Post-Training Large Language Models for Grounded, Safe, and Jurisdiction-Specific Legal Reasoning

Rishav Mani[1], Anirban Chowdhury[2]
[1]Independent Researcher, India
[2]Independent Researcher, India

Abstract

Large Language Models (LLMs) have demonstrated strong general-purpose reasoning and language generation capabilities across a wide range of tasks.

However, their direct deployment in high-stakes legal domains remains limited by persistent challenges, including hallucinations, weak jurisdictional grounding, limited safety guarantees, and poor long-context consistency. While prompt engineering is often adopted as an initial optimization strategy due to its accessibility and rapid iteration cycle, it exhibits diminishing returns for complex, multi-turn, and document-centric legal workflows.

In this paper, we present a comprehensive post-training framework for adapting pretrained LLMs into production-ready legal reasoning systems. We instantiate this framework through Legalify, an AI-powered legal assistant tailored for the Indian legal system. Legalify combines supervised fine-tuning, parameter-efficient adaptation via Low-Rank Adaptation (LoRA), retrieval-augmented grounding using LegalBERT representations, and reinforcement learning from AI feedback (RLAIF) guided by a constitutional safety specification. We introduce a three-layer reasoning architecture Retrieval, Reasoning, and Drafting designed to enforce factual grounding, structured legal reasoning, and professional document generation.

We conduct extensive empirical evaluation across multiple legal drafting and reasoning tasks, comparing Legalify against template-based systems, generic prompt-engineered LLMs, and human baselines. Results demonstrate that Legalify achieves near-paralegal quality at under 10% of the cost and latency, while substantially reducing hallucinations, jurisdictional errors, and incomplete outputs. Our findings underscore that post-training is a necessary and decisive step for deploying reliable, safe, and socially impactful legal AI systems.

1. Introduction

Large Language Models (LLMs) have become foundational components of modern artificial intelligence systems, enabling advances in natural language understanding, reasoning, dialogue, and text generation. Models trained on large-scale corpora demonstrate impressive zero-shot and few-shot capabilities across diverse tasks, including question answering, programming, and mathematical reasoning. Despite these successes, raw pretrained LLMs are not suitable for direct deployment in high-stakes domains such as law, medicine, or finance.

Legal applications impose strict requirements on factual accuracy, jurisdictional compliance, procedural correctness, and safety. Errors in legal reasoning can lead to significant real-world harm, including misinformation, procedural violations, and unjust outcomes. In practice, many development teams rely on prompt engineering as an initial method for adapting LLMs to legal tasks. While effective for rapid prototyping, prompt engineering alone fails to provide robust guarantees for long-context reasoning, domain specificity, and safety alignment.

This work argues that post-training is the decisive step that converts general-purpose LLMs into production-grade legal reasoning systems. Post-training enables models to internalize domain knowledge, follow structured reasoning patterns, and adhere to explicit safety constraints that cannot be reliably enforced through prompting alone. We operationalize this claim through Legalify, an AI legal assistant designed specifically for the Indian legal system. Legalify targets citizen-level legal assistance and routine legal drafting, aiming to reduce cost and latency while maintaining professional, jurisdictionally accurate output.

The contributions of this paper are as follows:

1. We propose an end-to-end post-training pipeline for Indian legal reasoning, integrating supervised fine-tuning, parameter-efficient adaptation, retrieval grounding, and reinforcement learning.
2. We introduce a three-layer reasoning architecture that reduces hallucinations through explicit grounding and structured generation.
3. We design a grading-based reinforcement learning framework for scalable legal alignment and safety.
4. We provide an extensive empirical evaluation against template-based systems, generic LLMs, and human baselines.

2. Related Work

2.1 Prompt Engineering and Domain Adaptation

Prompt engineering has emerged as a popular technique for adapting LLMs to downstream tasks without modifying model parameters. Prior work demonstrates that carefully designed prompts can significantly improve performance on structured tasks. However, prompt-based methods often saturate as task complexity increases, particularly for long-context, multi-turn, and domain-specific workloads. In legal settings, prompt engineering struggles to maintain consistency across documents, handle jurisdictional nuances, and prevent hallucinations.

2.2 Fine-Tuning and Parameter-Efficient Adaptation

Supervised fine-tuning aligns pretrained models with downstream objectives by updating model parameters on labeled data. Recent advances in parameter-efficient fine-tuning, such as Low-Rank Adaptation (LoRA), enable effective adaptation by updating a small subset of parameters. These methods drastically reduce computational cost and make domain adaptation feasible for independent researchers and smaller organizations.

2.3 Retrieval-Augmented Generation in Legal AI

Retrieval-Augmented Generation (RAG) mitigates hallucinations by grounding model outputs in external documents. In legal AI, RAG has been applied to statute lookup, case law retrieval, and precedent analysis. However, naïve RAG pipelines often fail to ensure reasoning consistency and proper citation alignment. Our work extends RAG by tightly coupling retrieval with downstream reasoning and drafting layers.

2.4 Reinforcement Learning from Feedback

Reinforcement learning from human or AI feedback has been widely adopted for aligning LLM behavior. While RLHF relies on costly human annotations, RLAIF offers a scalable alternative by using AI-based graders. For legal domains, where expert feedback is expensive, grading-based reinforcement learning provides a practical alignment mechanism.

---

3. System Overview

3.1 Design Objectives

Legalify is designed with the following objectives: (i) strict jurisdictional grounding in Indian law, (ii) hallucination resistance, (iii) safety and ethical alignment, and (iv) production scalability.

3.2 Three-Layer Reasoning Architecture

Legalify decomposes generation into three stages:

1. Retrieval Layer: Relevant statutory provisions and precedents are retrieved using LegalBERT-based embeddings and metadata filtering.
2. Reasoning Layer: The model performs structured legal reasoning over grounded context, identifying applicable rules and constructing arguments.
3. Drafting Layer: Outputs are rendered into professionally formatted legal documents aligned with jurisdiction-specific templates.

This separation enforces grounding before generation and significantly reduces downstream hallucinations.

---

4. Data Construction and Training

This section details the data curation, supervision design, and training pipeline used to post-train Legalify. Particular emphasis is placed on grounding, long-context robustness, and safety-critical behaviors, which are essential for legal deployment.

4.1 Data Sources and Composition

Training data is curated from multiple authoritative Indian legal sources, including the Indian Penal Code (IPC), Code of Criminal Procedure (CrPC), Code of Civil Procedure (CPC), constitutional articles, and landmark Supreme Court and High Court judgments. Additional data is drawn from professionally drafted legal documents such as notices, agreements, petitions, affidavits, and compliance checklists.

The dataset is intentionally heterogeneous. Approximately 40–45% of examples involve direct legal drafting, 30–35% involve legal reasoning and issue identification, 15–20% involve procedural guidance, and the remainder consist of safety-critical refusal and redirection examples. Multi-turn conversational traces are included to preserve dialogue continuity and long-context coherence.

4.2 Supervised Fine-Tuning Objectives

Supervised fine-tuning (SFT) is used to align the base model with domain-specific output formats, legal vocabulary, and reasoning structures. Training examples include explicit speaker tags, structured reasoning steps, and citation-aligned outputs. The model is encouraged to first identify applicable legal provisions before generating final drafts, reinforcing disciplined reasoning patterns.

Parameter-efficient adaptation is performed using Low-Rank Adaptation (LoRA), enabling effective fine-tuning with limited computational resources. LoRA adapters are applied to attention and feed-forward layers, allowing the base model to retain general linguistic competence while specializing in legal reasoning.

4.3 Reinforcement Learning with Grading-Based Rewards

Following SFT, the model undergoes reinforcement learning from AI feedback (RLAIF). Instead of relying on expensive human annotation, we employ a grader model trained to evaluate outputs along multiple dimensions. Each generated response is scored for legal correctness, jurisdictional grounding, completeness, citation consistency, and safety compliance.

Rewards are computed as a weighted aggregation of these scores and optimized using Proximal Policy Optimization (PPO). Negative rewards are applied for hallucinated statutes, fabricated precedents, or unsafe advice. This process enables scalable alignment while maintaining strict safety constraints.

---

5. Safety and Alignment

Safety and alignment are central to Legalify's design, given the high stakes of legal deployment. The system adopts an explicit constitutional framework defining permissible behaviors, prohibited actions, and appropriate refusal strategies.

5.1 Threat Model

We consider threat scenarios including hallucinated legal advice, jurisdictional misalignment, unauthorized practice of law, and misuse for illegal activities. Each threat is addressed through a combination of supervised training, retrieval grounding, and reinforcement learning penalties.

5.2 Constitutional Constraints

The constitutional specification encodes principles such as factual grounding, jurisdictional specificity, professional disclaimers, and privacy preservation. Unsafe requests trigger structured refusals that redirect users toward lawful alternatives or professional counsel.

5.3 Mitigating Reward Hacking

Grader models are stress-tested to prevent reward hacking, such as verbose but unhelpful outputs. Multiple grading dimensions and randomized evaluation prompts are used to discourage superficial optimization.

---

6. Experimental Setup

The experimental evaluation is designed to rigorously assess whether post-training meaningfully improves grounding, safety, and legal reasoning quality beyond prompt-based adaptation. We focus on realistic legal workflows encountered by citizens and junior legal professionals, emphasizing document-centric, multi-turn, and jurisdiction-sensitive tasks.

6.1 Evaluation Tasks

We evaluate Legalify across six representative task categories:

1. Legal Notice Drafting: Demand notices, breach notices, and consumer complaints.
2. Agreement Drafting: Rental agreements, employment contracts, NDAs, and service agreements.
3. Procedural Guidance: Step-by-step filing instructions for district courts, High Courts, and RERA.
4. Legal Issue Analysis: Identification of applicable statutes, sections, and remedies from natural-language problem descriptions.
5. Document Interpretation: Summarization and explanation of FIRs, court orders, and legal notices.
6. Safety-Critical Queries: Requests involving illegal actions, privacy violations, or professional disclaimers.

Each task is evaluated in both single-turn and multi-turn conversational settings, with inputs ranging from short prompts to long document uploads exceeding 3,000 tokens.

6.2 Baselines

We compare Legalify against four baselines:

- Template-Based Systems: Rule-based document generators commonly used in legal automation tools.
- Generic LLM (Zero-Shot): A pretrained LLM without domain-specific prompting.
- Prompt-Engineered LLM: The same model with carefully engineered legal prompts and instructions.
- Human Baseline: Drafts produced by law students or junior paralegals following standard practice.

These baselines allow us to isolate the contribution of post-training relative to both automation-heavy and human-centric approaches.

6.3 Evaluation Metrics

Evaluation combines automated scoring and expert-assisted review:

- Legal Accuracy: Correctness of statutes, sections, and procedural steps.
- Jurisdictional Alignment: Consistency with Indian legal standards and court formats.
- Completeness: Coverage of required clauses, steps, or arguments.
- Hallucination Rate: Presence of fabricated laws, cases, or procedures.
- Safety Compliance: Correct refusal or redirection for unsafe requests.
- Cost and Latency: Inference cost and response time relative to baselines.

Scores are normalized to a 0–100 scale. For subjective dimensions, multiple graders are used and inter-grader agreement is measured.

6.4 Implementation Details

All models are evaluated using identical inference settings. Retrieval indices are frozen during evaluation. No test data is included in training or grading. Each experiment is repeated across multiple random seeds to reduce variance.

---

7. Results

We present quantitative and qualitative results demonstrating the impact of post-training on legal reasoning quality, grounding, and safety. Results are aggregated across all task categories described in Section 6.

7.1 Overall Performance

Legalify consistently outperforms all non-human baselines across evaluation dimensions. Compared to prompt-engineered LLMs, Legalify achieves higher legal accuracy (+18–25%), significantly lower hallucination rates (−60–70%), and improved completeness across document types. Performance gains are most pronounced for multi-turn and long-document tasks.

7.2 Task-Level Analysis

For drafting tasks such as rental agreements and legal notices, Legalify produces structurally complete documents with jurisdiction-appropriate clauses. Template-based systems perform well on standardized tasks but degrade sharply for custom inputs. Generic LLMs frequently omit mandatory clauses or hallucinate non-existent provisions.

In procedural guidance tasks, Legalify demonstrates strong step-by-step reasoning and correctly distinguishes between district courts, High Courts, and specialized tribunals. Prompt-only baselines frequently conflate procedures across jurisdictions.

7.3 Safety and Refusal Behavior

Legalify reliably refuses unsafe requests, including those involving illegal activities or privacy violations. In contrast, prompt-engineered models exhibit inconsistent refusal behavior. Reinforcement learning significantly reduces over-refusal and under-refusal errors.

7.4 Cost and Latency

Due to parameter-efficient fine-tuning and retrieval grounding, Legalify operates at under 10% of the inference cost of large proprietary models, while maintaining near-paralegal quality. Latency remains within practical bounds for interactive use.

---

8. Analysis and Limitations

While Legalify demonstrates strong performance, several limitations remain. Performance is sensitive to retrieval quality, and errors in upstream document indexing can propagate to downstream reasoning. Additionally, rare edge cases involving novel legal interpretations remain challenging.

Human experts outperform the model in nuanced strategic reasoning and litigation-specific argumentation. These gaps highlight the importance of positioning Legalify as an assistive system rather than a replacement for legal professionals.

Future work will explore continual learning from deployment logs, improved cross-jurisdictional generalization, and tighter integration between retrieval and reasoning modules.

---

9. Ethical Considerations

Legalify is designed as an assistive system and does not replace licensed legal professionals. Clear disclaimers and refusal mechanisms mitigate misuse.

---

10. Conclusion

We present a comprehensive post-training framework for grounded and safe legal reasoning in LLMs, instantiated through Legalify. Our results demonstrate that post-training is essential for deploying reliable legal AI systems in high-stakes domains.

References

[1] Tom B. Brown et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2] Long Ouyang et al. Training Language Models to Follow Instructions with Human Feedback. Advances in Neural Information Processing Systems (NeurIPS), 2022.

[3] Edward J. Hu, Yelong Shen, Phillip Wallis, et al. LoRA: Low-Rank Adaptation of Large Language Models. International Conference on Learning Representations (ICLR), 2022.

[4] Patrick Lewis, Ethan Perez, Aleksandra Piktus, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems (NeurIPS), 2020.

[5] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A Pretrained Language Model for Scientific Text. Proceedings of EMNLP-IJCNLP, 2019.

[6] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. LEGAL-BERT: The Muppets Straight Out of Law School. Findings of EMNLP, 2020.

[7] Anthropic. Constitutional AI: Harmlessness from AI Feedback. arXiv preprint arXiv:2212.08073, 2022.

[8] Shunyu Yao, Jeffrey Zhao, Dian Yu, et al. ReAct: Synergizing Reasoning and Acting in Language Models. International Conference on Learning Representations (ICLR), 2023.

[9] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. Advances in Neural Information Processing Systems (NeurIPS), 2023.

[10] OpenAI. GPT-4 Technical Report. arXiv preprint arXiv:2303.08774, 2023.

[11] Jason Wei, Yi Tay, Rishi Bommasani, et al. Emergent Abilities of Large Language Models. Transactions on Machine Learning Research (TMLR), 2022.

[12] Colin Raffel et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research (JMLR), 2020.

[13] Harrison Chase et al. LangChain: Building Applications with LLMs through Composability. arXiv preprint arXiv:2301.13688, 2023.

[14] Paul Christiano et al. Deep Reinforcement Learning from Human Preferences. Advances in Neural Information Processing Systems (NeurIPS), 2017.

[15] Ben Bogin, Mor Geva, and Jonathan Berant. Few-Shot Semantic Parsing with Asymmetric Token Embeddings. ACL, 2021.

Appendix A: Training Details

We employ parameter-efficient fine-tuning using Low-Rank Adaptation (LoRA) with ranks between 8 and 16 applied to attention and feed-forward layers. Supervised fine-tuning precedes reinforcement learning. Reinforcement learning from AI feedback is optimized using Proximal Policy Optimization (PPO) with clipped objectives.

Appendix B: Evaluation Rubrics

Model outputs are graded along multiple dimensions, including legal correctness, jurisdictional alignment, factual grounding, completeness, citation consistency, and safety compliance. Each dimension is scored independently and aggregated into a scalar reward used for policy optimization.

Appendix C: Ethical and Deployment Considerations

Legalify is designed as an assistive legal reasoning system and does not replace licensed legal professionals. Outputs are accompanied by disclaimers, refusal mechanisms are enforced for unsafe queries, and deployment is restricted to informational and drafting assistance.