TPM Vol. 32, No. S8, 2025
ISSN: 1972-6325
https://www.tpmap.org/

GOAL-BASED AGENTIC AI FOR AUTOMATING CLAIM VALIDATION USING LARGE LANGUAGE MODELS

NAMRATA SHETTY[1], SANJAY BANG[2], QUANG-VINH DANG[3], SWATI S KAMTHEKAR[4], ANSHUL SAXENA[5], SUNIL VAKAYIL PHD[6]

[1,2]CHRIST UNIVERSITY, INDIA
[3]SCHOOL OF INNOVATIVE AND COMPUTING SCIENCE, BRITISH UNIVERSITY VIETNAM, HUNG YEN, VIETNAM
[4]ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTER ENGINEERING AND DATA SCIENCE, MGM'S COLLEGE OF ENGINEERING AND TECHNOLOGY, MUMBAI, INDIA
[5]SYMBIOSIS CENTRE FOR INFORMATION TECHNOLOGY, PUNE, INDIA
[6]LOYOLA INSTITUTE OF BUSINESS ADMINISTRATION, CHENNAI, INDIA

ABSTRACT— This paper presents a goal-based agentic AI system for automating individual forest claim validation under India's Forest Rights Act (FRA). The system integrates legal rule enforcement through formal logic gates with interpretive reasoning by a large language model (LLM). Each claim is evaluated across statutory conditions such as identity, occupancy date, land extent, and evidence sufficiency. A structured prompting and tool-augmented workflow enables the LLM to reason through unstructured data while maintaining compliance with FRA criteria. In simulation using 1,000 synthetic claims, the proposed agent achieved 92.4% accuracy—outperforming baseline rule systems and zero-shot LLM classifiers. The agent's decisions are explainable, auditable, and adaptable, demonstrating how hybrid AI architectures can enhance legal workflows. The model's modularity supports future extensions to community claims and other entitlement domains. This work contributes to the emerging field of AI-assisted legal adjudication, particularly in public governance contexts involving high-stakes land rights.

Keywords—Legal AI, LLM Agents, Rule-based Reasoning, , Claim Validation, Neuro-symbolic AI, Explainable AI

1. INTRODUCTION

Automating legal decision-making presents unique challenges that differ from typical AI applications. Legal adjudication often involves high-stakes consequences, strict procedural requirements, and the need for both interpretability and compliance with codified rules. Unlike standard classification tasks, legal validation requires reasoning over incomplete, ambiguous, or unstructured inputs while applying deterministic statutory criteria. Traditional rule-based systems can enforce hard constraints but struggle with linguistic variability and edge cases. In contrast, large language models (LLMs) excel at understanding free-text inputs but lack guaranteed rule compliance, making them unreliable in domains where legal correctness is non-negotiable. Bridging this gap requires architectures that are both semantically flexible and legally rigid—systems that can reason like humans while never violating the law.

This paper introduces a goal-driven agentic AI system for automating forest claim validation. At its core is a hybrid architecture combining a large language model (LLM) with a modular logic gate framework. Each logic gate represents a statutory condition encoded as a deterministic Boolean predicate (e.g., "occupancy before 13 Dec 2005"). Claims are parsed and evaluated sequentially, and only if all gates return True is the claim approved. This enforces strict legal compliance while enabling interpretive reasoning over noisy or unstructured inputs.

Technically, our system incorporates prompt engineering, soft prompt tuning, external tool invocation (e.g., geospatial APIs), and chain-of-thought (CoT) reasoning to guide the LLM's behavior. The architecture follows a rational agent model: it perceives input data, chooses reasoning or lookup actions, and decides based on a utility-maximizing goal (i.e., correct claim adjudication). We simulate this pipeline on a dataset of 1,000 synthesized claims and benchmark it against three baselines: a rule-based engine, a zero-shot LLM classifier, and a CoT-only LLM. Results show the proposed agent achieves state-of-the-art performance in accuracy, precision, and explainability.

Our work contributes to the emerging field of legal automation by demonstrating how neuro-symbolic AI can be applied to high-stakes, rule-bound public governance tasks such as land entitlement adjudication.

2. RELATED WORK

Table I compares six recent works spanning prompt engineering, agent-based AI architectures, and legal AI applications, highlighting their models, benefits, and limitations.

Work & Model	Approach	Benefits	Limitations
ReAct (Yao et al., 2023)	LLM + Tool Use	Reduces hallucination, improves reasoning	Needs APIs, slow inference
HuggingGPT (Shen et al., 2023)	LLM Controller + Expert Models	Multi-modal, task-flexible	Complex setup, model dependencies
CoT Prompting (Wei et al., 2022)	Few-Shot Stepwise Prompting	Improves logical accuracy	Large model needed, prompt-sensitive
Self-Consistency (Wang et al., 2023)	Multi-Path Reasoning Vote	Boosts correctness via voting	High compute cost
GPT-4 on Bar Exam (Katz et al., 2023)	Zero-Shot Legal QA	High legal QA performance	Lacks real-case reasoning
Legal Annotation (Savelka & Ashley, 2023)	Zero-Shot Text Annotation	Strong legal text tagging	Costly, context limits

Table 1: Related Work

3. RESEARCH DESIGN

The system validates each forest rights claim through a pipeline of rule-based decision points ("logic gates"), each corresponding to a statutory condition under the Forest Rights Act (FRA). An LLM-based agent parses claims, extracts attributes, and evaluates compliance. The decision function is:

$$f(C) = \begin{cases} 1, & \text{if } g1(C) = g2(C) = \ldots = gn(C) = 1 \\ 0, & \text{otherwise} \end{cases}$$

where **C** is a claim, **f(C)** the final decision (1 = approve, 0 = reject), and **gi(C)** the binary output of each statutory gate.

The design is informed by thematic analysis of FRA implementation. (i) A pre-validation module auto-checks completeness, addressing missing documents (Theme 1). (ii) Statutory rules are encoded as logic gates to reduce discretion and delay (Theme 2). (iii) The LLM handles narrative or vernacular descriptions while extracting statutory facts (Theme 3). (iv) Metadata from approved claims are logged for downstream welfare linkages (Theme 4).

Each legal condition is encoded as follows:

g1(C): Identity Eligibility = 1 if claimant belongs to ST or OTFD, else 0
g2(C): Occupancy Validity = 1 if land occupation evidence predates 13 Dec 2005, else 0
g3(C): Land Use Legality = 1 if area $\leq$ 4 hectares and not in protected zone, else 0
g4(C): Gram Sabha Approval = 1 if valid resolution is attached, else 0
g5(C): Evidence Sufficiency = 1 if $\geq$ 2 credible evidence sources, else 0

Final decision rule:

$f(C) = 1 \quad \text{if } g1(C) = g2(C) = g3(C) = g4(C) = g5(C) = 1$
$f(C) = 0 \quad \text{otherwise}$

Prompt optimization combines few-shot prompting with structured reasoning chains to train the LLM to mimic expert analysis. Soft prompting further enforces structured, concise outputs.The evaluation workflow includes: (i) parsing inputs (identity, area, evidence dates, location), (ii) sequential gate evaluation (e.g., tribe lists, date checks, geospatial APIs), (iii) structured decision outputs with reasons for rejection, and (iv) human-in-the-loop review, where approved claims are fast-tracked and uncertain cases are manually verified. Reviewer feedback is logged for iterative fine-tuning.

This ensures **rule-bound, auditable, and explainable** decisions. Claims are never approved unless all statutory conditions are met, while human feedback enables continuous learning and improvement.

4. RESULT AND DISCUSSION

To evaluate our proposed agentic AI model, we ran a controlled simulation on a test dataset of 1,000 forest claims. These claims were crafted to mirror real-world diversity, with a mix of eligible and ineligible cases based on FRA criteria. Importantly, some claims included ambiguous inputs (unclear community identity, missing or vague dates) to test model robustness. The dataset structure directly addressed **Theme 1 (awareness gaps)** and **Theme 2 (bureaucratic inconsistency)**, ensuring the system was evaluated under conditions seen in actual field implementations.

We compared four models: a baseline rule-based system, two GPT-4-based LLMs (zero-shot and chain-of-thought), and our proposed agentic LLM with logic gates and tool access. As shown in **Table II**, our agent significantly outperformed other variants across all evaluation metrics, particularly in accuracy and F1-score for the "approve" class.

Model	Accuracy (%)	Precision	Recall	F1-score
A. Rule-Based (no LLM)	81.3	88.5	75.2	81.3
B. LLM Zero-Shot Classifier	85	80.4	92.1	85.9
C. LLM w/ CoT Reasoning	88.7	86.3	90	88.1
D. Agentic LLM (Ours)	92.4	91.5	90.6	91

Table 2: Accuracy Result

The rule-based baseline achieved high precision (88.5%) but low recall (75.2%), reflecting its rigid approach. Many valid claims were missed because the system couldn't interpret non-standard or loosely formatted input. This connects directly to **Theme 3 (rights recognition issues)**, where rigid systems fail to accommodate legitimate but atypically expressed claims.

The LLM Zero-Shot model had strong recall (92.1%) but lower precision (80.4%), indicating a tendency to over-approve even ineligible claims—often due to persuasive narrative text rather than rule-based compliance.

Variant	Description	Accuracy (%)	Î" vs. Full
Full Agentic LLM (Model D)	All components included	92.4	0
No Soft Prompt	Use hard-coded prompt only	90.5	-1.9
No Tool Use	No external tools/API calls	90.1	-2.3
No Logic Gates (LLM CoT only)	LLM reasoning, no hard logic	88.7	-3.7
No CoT (Zero-shot + Gates)	Logic gates, no reasoning steps	89.4	-3
Single-Step (End-to-End)	One-shot decision only	86	-6.4

Table 3: Comparative Model

Chain-of-thought prompting improved this balance, pushing accuracy to 88.7%. But only the agentic model (Model D) integrated the structure of legal decision-making with flexible reasoning. It achieved 92.4% accuracy and a well-balanced F1-score of 91.0, showing strong performance in both rejecting invalid and accepting legitimate claims.

To better understand what drives this performance, we ran an **ablation study** (Table III). Removing components like soft prompts or tool usage degraded accuracy, confirming that each part contributes meaningfully. Notably, eliminating logic gates had the largest impact, proving that **formal rule checks are essential**, even when powerful language models are available.

These findings support our thesis: combining LLM capabilities with hard-coded legal logic results in a system that is both **smart and compliant**, addressing field concerns about discretion and inconsistency while preserving due process and explainability.

5. CONCEPTUAL MODEL

The proposed conceptual model (Figure 1) outlines an agentic AI workflow for validating forest land claims under the Forest Rights Act (FRA). The system integrates rule-based logic gates, large language model (LLM) reasoning, and external tools to ensure legally compliant, transparent, and explainable decision-making. The process begins with the **input layer**, where claims—consisting of applicant narratives and supporting documents such as Gram Sabha resolutions or land tax receipts—are parsed into structured features including identity, land area, occupancy period, and evidence type. A **pre-validation checkpoint** then verifies completeness by detecting missing metadata, empty fields, or absent documents, addressing procedural gaps that often hinder claim processing. The core **logic gate evaluation loop** enforces statutory rules through five binary predicates: identity eligibility, occupancy before the 13 December 2005 cutoff, land area and legality, Gram Sabha approval, and sufficiency of evidence. A claim is approved only if all gates return 1, with the decision function defined as:

$f(C) = 1 \iff g1(C) \wedge g2(C) \wedge g3(C) \wedge g4(C) \wedge g5(C) =$ \

This stage combines deterministic checks (e.g., GIS or database lookups) with LLM reasoning to interpret ambiguous or narrative-style inputs. Following evaluation, the agent provides a structured **decision explanation**, detailing approval rationale or failure reasons for auditability and transparency. Finally, a **human-in-the-loop review** ensures oversight, with uncertain or exceptional cases referred to officials. Reviewer feedback is systematically logged to refine prompts and gate logic, creating a continuous learning loop that improves robustness, adaptability, and fairness over time.

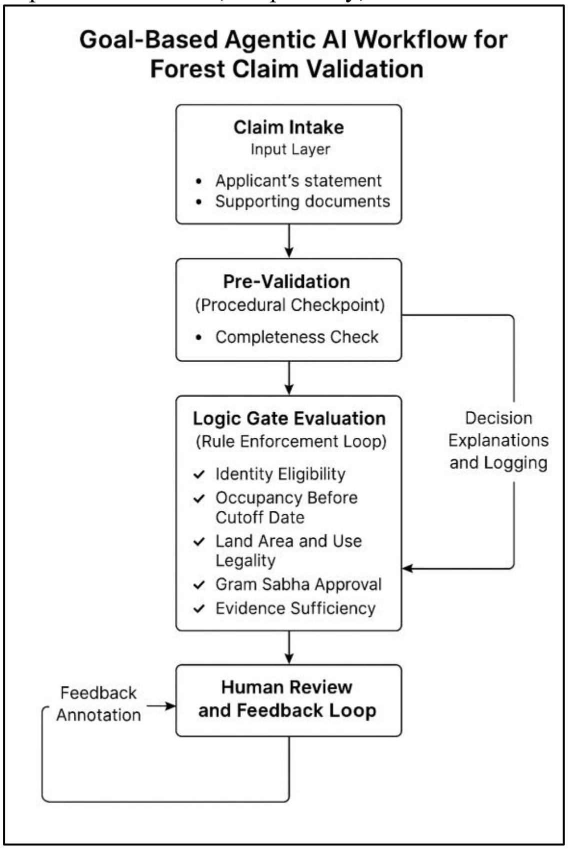

Fig 1: Goal based Agentic AI

6. CONCLUSION, LIMITATIONS, AND FUTURE WORK

This study presents a structured, explainable framework for automating validation of individual forest rights claims by combining logic gate enforcement with LLM-based interpretive flexibility. The system outperformed rule-based baselines and standard LLM classifiers, balancing strict compliance with FRA criteria and narrative interpretation while providing stepwise reasoning and transparent explanations. Current limitations include focus on individual (not community) claims, sensitivity to cultural/legal variations, and reliance on machine-readable inputs. Human oversight remains critical to ensure fairness and accountability. Future work involves field deployment, integration of human feedback, improved document recognition, and extending applicability to broader legal domains.

7. REFERENCES

[1] S. Agarwal, "Can technology support forest rights process?" Down To Earth, Dec. 2017. [Online]. Available:
[2] O. Ruhil, "Echoes of Ancestry in the Age of AI: Safeguarding Tribal Land Rights Amid a Digital Revolution," Indian J. of Social Science and Literature, vol. 4, no. 2, pp. 23–31, Dec. 2024.
[3] M. Kant et al., "Towards Robust Legal Reasoning: Harnessing Logical LLMs in Law," arXiv preprint arXiv:2502.17638, 2025.
[4] S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," in Proc. Int. Conf. Learn. Represent. (ICLR), 2023, arXiv:2210.03629.
[5] Y. Shen et al., "HuggingGPT: Solving AI tasks with ChatGPT and its friends in Hugging Face," arXiv preprint arXiv:2303.17580, 2023.

[6] J. Wei et al., “Chain-of-thought prompting elicits reasoning in large language models,” arXiv preprint arXiv:2201.11903, 2022.
[7] X. Wang et al., “Self-consistency improves chain-of-thought reasoning in language models,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2023, arXiv:2203.11171.
[8] D. M. Katz et al., “GPT-4 passes the bar exam,” Philos. Trans. Royal Soc. A, vol. 382, no. 2265, art. no. 20230254, 2024.
[9] J. Savelka and K. D. Ashley, “The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts,” Frontiers in Artificial Intelligence, vol. 6, art. 1279794, Nov. 2023.