

LEXIS - A LEGAL ANALYSIS SYSTEM USING AI

Presented by:

Kushal Adhyaru
Brendan Chao
Mayuka Kothuru
Albert Ong
Prem Shah

Project Advisors:

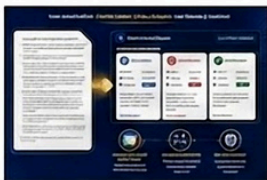
Dr. Simon Shim
Dr. Lee C. Chang

Introduction

Legal contract review is manual, slow, and costly. Lawyers spend hours identifying obligations, flagging risks, and verifying compliance across diverse legal documents. Standard LLMs fail to reliably distinguish between obligations, prohibitions, and permissions without domain-specific fine-tuning. Commercial platforms are proprietary and opaque.

LEXIS addresses this gap with an open, end-to-end legal intelligence platform. For external stakeholders, it extracts obligations, assesses risk, notifies similar provisions, and integrates directly into legal workflows through Google Calendar and Gmail.

Built on the CUAD dataset with 510 contracts and 41 clause types, LEXIS fine-tunes two open-weight LLMs using QLoRA and uses a six-step analysis pipeline that transforms raw contract text into workflows, queryable obligation results.



Methodology

Six-Step Pipeline

- Step 1: Span Extraction using ModernBERT with LegalBERT tokenizer.
 - Step 2: Deontic Classification into OBLIGATION, PROHIBITION, or PERMISSION.
 - Step 3: Structured JSON Connection extracting action, owner, deadline, condition, consequence, and high level.
 - Step 4: Risk Assessment scoring HIGH, MEDIUM, or LOW.
 - Step 5: Semantic Search using Jaccard similarity against stored clauses.
 - Step 6: Calendar Integration mapping deadlines to compliance calendar.
- All services include in-context and regex fallbacks ensuring results are returned even without GPU access.



Analysis and Results

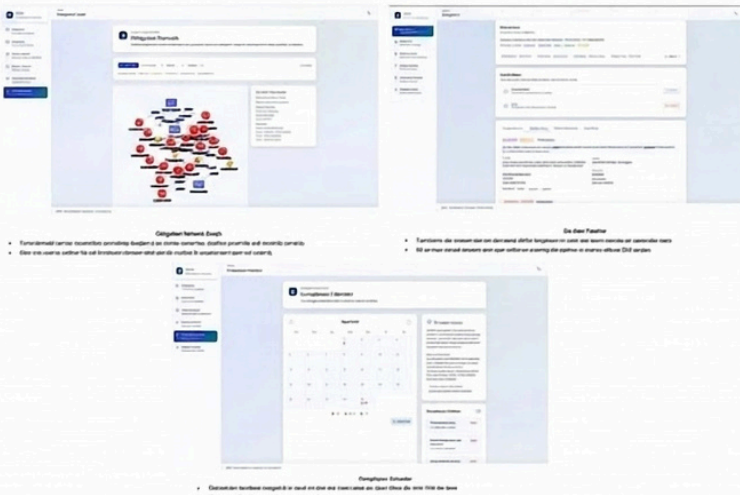


Model Performances

- Phi-4 and Qwen2.5-7B achieved the strongest post-fine-tuning on O/U Deontic Accuracy and <28.9% in ROUGE-L.
- QLoRA fine-tuning on CU Deontic Accuracy and <28.9% in ROUGE-L.
- All models reached near perfect Field Completion (87.8-100%), confirming consistent structured JSON output.

Baseline vs Fine Tune Model

Model	Generation	ROUGE-L	F1 Score
Phi-4	0.87	0.87	0.87
Qwen2.5-7B	0.87	0.87	0.87
Qwen2.5-72B	0.87	0.87	0.87
Phi-4	0.87	0.87	0.87
Qwen2.5-72B	0.87	0.87	0.87



Key Features

- Six-step automated obligation extraction pipeline processing PDF, DOCX, RTF, and TXT contracts in under one second with 84% deontic classification accuracy.
- Interactive force-directed obligation network graph and compliance timeline visualizing obligations and deadlines coded by risk level across the full contract portfolio.
- 175 passing tests covering unit, integration, and ton real-world contract scenarios including NDA, SaaS, Employment, GDPR, and Mergers and Acquisitions.
- Acquisitions.

Summary/Conclusions

LEXIS proves that open-weight fine-tuned LLMs can deliver accurate, structured legal analysis without proprietary systems. QLoRA fine-tuning produced a 23-point deontic accuracy improvement and 100% 260% validity across all models. The six-step pipeline with graceful degradation ensures reliable performance in both development and production environments.

LEXIS work includes multi-language support, human-in-the-loop feedback, knowledge distillation for smaller deployment, and real-time error monitoring via workflows.

Key References

- [1] Hendrycks et al. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. 2021.
- [2] Tugener et al. LEDGAR: A Large-Scale Multi-Label Corpus for Legal Provisions. LREC 2020.
- [3] Chalkidis et al. LevGLUE: A Benchmark for Legal Language Understanding. ACL 2022.
- [4] Chalkidis et al. LEGAL-BERT: The Muppets Straight Out of Law School. 2020.
- [5] Sino et al. Exploring LLMs Applications in Law. IEES Access 2020.
- [6] Shu et al. LawLLM: Law Large Language Model for the US Legal System. CIKM 2021.

Acknowledgements

We are grateful to the Applied Data Science Department for the resources and academic support that made the development of the LEXIS project possible.

GitHub: github.com/mayuka-reddy/298B

Interested in partnering with us on a capstone project or exploring other ways to collaborate with the College of IDS? Contact CIDS_CAP@sjsu.edu to discuss potential projects and partnership opportunities.