviations	83
Appendix 2: Sample Test Queries and Model Responses	84
Appendix 3: External Expert Declarations	87

List of Figures

Figure 2.1 ChatLaw architecture.....	10
Figure 3.1 System architecture of the RAG pipeline for the proposed AI Legal Assistant.....	30
Figure 4.1 Canonical RAG pattern	39
Figure 4.2 RAG pipeline structure used across LLM backends	43
Figure 5.1 T4 GPU Memory	46
Figure 5.2 L4 GPU Memory	46
Figure 5.3 A100 GPU Memory	47
Figure 5.4 GPT-3.5-Turbo average latency	48
Figure 5.5 T4 GPU usage rate.....	49
Figure 5.6 Mistral-7B-v0.1 average latency on T4 GPU	49
Figure 5.7 Mistral-7B-v0.1 average latency on A100 GPU.....	49
Figure 5.8 A100 GPU usage rate	49
Figure 5.9 Mistral-7B-v0.1 average latency on L4 GPU	50
Figure 5.10 L4 GPU usage rate.....	50
Figure 5.11 GPT-3.5-Turbo average cost	51
Figure 5.12 Mistral-7B-v0.1 average cost on L4 GPU	52
Figure 5.13 Mistral-7B-v0.1 average cost on T4 GPU	52
Figure 5.14 Mistral-7B-v0.1 average cost on A100 GPU.....	52
Figure 5.15 Trade-off between cost and latency for GPT3.5-Turbo.....	55
Figure 5.16 Trade-off between cost and latency for Mistral-7B-v0.1	56
Figure 5.17 Significance test.....	59
Figure 5.18 Confidence Interval	59
Figure 5.19 Latency T-Test.....	60
Figure 5.20 Cost T-Test	60

List of Equations

Equation 1 Cost per query.....	51
--------------------------------	----

List of Tables

Table 2.1 Overview of legal AI and chatbot systems by architecture, jurisdiction, and performance.....	10
Table 2.2 Core components of a RAG architecture and relevance in legal applications.....	15
Table 2.3 Key Trade-offs between proprietary and open-source LLMs.....	19
Table 5.1 Model accuracy on test queries.....	47
Table 5.2 Measured average latency per query.....	48
Table 5.3 Cost Estimates.....	50