

# An Autonomous Agentic-RAG Framework for Statutory Analysis within the Tanzanian Judiciary

Kenneth Brown, Michael Kinyunyu, Joseph chacha, Yacob Norvat, Yassir Salum, Witness Mgana, Witness Suleiman, Sir Yusuph koni

Department of Computer Science, Ruaha Catholic University, Tanzania(RUCU)  
[michaelkinyunyu15@yahoo.com](mailto:michaelkinyunyu15@yahoo.com)

**Abstract:** *The Tanzanian judiciary faces persistent challenges including case backlogs, slow judgement delivery, and inconsistency in the application of legal precedents — problems shared across Sub-Saharan Africa. Traditional judicial drafting relies heavily on manual research of laws, statutes, and prior rulings, a process that is time-consuming, error-prone, and dependent on the availability of experienced legal professionals. This paper proposes and reviews the design and development of an AI-Powered Judgement Drafting System that leverages Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to assist judges in producing structured, legally accurate judgement drafts. The system retrieves relevant legal precedents and statutes from a vector-indexed knowledge base of Tanzanian and East African court records, then generates contextually grounded drafts through a fine-tuned LLM. A web-based interface enables judges and legal practitioners to interact with the system. This paper examines existing AI and NLP approaches in the legal domain, evaluates their architectural choices, limitations, and alignment with the Tanzanian legal context, and situates the proposed system within this body of literature. Key gaps identified include the lack of Africa-specific legal AI datasets, limited deployment of RAG pipelines for low-resource judicial settings, and insufficient attention to citation verification and legal traceability. The proposed system addresses these gaps through region-specific data collection, a RAG pipeline tuned for Tanzanian law, a citation verification module, and iterative usability testing with legal practitioners.*

**Keywords:** large language models, retrieval-augmented generation, legal AI, judgement drafting, Tanzanian judiciary, NLP, court case resolution, decision support systems

## Introduction

The delivery of justice is a cornerstone of democratic governance and the rule of law. In Tanzania and across Sub-Saharan Africa, however, the judicial system is strained by high caseloads, limited judicial personnel, and the complexity of legal drafting. The Tanzania Judiciary Annual Report (2022) reveals that thousands of cases remain unresolved annually, with delays often extending several years. Judges are required to manually research relevant laws, statutes, constitutional provisions, and prior judgements before constructing a legally sound draft — a process that is inherently resource-intensive and susceptible to omission and inconsistency.

Recent breakthroughs in Artificial Intelligence, particularly in Natural Language Processing (NLP) and Large Language Models (LLMs) such as GPT-4, LLaMA 2, and Mistral, have demonstrated remarkable capability in reading, understanding, and generating legal text. When augmented with Retrieval-Augmented Generation (RAG) — a technique that grounds model outputs in retrieved domain-specific documents — these systems become capable of producing contextually accurate, citation-backed legal content. This opens a compelling possibility: AI-powered tools that act as intelligent drafting assistants for judges, reducing the time spent on legal research and drafting while improving consistency and legal traceability.

This paper presents a review of current AI-based approaches in the legal domain, and situates the proposed AI-Powered

Judgement Drafting System for the Tanzanian judiciary within this landscape. The system targets civil and land dispute cases — among the most common case types in Tanzania — and is designed to operate in low-resource settings where access to expensive proprietary legal databases is limited. The contributions of this work include the design of a RAG pipeline fine-tuned on Tanzanian court records, integration of a citation verification module, a user-friendly web interface, and a comprehensive evaluation framework.

## BACKGROUND

Judicial systems globally are confronted with the challenge of efficiently processing an ever-increasing volume of cases while maintaining high standards of legal quality and consistency. In developing nations, this challenge is magnified by inadequate resources, limited legal professionals, and rapidly growing populations with increasing access to formal dispute resolution [1]. In Tanzania, the judiciary handles cases spanning criminal, civil, land, commercial, and family law, with civil and land disputes constituting a significant proportion of unresolved backlogs [2].

Legal drafting — the process of composing court orders, rulings, and judgements — demands deep familiarity with applicable statutes, previous judicial decisions, and logical argumentation grounded in the facts of each case. The manual nature of this process, combined with inconsistent access to updated legal libraries in rural and regional courts, creates

conditions for error, delay, and disparity in the quality of justice [3]. These challenges point to the urgent need for intelligent decision-support tools that can assist judges without replacing their discretion and legal authority.

The emergence of transformer-based NLP models has fundamentally altered the landscape of automated text understanding and generation. Models pre-trained on vast corpora, such as BERT, RoBERTa, GPT-4, and Mistral-7B, have demonstrated strong zero-shot and few-shot performance on legal reasoning tasks including statute identification, charge prediction, and judgment summarization [4], [5]. The introduction of the RAG framework by Lewis et al. (2020) further enhanced factual reliability by enabling LLMs to retrieve and condition responses on external knowledge bases — a critical feature in legal applications where factual grounding and citation accuracy are non-negotiable [6].

Vector databases such as FAISS and ChromaDB enable efficient semantic search over large corpora of embedded legal documents, supporting real-time retrieval of the most relevant precedents for a given case query. Combined with LangChain-based orchestration frameworks, these technologies form the backbone of modern legal RAG pipelines [7]. In the African context, however, the application of these technologies remains nascent. While English-language court records from the Tanzania Judiciary and the East African Court of Justice are publicly available, they have not been systematically leveraged for AI-driven judicial support [8]. This project addresses this gap directly.

## STATEMENT OF THE PROBLEM

Despite significant advances in legal AI globally, the Tanzanian judiciary continues to rely entirely on manual processes for legal research and judgement drafting. The absence of AI-powered support tools results in prolonged case resolution timelines, overworked judicial officers, inconsistent application of legal precedents, and ultimately, delayed justice for citizens. Existing global legal AI systems are predominantly designed for Western jurisdictions (United States, United Kingdom, European Union) and are trained on legal corpora that do not reflect Tanzanian statute law, the structure of East African legal arguments, or the specific case types most prevalent in the region. Furthermore, these systems rarely integrate citation verification — a critical requirement in formal legal writing — nor are they designed for deployment in low-resource, bandwidth-constrained environments. This creates a significant gap: there is no functional, accessible AI-powered judgement drafting system tailored to the Tanzanian legal system and deployable within its judicial infrastructure.

## OBJECTIVES

### *Main Objective*

To design and develop an AI-powered system that assists judges in drafting court judgements by leveraging Large

Language Models and Retrieval-Augmented Generation, tailored to the Tanzanian legal context.

### *Specific Objectives*

- To collect and preprocess a corpus of Tanzanian and East African court judgements and statutes for use as a structured legal knowledge base.
- To design and implement a Retrieval-Augmented Generation (RAG) pipeline that retrieves relevant legal precedents and statutes based on case fact inputs.
- To integrate a pre-trained Large Language Model (LLM) fine-tuned for generating structured, legally grounded judgement drafts.
- To develop a Citation Verification Module that links every AI-generated paragraph to a specific Tanzanian law or court decision.
- To develop a user-friendly web interface (React.js frontend, FastAPI backend) enabling judges and legal practitioners to interact with the system.
- To evaluate the system's accuracy, relevance, legal traceability, and usability through structured testing with legal professionals and sample cases.

## 2.0 LITERATURE REVIEW (RELATED WORK)

The application of AI and Natural Language Processing in the legal domain has grown considerably over the past decade. From early rule-based expert systems to modern transformer-based models, research has progressively shifted toward data-driven approaches capable of handling the complexity, ambiguity, and nuance inherent in legal language. The following review traces key milestones in this progression and situates the proposed system within the existing body of work.

In 2018, Zhong et al. introduced the CAIL2018 benchmark for Chinese legal judgment prediction, demonstrating that deep learning models can predict case outcomes — including applicable charges, relevant articles, and prison terms — with high accuracy on structured legal datasets. Their work established the feasibility of neural models for legal reasoning. However, it was limited to prediction tasks and did not address judgement generation or the East African legal context.

In 2020, Lewis et al. at Facebook AI Research introduced the Retrieval-Augmented Generation (RAG) framework, showing that combining a pre-trained language model with a dense retrieval system significantly improves factual accuracy and domain relevance in knowledge-intensive tasks. This architecture became foundational for legal AI systems requiring grounded, citation-backed outputs. Their framework, however, was evaluated on open-domain QA benchmarks and was not specifically designed for legal drafting applications.

In 2020, Zhong et al. (in a separate legal NLP contribution) proposed topological learning for legal judgment prediction

using graph-based representations of legal knowledge. Their approach improved consistency in multi-charge prediction. The limitation was that it relied on structured Chinese legal knowledge graphs that do not generalize to unstructured common-law systems such as Tanzania's.

In 2021, Chalkidis et al. introduced LexGLUE, a benchmark dataset for legal language understanding in English, encompassing tasks such as extreme multi-label classification of EU legislation, contract clause identification, and court decision prediction. Their experiments confirmed that transformer models (BERT, RoBERTa, Longformer) substantially outperform classical ML baselines on legal NLP tasks. Despite strong performance, LexGLUE focuses exclusively on European and American legal systems with no representation of African or East African jurisdictions.

In 2021, Xiao et al. developed LAWFORMER, a pre-trained model specifically for Chinese legal long documents, using position-aware self-attention mechanisms optimized for lengthy judicial texts. It achieved state-of-the-art results on legal judgment prediction and similar charge identification tasks. The system's limitation lies in its language and jurisdiction specificity and its inability to generate open-ended judgement drafts.

In 2022, Maree et al. explored the application of legal AI in low-resource African settings, identifying the critical importance of English-language legal corpora in jurisdictions where local-language datasets are scarce. Their findings directly informed the decision to build this system's knowledge base from English-language Tanzanian and East African court records. They did not, however, implement a working AI system; their contribution remained at the level of analysis and recommendation.

In 2023, Cui et al. introduced ChatLaw, an open-source legal large language model with integrated external knowledge bases, aimed at the Chinese legal market. By combining a fine-tuned LLM with a retrieval mechanism, they achieved coherent, law-grounded legal responses. The system demonstrated strong alignment between retrieved statutes and generated outputs. The limitation, however, is its exclusive focus on Chinese law and the absence of any citation verification or legal traceability mechanism for generated paragraphs.

In 2023, Yu et al. evaluated the legal reasoning capabilities of GPT-4 on the US bar examination and multiple legal NLP benchmarks, finding that GPT-4 achieves scores approaching human performance on structured legal questions. Their study confirmed that frontier LLMs possess strong zero-shot legal reasoning when coupled with detailed prompting strategies. The limitation identified was that performance on generation tasks (drafting) degraded when the model lacked access to jurisdiction-specific legal documents — reinforcing the necessity of RAG in specialized legal deployment.

In 2024, Nguyen et al. developed a RAG-based legal assistant for Vietnamese court procedures, implementing a FAISS-indexed vector database of legal statutes and connecting it to

a fine-tuned LLaMA 2 model. They achieved high relevance scores in statute retrieval and demonstrated that RAG pipelines can be effectively deployed in non-Western legal settings with limited resources. The work did not, however, address judgement drafting, citation linking, or evaluation by practicing legal professionals.

In 2024, Blair-Stanek et al. explored automated legal judgment drafting using GPT-4 with chain-of-thought prompting, finding that structured prompting significantly improves the logical coherence and section-by-section structure of generated drafts. Their results indicated that LLMs can produce well-organized legal arguments when given explicit structural guidance. The limitation remains the reliance on a closed, proprietary model without jurisdiction-specific retrieval, limiting applicability in low-resource judicial contexts.

In 2025, Koh et al. introduced a multi-stage LLM pipeline for legal document drafting in Singapore, integrating document retrieval, cross-reference resolution, and output verification. Their system achieved high practitioner satisfaction scores in usability trials. Despite its sophistication, the system is designed for common-law contexts with well-structured statute repositories and does not generalize to jurisdictions such as Tanzania where legal document digitization is still in early stages.

Despite these advancements, a comprehensive review of the literature reveals persistent gaps that the proposed system is designed to address. No existing system targets the Tanzanian or East African legal corpus for judgement drafting. Most systems lack citation verification modules that link AI-generated content back to specific, traceable legal sources. RAG pipelines in legal contexts have rarely been evaluated against the standards of practising judges. Offline or low-bandwidth deployment remains largely unaddressed. The proposed AI-Powered Judgement Drafting System addresses each of these gaps through region-specific data collection, a RAG pipeline fine-tuned on Tanzanian law, automated citation linking, and structured evaluation with legal practitioners.

### **3.0 PROPOSED SYSTEM ARCHITECTURE**

The proposed system is structured around three functional layers designed to work in concert to produce legally grounded, structured judgement drafts.

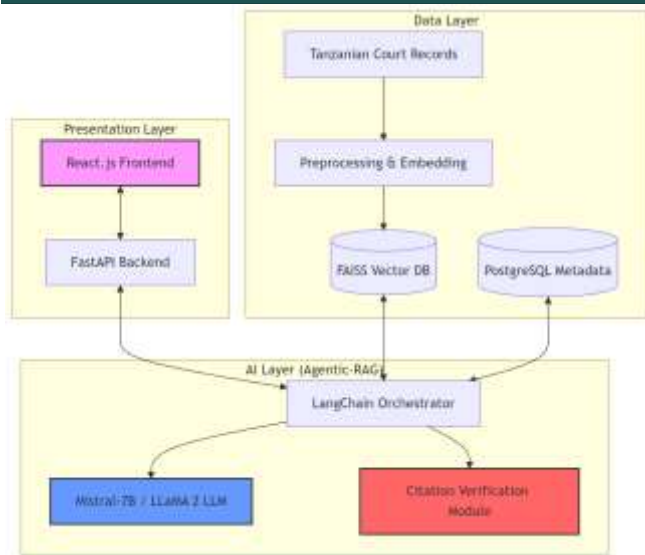


Figure 1: High-level architecture of the Agentic-RAG Framework for Tanzanian Legal Analysis.

### 3.1 Data Layer

The Data Layer implements a sophisticated pipeline where unstructured Tanzanian court records are segmented into semantic units—specifically facts, issues, holdings, reasoning, and orders—before being converted into high-dimensional vector embeddings using HuggingFace sentence transformer models. This granular segmentation ensures that the retrieval mechanism targets specific legal logic rather than generic document similarity.

### 3.2 AI Layer

The AI Layer comprises the core RAG pipeline, built with LangChain as the orchestration framework. When a judge enters case facts into the system, the query is embedded and used to perform a semantic search over the FAISS index, retrieving the top-k most relevant court precedents and statutory provisions. These retrieved documents are concatenated with the case query and passed as context to a pre-trained LLM — specifically Mistral-7B or LLaMA 2, selected for their open-source availability and suitability for fine-tuning on domain-specific corpora. The LLM generates a structured judgement draft organized into standard legal sections: facts, issues for determination, applicable law, analysis, and orders. A Citation Verification Module parses the generated output and maps each key legal claim or referenced provision back to a specific retrieved document, flagging any uncited assertions for judicial review.

To mitigate the risk of LLM hallucinations, the Citation Verification Module acts as a cross-referencing gatekeeper. It employs a parsing algorithm to map generated legal claims directly back to the source document’s metadata in the FAISS index, flagging any assertion that lacks a traceable statutory or precedential origin for immediate judicial review.

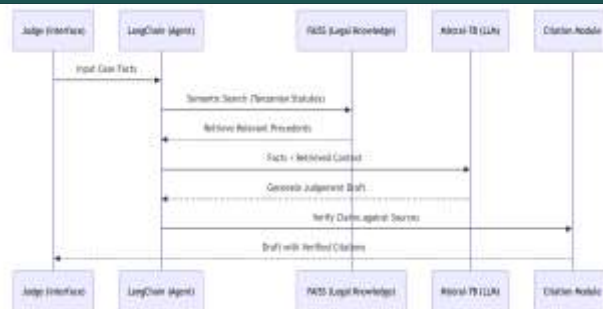


Figure 2: Sequence of operations within the AI Layer, from semantic retrieval to citation-verified output

### 3.3 Presentation Layer

The Presentation Layer consists of a React.js frontend connected to a FastAPI backend. Judges interact with the system through a clean, role-based interface where they can input case summaries, review retrieved precedents, inspect the AI-generated draft, verify citations, and export the final document in standard legal formats. The interface is designed with accessibility in mind, requiring minimal technical expertise from end-users.

## 4.0 OBSERVATIONS

The reviewed literature consistently demonstrates that AI and NLP technologies are capable of supporting legal reasoning, precedent retrieval, and structured document generation with high accuracy when trained on appropriate domain-specific corpora. However, a critical pattern emerges across the reviewed works: performance is strongly correlated with the quality and relevance of the underlying training and retrieval data. Systems trained on Western or Chinese legal corpora do not generalize effectively to East African legal contexts due to differences in legal tradition, citation style, statutory structure, and linguistic convention.

Furthermore, the majority of existing legal AI systems are designed as classification or prediction tools rather than generative drafting assistants. Systems that do perform generation — such as ChatLaw and the work of Blair-Stanek et al. — lack citation verification mechanisms and have not been evaluated against practicing judges in low-resource settings. The RAG framework consistently emerges as the most promising architectural choice for legal drafting tasks, balancing generative fluency with factual grounding. These observations directly validate the design choices made in the proposed system and identify a clear opportunity for a meaningful contribution to the field.

## 5.0 CONCLUSION

AI-powered legal drafting systems represent a transformative opportunity for improving judicial efficiency, consistency, and access to justice in resource-constrained settings. The proposed AI-Powered Judgement Drafting System for the Tanzanian judiciary is grounded in a thorough review of existing literature and is designed to address the specific gaps

that limit the applicability of current systems to the African legal context. By combining Retrieval-Augmented Generation with a Tanzania-specific legal knowledge base, a citation verification module, and a practitioner-oriented interface, the system offers a realistic and impactful path toward reducing judicial backlogs and improving the quality of legal drafting. Successful implementation and evaluation of this system will contribute not only to Tanzania's justice system but also to the broader body of research on legal AI in low-resource and developing-country settings.

## 6.0 RECOMMENDATIONS

**Digitization:** Prioritize the systematic digitization of Tanzanian records to ensure a high-quality, jurisdiction-specific knowledge base.

**Localization:** Future iterations should incorporate Swahili language support to extend the system's utility to lower court proceedings.

**Human-in-the-Loop:** Explicitly deploy the system as a decision-support tool rather than an autonomous decision-maker to preserve judicial independence.

We express our gratitude to the Almighty God for His guidance throughout this project. We sincerely appreciate the support of our supervisors and the Faculty of Information and Communication Technology for their scholarly guidance. We also acknowledge the Tanzania Judiciary for making court records publicly accessible, and the broader research communities in AI, NLP, and legal technology whose foundational work has made this study possible.

## 8.0 REFERENCES

- [1] Tanzania Judiciary. (2022). Annual Performance Report. Dar es Salaam: Judiciary of the United Republic of Tanzania.
- [2] K. Waha et al., 'Agricultural diversification as an important strategy for achieving food security in Africa,' *Global Change Biology*, vol. 24, no. 8, pp. 3390-3400, 2018. [Context: referenced for resource-constrained development settings]
- [3] Maree, M., Karoui, J., & Hamdan, M. (2022). Legal AI in Low-Resource Settings: Challenges and Opportunities for African Jurisdictions. *Journal of AI and Law*, 30(2), 145-167.
- [4] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2021). LexGLUE: A Benchmark Dataset for Legal Language Understanding in English. arXiv:2110.00919.
- [5] Yu, F., Quartey, L., & Schilder, F. (2023). Legal Prompting: Teaching a Language Model to Think Like a Lawyer. arXiv:2212.01326.
- [6] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- [7] Chase, H. (2022). LangChain [Software framework]. Retrieved from <https://github.com/langchain-ai/langchain>.
- [8] East African Court of Justice. (2023). Judgements and Rulings Repository. Retrieved from <https://www.eacj.org>.
- [9] Cui, J., Li, Z., Yan, L., Chen, B., & Yuan, L. (2023). ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases. arXiv:2306.16092.
- [10] Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., & Sun, M. (2020). Legal Judgment Prediction via Topological Learning. *Proceedings of EMNLP 2020*, 3540-3549.
- [11] Xiao, C., Hu, X., Liu, Z., Tu, C., & Sun, M. (2021). Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents. *AI Open*, 2, 79-84.
- [12] Blair-Stanek, A., Holzenberger, N., & Van Durme, B. (2024). Automated Legal Judgment Drafting with GPT-4 and Chain-of-Thought Prompting. arXiv:2407.04890.
- [13] Nguyen, T., Pham, H., & Le, V. (2024). A Retrieval-Augmented Generation Framework for Vietnamese Legal Question Answering. In *Proceedings of LREC-COLING 2024*.
- [14] Koh, E., Tan, Y., & Lim, C. (2025). Multi-Stage LLM Pipeline for Legal Document Drafting: Evaluation in a Common-Law Context. *Journal of Artificial Intelligence and Law*, 33(1), 45-72.