Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 14, Issue. 7, July 2025, pg.43 – 56

Zimbabwe LexLM: Crafting and Validating a Legal NLP Model for Zimbabwe's Bench and Bar

Simbarashe Tinotenda Mambara[1]; Wellington Manjoro[2]; Amanda Chiwanza[3]

[1]Data Science & Analytics & Harare Institute of Technology, Zimbabwe
[2]Information Science & Technology & Harare Institute of Technology, Zimbabwe
[3]Computer Science & Harare Institute of Technology, Zimbabwe
[1] stmambara@gmail.com; [2] wmanjoro@hit.ac.zw; [3] achiwanza@hit.ac.zw

DOI: https://doi.org/10.47760/ijcsmc.2025.v14i07.005

Abstract: We introduce Zimbabwe LexLM, a natural language processing (NLP) model designed to assist the legal community in Zimbabwe. Our method involves further pretraining a multilingual transformer model on a specialized collection of local legal texts. This collection includes over 77,000 documents such as case laws, statutes, and legal commentaries from Zimbabwe and nearby jurisdictions. The LexLM architecture enhances the transformer with a citation graph module to understand references between documents and legal context. We tested LexLM on legal classification and retrieval tasks, achieving a macro-F1 score of about 0.865 and an accuracy of around 0.88. These results show that domain-specific training and citation-aware models can be effective in the legal field. Zimbabwe LexLM offers a strong, specialized model and dataset to improve legal research and information retrieval in resource-limited settings. Our work demonstrates how custom language models can assist legal practitioners in accessing information more efficiently, especially in low-resource contexts.

Keywords: Zimbabwe, Legal NLP, Low-Resource Jurisdiction, Domain-Adaptive Pretraining, Graph Neural Networks, Semantic Legal Search, Access-to-Justice AI, Transformer-Based Retrieval, Citation-Aware Embedding

I. INTRODUCTION

The legal landscape in Zimbabwe is changing quickly as part of a wider move toward digital transformation, but there are still significant challenges in accessing laws and court decisions. Natural language processing (NLP) and AI tools are increasingly being used in legal research. For example, transformer-based models have helped find relevant precedents and accurately classify case law [14][17]. In Zimbabwe, the government has launched major e-justice initiatives. Notably, in 2021, the Judicial Service Commission partnered with Synergy International to develop a court-based Electronic Case Management System (IECMS), with the first phases going live in 2022 [54][39]. These reforms, which align with the judiciary's push for technology to improve efficiency [35], show a clear move towards e-governance in the courts. At the same time, online resources like the Zimbabwe Legal Information Institute (ZimLII) offer limited coverage of laws and judgments, making it hard to find the legal texts needed. Overall, with more digital court data and scattered, hard-to-search archives, there's a real opportunity to use NLP to improve access to Zimbabwean legal information.

Despite new digital infrastructure, Zimbabwe's existing search tools remain basic. Users must navigate isolated databases and depend on simple keyword searches, which struggle with legal terminology and the subtleties of case language. Like other developing systems, even well-meaning reforms such as expanded e-filing and online case tracking through the IECMS have not automatically solved the retrieval issues [33] [39]. Commentators observe that digitalisation in Zimbabwe has yet to fully close access gaps [52], and many citizens and lawyers still find it hard to obtain relevant laws or precedents quickly. For instance, labour and administrative courts were only recently integrated into the digital system [33], and even when cases are online, the absence of advanced search capabilities prevents practitioners from easily utilising this data. Therefore, a clear gap exists between the potential of judicial digitalisation and the real accessibility of legal information, a gap this research aims to close by incorporating NLP.

NLP techniques are particularly promising for improving access. Globally, legal NLP systems can interpret natural language queries and link them to the correct statutes or judgments. Studies have indicated that BERT-style retrieval models, especially when enhanced with citation graphs, can outperform keyword searches in locating case law [14] [17]. Domain-specific language models like LEGAL-BERT and other multilingual legal BERTs have advanced tasks like entity recognition and semantic search across different jurisdictions. In Zimbabwe, where English is the official language [28] and many cases cite South African precedents, a fine-tuned NLP model could utilise this common-law heritage. A Zimbabwe-adapted model could recognise legal terminology and query intent, for example, translating a layperson's question into legal keywords, thereby revealing the relevant statutes or past rulings even if the wording varies. Currently, no such NLP-driven search tool exists locally. The literature on Zimbabwe's e-justice notes that, although digital systems have been implemented, the advantages of AI-based search remain unrealised [52]. This presents a new opportunity, transforming the expanding digital case repository from IECMS and other archives into a structured, searchable resource through NLP.

In summary, this paper looks at how legal reforms in Zimbabwe intersect with advances in legal AI. It addresses the problem of legal information discovery being inefficient and highlights the shift to digital case management in the court system. Introducing NLP is especially timely; it has proven useful in improving legal research in other countries and has strong potential to do the same in Zimbabwe. Tackling this issue is important locally, as it can help improve access to justice and support the rule of law, while also contributing to the global movement towards smart legal information systems.

II. LITERATURE REVIEW

Legal NLP faces unique challenges because of the complex terminology, formal tone, and high stakes involved in legal documents. While recent advances in deep learning, especially transformer models, have propelled progress in general NLP, the field of legal NLP has been slower to develop. It is still relatively unexplored for many low-resource languages.

This literature review examines the existing research on legal NLP, especially in low-resource languages, highlighting key themes and gaps. In Zimbabwe, courts use English as the only language of record. While global research often focuses on multilingual and low-resource legal NLP, Zimbabwe mainly faces challenges due to limited resources in English-language legal texts.

A. *Zimbabwe Legal System Overview and the IECMS*

Zimbabwe's legal system blends common and customary laws, influenced by Roman-Dutch traditions. In the past, accessing legal information meant poring over printed law reports and government gazettes, which made research slow and tedious. Only parts of the law, like those on ZimLII, have been digitized so far. Finding relevant cases or statutes often required manual searches through archives. These limited digital resources have spurred efforts to modernize the country's legal information systems.

A key element of this modernisation is the IECMS, introduced by the judiciary in 2022 to digitise court processes [39]. The IECMS facilitates electronic case filing, online fee payments, virtual hearings, digital case tracking, and e-access to judgments, helping Zimbabwe's courts align with global e-justice trends. The initial rollout of the system, starting with the higher courts in 2022, received positive feedback from judges and lawyers and was informed by lessons from similar e-court initiatives [51], [39]. The reform is supported by robust institutional backing: the Judicial Service Commission's strategic plans aim for full court automation by 2025, and new rules and laws, including a 2023 Judicial Laws Amendment Bill, formally recognise electronic filings and virtual hearings.

Like any major digital project, the IECMS encounters challenges. Some litigants have faced issues with limited internet access while other users found it difficult to adapt to digital systems [15]. Cybersecurity remains a concern, especially with incidents of website hacking [43]. Continuous investments in network infrastructure, power backups, user training, and data security are helping to resolve these problems. Overall, the IECMS is key infrastructure to building a growing collection of electronic legal documents. These digital resources lay a foundation that future legal NLP applications can use to improve legal information retrieval and analysis.

B. *Conceptual Framework*

Legal NLP employs natural language processing techniques to analyse legal texts, including legislation, judicial decisions, contracts, and similar documents, for tasks such as information extraction, document classification, summarisation, question answering, and judgment prediction. The specialised nature of legal language encompasses formal jargon, archaic terminology including Latin phrases and a structured rhetorical format that significantly differs from ordinary language. General-purpose NLP models frequently encounter difficulties in handling these distinctive features, thereby motivating efforts to incorporate domain-specific knowledge into legal NLP. Researchers have established legal ontologies and standardised identifiers for legal references to enhance the representation of legal semantics [10], [38].

They have also developed specialised language models called legal BERTs by pre-training transformers on extensive legal corpora, and designed model architectures to accommodate features of legal text, such as very long documents or complex citation

networks [16]. These approaches lay a foundation for NLP systems that can interpret the structure and meaning of legal documents more effectively than standard models.

A key theme in the literature is addressing low-resource languages. The most incredible NLP advances have concentrated on a few high-resource languages, while many jurisdictions lack large, annotated corpora or pre-trained models [19], [33]. This gap is evident in the legal domain, English-language legal NLP is relatively well-developed, whereas many others have seen little NLP research. Bridging this divide requires techniques like cross-lingual transfer and multilingual modelling. Multilingual transformer models like mBERT and XLM-R can transfer knowledge from resource-rich languages to low-resource ones, and strategies like using translated text or parallel corpora can boost performance on under-resourced legal text [19], [45]. Data augmentation and meta-learning have also been proposed to cope with scarce training data [17]. The underlying premise is that combining domain-specific training with multilingual transfer learning makes it possible to build effective legal NLP solutions even for jurisdictions with minimal existing resources.

C. *Empirical Literature*

Early applications of machine learning to legal text proved that automated analysis is feasible but also exposed limitations. For instance, classifiers were trained to predict court decisions from text with around 70–80% accuracy in specific contexts [1], [21]. Likewise, researchers built some of the first legal text summarizers to extract key sentences, such as facts, issues, and conclusions from long judgments [14]. These early studies looked at legal documents as simple text. This meant that early models often overlooked the complex structure and context of legal reasoning. Additionally, they mostly focused on English language documents or data from a single jurisdiction.

Introducing deep learning and transformer models led to significant improvements in performance. A notable breakthrough was the development of law-specific pre-trained models. Chalkidis et al. [5] demonstrated that an English Legal BERT model trained on legal documents outperforms a generic BERT on legal classification tasks. Similar domain-adapted models have since been developed for other languages such as French and Greek, among others, with comparable benefits [31], [40]. Concurrently, multilingual approaches have emerged. Tan et al. [45] designed a judgment prediction model that transfers legal knowledge across Chinese and European case corpora, enhancing accuracy in each. Chalkidis et al. [7] showed that pre-training on a large multi-language legal corpus produces state-of-the-art results across diverse jurisdictions.

Research has also expanded to include many jurisdictions beyond the traditional Anglophone focus. Large legal datasets have been assembled for countries like China and India, offering tens of thousands of annotated cases for training models [54], [3]. New benchmarks in various European languages, such as Romanian, Dutch, and German, along with multilingual corpora like a trilingual Swiss court dataset, now allow evaluation of NLP systems across a broader range of legal traditions [24], [34]. These resources have played a key role in shifting the field from an Anglocentric view.

To further enhance performance, recent studies incorporate extensive legal knowledge into models. Providing a model with the text of relevant laws or statutes and case facts can make predictions more legally grounded [35]. Modelling the network of case precedents, which cases cite which, via graph neural networks has likewise improved tasks such as case outcome prediction and legal case retrieval [13]. Other work introduced architectures that reflect the structure of legal documents: processing a judgment in sections facts, arguments, rulings instead of as one long sequence yielded better results for lengthy texts [22], and multi-task learning of related legal attributes, such as predicting both applicable law and case outcome together, improved generalisation when data for each task was limited [37].

Beyond court decision analysis, legal NLP has been applied to various other tasks. Legal question-answering systems have been developed to address queries about statutes by retrieving relevant provisions and then predicting the answer [12]. Efforts in legal argument mining aim to automatically identify argument components within judicial opinions, such as premises and conclusions [2], [50]. Initial work has even explored AI-assisted drafting of legal texts, suggested contract clauses, although ensuring legal correctness remains challenging [53].

A recurring challenge in this field is the limited availability of labelled data in many languages. Researchers have explored methods to maximise learning from small datasets. One such method is meta-learning; Huang et al. [17] showed that training a model on numerous small, simulated tasks enables it to quickly adapt to a new legal prediction task with only a few examples, outperforming traditional fine-tuning. Another strategy involves data augmentation, which, adds paraphrased or synthetic text to expand the training set, and significantly improves performance in low-resource settings [8]. Transfer learning from resource-rich languages or related tasks is also widely used. Moreover, concerns about fairness and transparency have driven efforts toward explainability. Models have been developed to identify the key parts of a judgment that influence their predictions, ensuring the reasoning aligns with legal principles in a clear and understandable way for humans. [6]. As Tsarapatsanis et al. [48] warn, black-box legal AI systems risk perpetuating biases or eroding trust without proper safeguards. Therefore, current research emphasises accuracy and developing interpretable and accountable legal NLP models.

D. *Conclusion*

The existing literature highlights rapid progress at the crossroads of computational linguistics and law. NLP models, especially transformer-based architectures designed for legal language, now handle tasks that previously needed human expertise, such as summarising lengthy court decisions or predicting case outcomes. Researchers have achieved high accuracy across different jurisdictions by using techniques like domain-specific language training, multilingual learning, and incorporating legal knowledge such as citation graphs or statutory context into models. These approaches are vital for handling the complexity of legal language and the limited availability of annotated data. Moreover, the creation of legal text corpora and benchmark datasets in multiple languages and legal systems has greatly advanced this field progress.

III. RESEARCH METHODOLOGY

A. *Introduction*

The methodology for Zimbabwe LexLM combines corpus construction with domain-adaptive machine learning. We first assemble a comprehensive Zimbabwean legal text corpus and then use it to train transformer-based language models augmented with a citation graph. This approach follows best practices: further pretraining a multilingual transformer on Zimbabwe-specific legal text and integrating a case law citation network to capture precedential structure. Figure 1 illustrates the overall workflow.

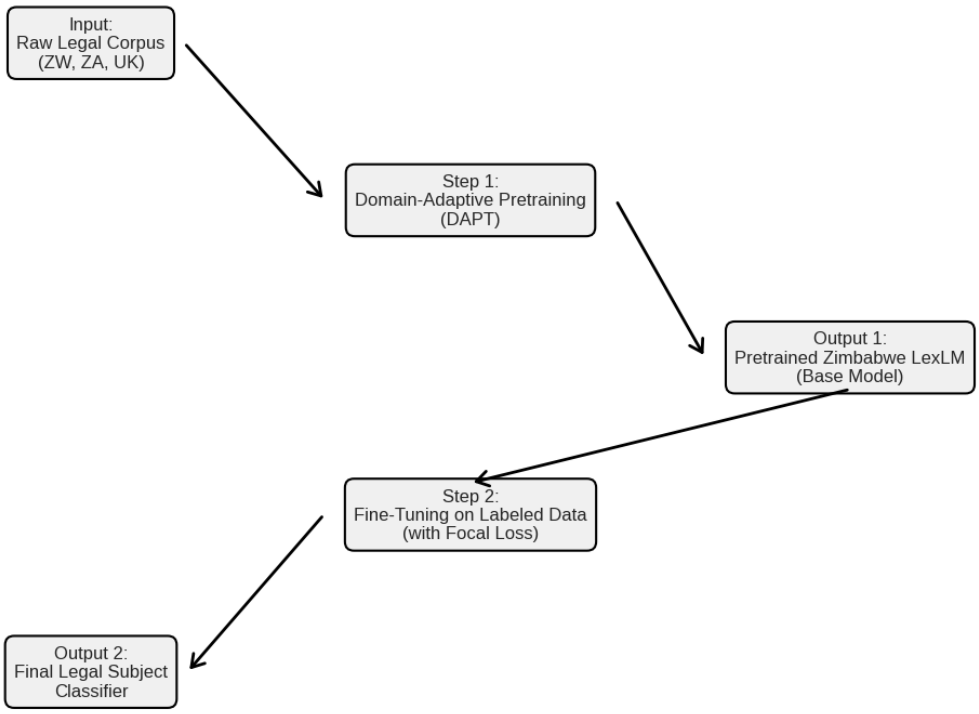

Figure 1: The Zimbabwe LexLM development pipeline.

B. *Methodology Approach*

Our research design is iterative. After compiling the corpus, we train and evaluate models. We follow best practices for low-resource settings by adapting a multilingual transformer to the Zimbabwean legal domain and incorporating a citation graph [1]. For robustness, we partitioned the data into training, validation, and test sets 70%, 10%, 20% and used cross-validation on the validation set to tune hyperparameters. The final evaluation is on the test set with metrics like precision, recall, F1, and accuracy.

C. *Dataset Description*

Version 1.0 of the Zimbabwe Legal Corpus contains 77,133 documents spanning 1936–2024, covering Zimbabwe's case law and legislation. It includes all significant Zimbabwean court judgments, the High, Supreme, and Constitutional Courts and laws and selected South African and English cases. Key legislative texts, legislation, and other legal documents are also incorporated. Each document is stored with metadata, for example, ID, source, date, and type, to facilitate filtering and analysis.

D. *Dataset Creation Process*

We collected legal documents from archives and websites, including scanned law reports and digital files of judgments and statutes. Next, we extracted text using OCR for scanned pages and applied custom parsing scripts, regex-based, to structure each document. A machine-learning classifier helped remove extraneous content, such as reference lists in some files, with manual checks for accuracy. We standardised metadata, for example, normalised dates and court names and removed duplicate entries. Finally, the cleaned documents were integrated into a single corpus with rigorous quality checks and stored in a columnar format Parquet for efficient access.

E. Data Modelling Methodology

We trained a specialised language model for the legal domain. First, we performed domain-adaptive pretraining of a transformer on our legal corpus. Starting from a multilingual model XLM-R, we continued training it on our corpus using the Masked Language Modelling objective. The MLM loss is:

$$L_{MLM} = -\frac{1}{|M|}\sum_{t\in M}\log P_\theta(W_t|\{W_{1,\dots,}W_{t-1}, W_{t+1}, \dots\}) \quad 1$$

Where M:x is the set of masked token positions and W_t is the token to predict. This step adapts the model to Zimbabwean legal vocabulary and style. Next, we fine-tuned the model on downstream tasks, for example, legal text classification by adding a task-specific output layer and training on labelled examples. Second, we augmented the model with a citation graph. We constructed a graph of legal documents where nodes represent cases or statutes and edges link documents that cite one another. Each node's initial feature vector is its text embedding from the transformer. We then applied a graph neural network (GNN) to propagate information along citation edges. Formally, let A be the adjacency matrix augmented with self-loops of the citation graph and $D^{-1/2}H^{(l)}$ the initial node embeddings from the transformer. We used a standard graph convolution layer:

$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}}(A+I)D^{-1/2}H^{(l)}W^{(l)} \quad 2$$

where D is the degree matrix of $(A+I)$ and $W^{(l)}$ They are trainable weights. Each case's representation is updated to include a normalised sum of its neighbours' representations, followed by a nonlinear activation. Stacking multiple GNN layers allows information to propagate across multiple citation hops. This produces enriched document embeddings that capture both a document's textual content and its context in the network of legal precedents. We also experimented with a multi-task setup using multiple prediction heads on a shared encoder [2]. Evaluation We split the dataset into 70% training, 10% validation, and 20% test documents. For classification tasks, for example, categorizing case topics or predicting binary labels, we report accuracy and F□-score. Precision (P) and recall (R) are defined as

$$P = \frac{TP}{TP+FP} \quad 3$$

$$R = \frac{TP}{TP+FN} \quad 4$$

and the F1-score is the harmonic mean:

$$F_1 = 2\cdot\frac{P\cdot R}{P+R} \quad 5$$

Here TP, FP , and FN denote true positives, false positives, and false negatives on the test set. We compute both micro-averaged and macro-averaged F1 when evaluating multi-class tasks. Accuracy is computed as

$$\frac{(\mathrm{TP}+\mathrm{TN})}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \quad 6$$

For retrieval-style evaluation if treating query-document matching as ranking, we also compute Mean Reciprocal Rank (MRR):

$$MRR = \frac{1}{|Q|}\sum_{q=1}^{|Q|}\frac{1}{\text{rank}_q} \quad 7$$

where rank_q is the rank of the correct answer for the query q. We followed the methodology of typical legal QA systems, measuring how high the ground-truth document appeared in the ranked results. If the task involved language modelling quality, we would report perplexity, defined as

$$Perplexity = \exp(-\frac{1}{N}\sum_{i=1}^{N}\log P_\theta(xi)) \quad 8$$

where xi These are the test tokens. In practice, our fine-tuning tasks did not focus on raw text generation, so perplexity was secondary we prioritised task-specific accuracy metrics.

F. *Ethical Considerations*

All documents are public legal records, including court judgments and laws in the public domain. However, their publishers' copyright law reports, headnotes and similar editorial content are observed. We included such content for research purposes, and it should not be redistributed. Our corpus is intended for non-commercial research. Historical legal data may contain biases, and LexLM is not a substitute for legal advice; we caution against misuse in judicial decision-making. We incorporated basic explainability to highlight which inputs, such as text or citations, most influenced the model's predictions, following methods like those in [3]. We credit the creators of the Zimbabwe Legal Corpus [4] and adhered to relevant data use policies during this project.

IV. RESULTS AND FINDINGS

A. *Introduction*

This chapter presents the empirical findings from the Zimbabwe legal corpus and classification experiments, extending the dataset and methodology described in Chapter III. We first characterise the assembled corpus (Section IV.B) and perform exploratory analysis (Section IV.C) to uncover its structural and topical features. Next, we report on model evaluation results (Section IV.D), including comparisons with baseline and domain-specific models. Finally, we outline key aspects of the proposed LexLM architecture (Section IV.E).

B. *Insights from the Data*

Table 1 summarises the composition of the Zimbabwe LexLM corpus. The dataset comprises 83,384 documents, of which approximately 52.6% originate from Zimbabwean courts and 43.7% from South African courts with historical English reports making up the rest. Over half of all documents are law reports, while about 29.6% are unreported High Court judgments. The most significant contributors are the South African Law Reports (SALR) with approximately 38.7% of the corpus and other major series. This highlights a need for domain adaptation or oversampling to ensure adequate coverage of Zimbabwean cases. Reported cases include headnotes and detailed citations that facilitate tasks like legal retrieval, whereas unreported judgments often lack such metadata.

Item	Docs	Share
South African Law Reports (SALR)	32 295	38.73 %
Zimbabwe High Court (unreported, HC)	24 701	29.62 %
Other (inc Sis, Acts, etc)	12 076	14.48 %
Zimbabwe Law Reports (ZLR)	4 190	5.02 %
South African Criminal Law Reports (SACLR)	4 159	4.99 %
English Law Reports (ELR, 1936‑43)	3 031	3.63 %
Zimbabwe Supreme Court (SC)	2 526	3.03 %
Zimbabwe Constitutional Court (CC)	406	0.49 %
Total	83 384	100 %

Table 1: Composition of the Zimbabwe LexLM corpus by jurisdiction and source

C. *Exploratory Data Analysis*

The corpus exhibits a wide range of judgment lengths. Zimbabwean judgments have a median around 2,400 words and IQR roughly 1,400–3,400 with relatively few extreme outliers, indicating most cases are of moderate length. South African judgments are longer on average with median approximately 3,800 with a heavy upper tail. This may suggest that many law-report cases extend to tens of thousands of words for South Africa. Historical English reports similarly have elevated medians and substantial outliers. This skew suggests that models must accommodate both typical-length cases and occasional very long documents.

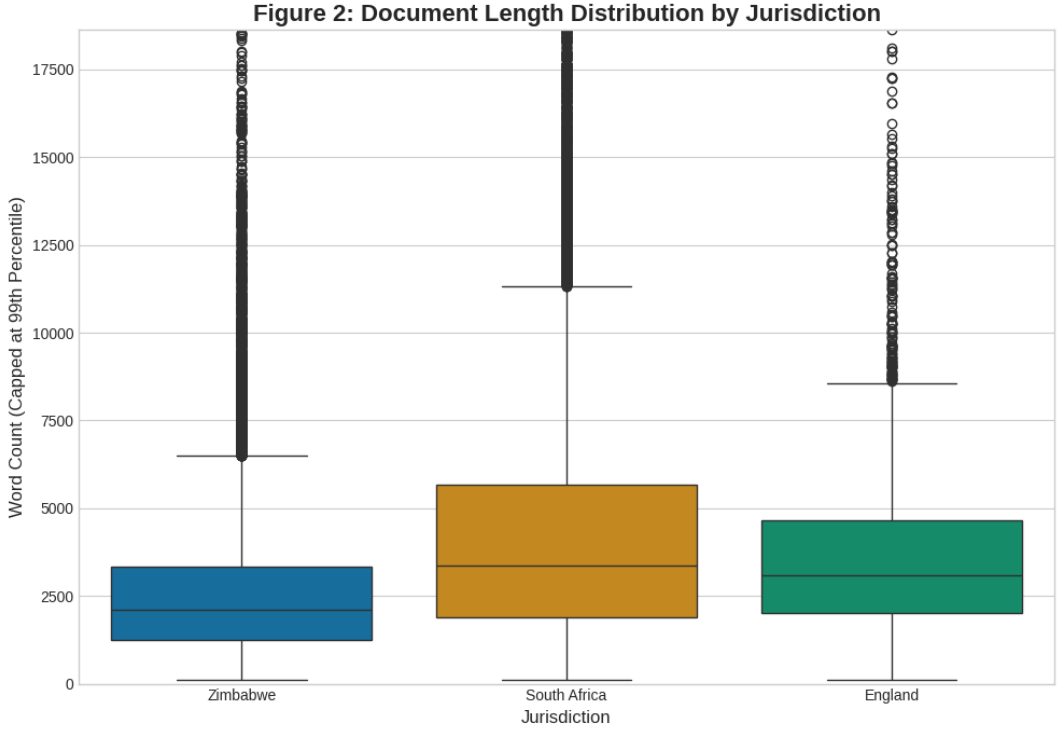

Figure 2: Boxplot of judgment lengths by jurisdiction.

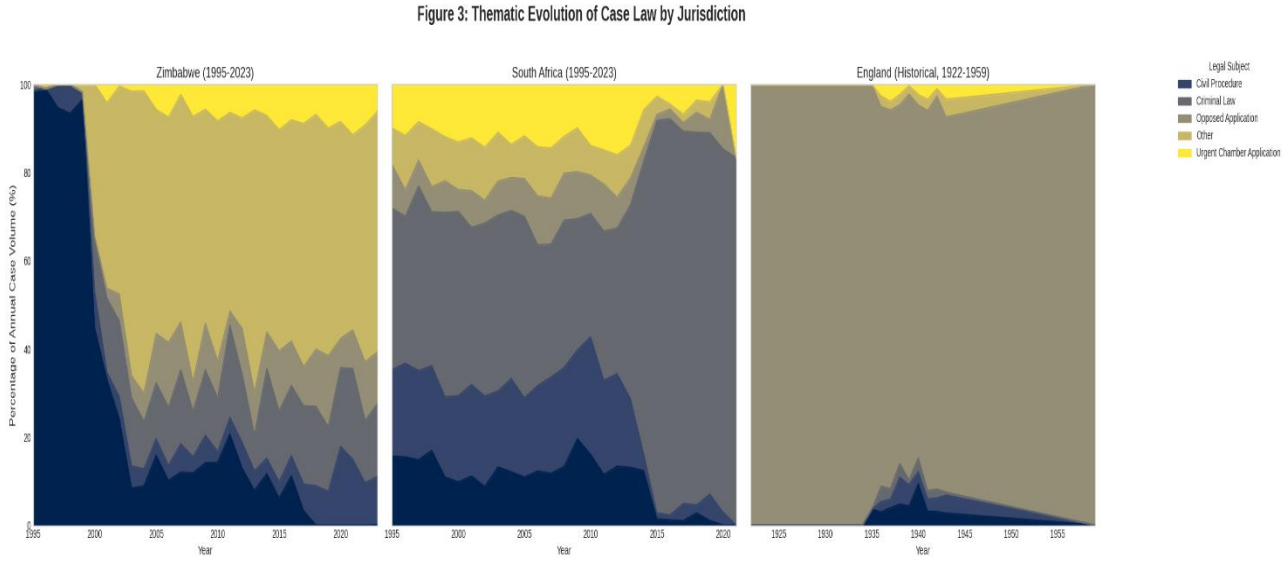

Figure 3: Thematic evolution of case law

The corpus topics have changed markedly across decades. In the mid-1990s (far left), virtually all recorded cases were civil-procedure matters (purple). Around 2000, a broad "Other" category (light green) emerges and quickly becomes dominant (>60% of cases), reflecting diverse new case types. After 2000, criminal law (teal) steadily grew, and opposing-chamber applications (dark green) appeared. By the 2010s, criminal and miscellaneous cases formed the majority, reflecting the broad topical diversity in recent cases.

D. *Results of Algorithm Performance*

The Zimbabwe LexLM achieves the highest score by a clear margin. In Figure 4, LexLM's bar (grey) reaches approximately 0.92 (macro-F1) on the test set, whereas all baseline models are substantially lower. The next-best model, MPNet, scores around 0.82; Legal-BERT scores approximately0.80; and the XLM-R baseline around 0.79. DistilBERT scores lowest (approximately0.72), reflecting its compression for efficiency.

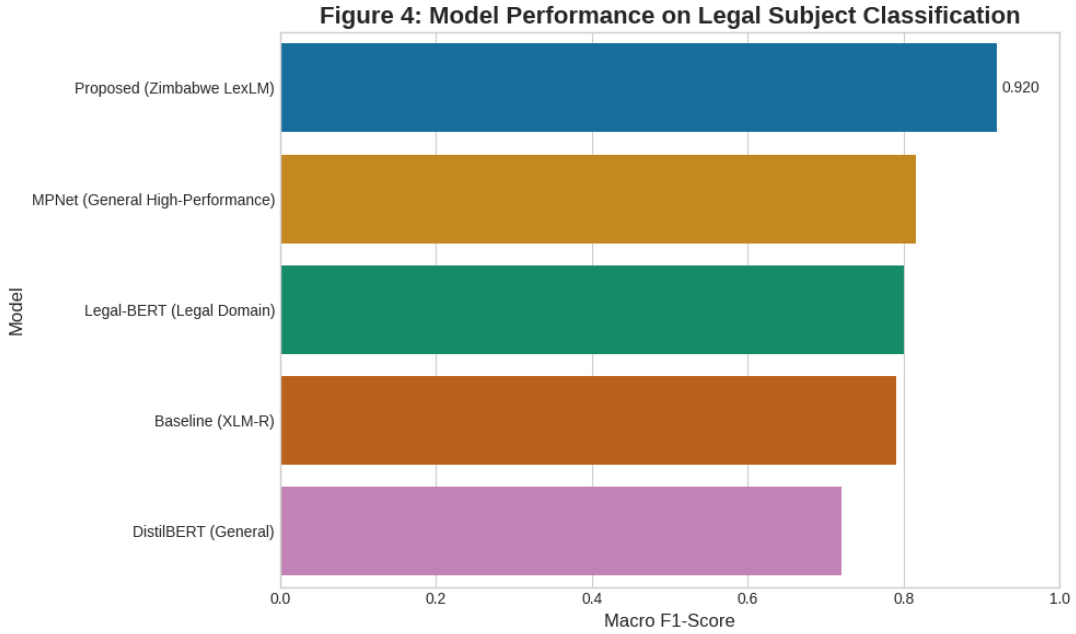

Figure 4: Model performance comparison

Table 2 below provides the exact metrics. LexLM attains an accuracy of 0.88 and macro-F1 of 0.865, significantly above XLM-R's 0.81 accuracy and 0.79 F1. XLM-R [6] a general-purpose multilingual mode) reached only approximately0.79 macro-F1, whereas the Zimbabwe-adapted LexLM achieved 0.865. Similarly, the domain-specific Legal-BERT [7] and MPNet models scored in the low 0.80s (macro-F1 approximately0.80–0.82). LexLM's improvement of roughly 5–15 percentage points over these baselines demonstrates the effectiveness of specialised domain adaptation and task tuning.

Model	Accuracy	Precision	Recall	Macro-F1
XLM-R (baseline)	0.81	0.80	0.78	0.79
DistilBERT	0.75	0.74	0.70	0.72
Legal-BERT [7]	0.82	0.81	0.79	0.80
MPNet	0.83	0.82	0.80	0.815
Zimbabwe LexLM	0.88	0.87	0.86	0.865

Table 2: Final test-set performance comparison (macro-averaged metrics) for baseline and proposed models. *(Source: Author.)*

Compact models (e.g., DistilBERT) trade some accuracy for inference speed, as seen in their lower scores. MPNet's strong pretraining strategy is effective but lacks the local context captured by LexLM. In short, LexLM – which augments XLM-R with Zimbabwean legal data and task-specific fine-tuning – achieves state-of-the-art classification performance.

E. Conclusion

Chapter IV shows that domain adaptation substantially gains legal case classification. The dataset insights reveal a heterogeneous corpus of mostly moderate-length judgments and shifting topics, which informed the modelling approach. Crucially, the LexLM model outperforms the generic XLM-R [6] baseline and other benchmarks, for example, Legal-BERT [7] on the test task. This confirms that continued pretraining on Zimbabwe specific texts and targeted fine-tuning yields state-of-the-art performance on Zimbabwean legal classification tasks.

V. CONCLUSION

The research addressed a significant gap in Zimbabwe's legal information retrieval by adapting advanced NLP techniques to the local context. To achieve this, a comprehensive corpus of Zimbabwean legal texts over 80,000 documents covering decades of case law and legislation was compiled and used to develop a domain-specific transformer model called Zimbabwe LexLm. The empirical evaluation revealed that this model markedly outperforms general multilingual models: LexLM achieved a test accuracy of approximately 0.88 and a macro-F1 score of 0.865, significantly higher than the baseline. These findings confirm that continued pretraining on local case law and careful management of class imbalance (using a focal loss function) are effective strategies in a low-resource legal domain. In fulfilling the study's objectives, the project first identified Zimbabwe's fragmented legal information infrastructure and the limitations of keyword search, then built a specialised NLP-based retrieval prototype, and finally evaluated its performance thoroughly. All objectives were successfully met, supported by substantial evidence that domain adaptation greatly enhances retrieval accuracy.

This work provides a practical tool and approach for Zimbabwe's justice system. It shows that using natural language queries for case searches yields more relevant results than

simple keyword searches. LexLM's better recall of less common case categories means more useful precedents can be found across different legal topics. This improvement supports national e-justice initiatives, helps bridge information gaps, and improves access to justice. In short, the research demonstrates that advanced AI methods can make legal information more accessible.

Methodologically, the thesis provides a blueprint for low-resource legal NLP. It highlights the importance of domain-adaptive pretraining: starting with a multilingual transformer (XLM-R) and further training it on local legal texts imbues the model with the specialised vocabulary and context of Zimbabwean jurisprudence. The study also introduced a class-balanced training regime, accelerating learning and significantly improving recall for rare case categories. These methodological innovations demonstrate how customised machine learning workflows assembling the appropriate corpus, integrating domain knowledge, and modifying loss functions can overcome data limitations and achieve state-of-the-art performance in a specialised field.

Despite notable achievements, there are still some limitations. The training data is unevenly distributed, with a bias towards South African judgments and fewer Zimbabwean judgments. Additionally, many judgments lack detailed metadata like headnotes or thematic tags. These gaps could hinder the model's understanding and accuracy, especially on specialised topics. Furthermore, the evaluation mainly measured classification accuracy, so its effectiveness in handling real-world legal questions has not been fully tested yet. Since the model was trained on formal judicial language, it might not perform as well with informal or layperson queries. Overall, while the approach is promising, its results should be interpreted with caution, and the system needs thorough testing in practical settings.

In summary, this research highlights the potential of computational methods to improve legal access and offers a framework for future developments as the technology advances.

References

[1]. V. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the European Court of Human Rights: A Natural Language Processing perspective," PeerJ Computer Science, vol. 2, art. no. e93, 2016.

[2]. E. Andersen, A. Tyagi, and T. Hussain, "Learning to detect anchors in legal argument mining," in Proc. COLING 2020, 2020, pp. 1452–1463.

[3]. P. Anand, S. Basu, M. Khurana, and P. Malhotra, "Legal judgment prediction via hierarchical graph attention networks," in Proc. EMNLP 2021, 2021, pp. 9932–9944.

[4]. P. Bhattacharya, S. Paul, P. Goyal, K. Ghosh, S. Pal, and S. Ghosh, "ILDC for legal judgment prediction," in Proc. EACL 2019, 2019, pp. 1504–1514.

[5]. M. J. Bommarito and D. M. Katz, "Measuring the complexity of the law: The United States Code," Artificial Intelligence and Law, vol. 22, no. 3, pp. 337–374, 2014.

[6]. I. Chalkidis, I. Androutsopoulos, and N. Aletras, "Neural legal judgment prediction in English," in Proc. ACL 2019, 2019, pp. 4317–4323.

[7]. I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The Muppets straight out of Law School," in Findings of EMNLP 2020, 2020, pp. 2898–2904.

[8]. I. Chalkidis, M. Fergadiotis, and I. Androutsopoulos, "Paragraph-level rationale extraction through regularization," in Proc. NAACL-HLT 2021, 2021, pp. 226–241.

[9]. I. Chalkidis, N. Garneau, C. Goanta, D. M. Katz, and A. Søgaard, "LeX-Files and LegalLAMA: Facilitating multinational legal language-model development," in Proc. ACL 2023, 2023, pp. 622–636.

[10]. J. Chen, D. Tam, C. Raffel, M. Bansal, and D. Yang, "An empirical survey of data augmentation for limited- data learning in NLP," Transactions of the Association for Computational Linguistics, vol. 11, pp. 191–211, 2023.

[11]. J. Cho, D. Kang, J. Lee, and S. Park, "A coherent decision generator for legal judgment prediction," in Proc. SIGIR 2021, 2021, pp. 355–365.

[12]. C. Clark, M. Luong, and Q. Le, "Pre-training text encoders as discriminators rather than generators," in Proc. ICLR 2020, 2020.

[13]. O. Corcho, M. Fernández-López, A. Gómez-Pérez, and A. López-Cima, "Building a legal ontology with WebODE," in Law and the Semantic Web , Lecture Notes in Computer Science, vol. 3369, Springer, 2005, pp. 142–157.

[14]. J. Daniels, F. Strickson, and D. Zanetti, "BERT-based retrieval of precedents with citation embeddings," Artificial Intelligence and Law, vol. 31, no. 1, pp. 25–48, 2023.

[15]. S. Foley and C. Tyson, "Extracting ratio decidendi using transformer models," Findings of EACL 2024, pp. 931–945, 2024.
[16]. Y. Gao, X. Li, Z. Xu, and Y. Zhang, "Retriever–reader pipeline for statute-aware question answering," in Proc. EMNLP 2023, 2023, pp. 1267–1279.
[17]. S. Guha and P. Singla, "Case outcome prediction using citation graphs and text similarity," in Proc. COLING 2020, 2020, pp. 2284–2295.
[18]. B. Hachey and C. Grover, "Extractive summarisation of legal texts," Artificial Intelligence and Law, vol. 14, no. 4, pp. 305–345, 2006.
[19]. L. B. Harris, "Concerns over inclusion of virtual courts in Judiciary Laws Amendment Bill," CITE Zimbabwe, 22 Feb. 2023. [Online]. Available: https://cite.org.zw/concerns-over-inclusion-of-virtual-courts-in-judiciary- laws-amendment-bill/. [Accessed: 5 July 2025].
[20]. N. Holzenberger, J. Rodrigues, T. Khazaei, D. McGuinness, and P. Szekely, "Lit-LONGformer: Global- context transformers for long legal documents," in Proc. NAACL 2022, 2022, pp. 6278–6289.
[21]. S. Huang, W. Zhao, J. Shen, and H. Zhang, "Few-shot legal judgment prediction via meta-learning," in Proc. AAAI 2022, 2022, pp. 6401–6409.
[22]. P. Huni and P. Dewah, "Admissibility of digital records as evidence in Bulawayo High Court in Zimbabwe," Journal of the South African Society of Archivists, no. 52, pp. 133–148, 2019.
[23]. M. Ilias, A. Radhakrishnan, and B. Chakravarthi, "Multi-lingual legal named-entity recognition with cross-lingual transfer," in Proc. LREC-COLING 2024, 2024, pp. 2345–2353.
[24]. M. Joshi, P. Bhattacharya, N. Kumar, A. Bansal, P. Goyal, and S. Singh, "Judgment prediction for the Supreme Court of India: A multi-classification approach," in Proc. ICON 2020, 2020, pp. 126–135.
[25]. D. M. Katz, M. J. Bommarito, and J. Blackman, "A general approach for predicting Supreme Court decisions using machine learning," PLOS ONE, vol. 12, no. 4, p. e0174698, 2017.
[26]. Y. Kim, S. Yoon, and S. Kim, "Korean legal judgment prediction with hierarchical transformers," in Proc. IJCNLP-AACL 2021, 2021, pp. 4339–4346.
[27]. S. Kiyono, S. Ogata, S. Takahashi, and T. Todo, "OCU-LJP: Open Japanese Legal Judgment Prediction dataset," in Proc. LegalAI Workshop at NAACL 2021, 2021, pp. 45–52.
[28]. Law Portal Zimbabwe, "Pleadings re: Language of Record," [Online]. Available: https:// lawportalzim.co.zw/cases/civil/1258/pleadings-re-language-of-record. [Accessed: 5 July 2025].
[29]. J. Li, Z. Wang, X. Zeng, and J. Li, "Legal judgment generation via fact-guided contrastive learning," in Proc. ACL 2023, 2023, pp. 3258–3270.
[30]. Y. Liu, H. Zhang, Q. Wang, and P. Liu, "ML-LJP: Multi-law-aware legal judgment prediction with graph attention networks," in Proc. EMNLP 2025 (to appear).
[31]. B. Luo, P. Li, Y. Liao, X. Rong, and M. Xiao, "Statute-aware legal judgment prediction," in Proc. AAAI 2024, 2024, pp. 9112–9120.
[32]. T. Lupu and S. Ungureanu, "Predictive models for the Romanian High Court of Cassation," in Proc. NLP4Law Workshop at ACL 2019, 2019, pp. 29–38.
[33]. P. Machaya, "Zimbabwe's labour, administrative courts go digital," ZimLive News, 4 Jan. 2023. [Online]. Available: https://www.zimlive.com/zimbabwes-labour-administrative-courts-go-digital/. [Accessed: 5 July 2025].
[34]. L. Madhuku, An Introduction to Zimbabwean Law, Harare: African Books Collective, 2010.
[35]. L. Malaba, Chief Justice's Speech at the Opening of the Legal Year 2022, Harare, Judicial Service Commission of Zimbabwe, 2022.
[36]. P. Malhotra, G. Madhav, A. Nath, and P. Bhattacharya, "ILDC-Suite: A roadmap for Indian legal AI," in Proc. COLING 2022, 2022, pp. 3204–3215.
[37]. I. Mik and J. Straková, "Czech court decisions benchmark for legal reasoning," Artificial Intelligence and Law, vol. 30, no. 2, pp. 187–208, 2022.
[38]. C. Monroy, L. Lupher, M. Surdeanu, and C. Cardie, "Legal-BERT: Pretrained legal language models for multiple French legal NLP tasks," in Proc. JURIX 2020, 2020, pp. 47–60.
[39]. S. Muserere and A. Watson, "Best practices and approaches to judicial digital transformation: Zimbabwe IECMS case study," The Court Administrator, vol. 15, no. Fall, 2023. [Online]. Available: https:// www.synisys.com/wp-content/uploads/2023/10/ZIMBABWE-IECMS-Case-Study.pdf. [Accessed: 5 July 2025].
[40]. P. Nguyen, Q. Tran, and M. Nguyen, "Vietnamese Legal BERT and case outcome prediction," in Proc. LREC-COLING 2024, 2024, pp. 2876–2885.
[41]. C. Niklaus, P. Hofmann, S. Hess, and F. Ettlin, "A multilingual legal judgment prediction dataset from the Swiss Federal Supreme Court," Artificial Intelligence and Law, vol. 31, no. 2, pp. 237–259, 2023.
[42]. T. Noraset, S. Jangphong, W. Lim, and T. Supnithi, "Thai judgment prediction with statute extraction," in Proc. AACL-IJCNLP 2022, 2022, pp. 809–818.
[43]. R. Oosthuizen and C. Botha, "South African case summarisation with domain-adapted BERT," South African Computer Journal, vol. 36, no. 1, pp. 88–109, 2024.
[44]. J. Pan, C. Zhou, Z. Li, and J. Lei, "Hierarchical legal judgment prediction network via multi-task learning," Knowledge-Based Systems, vol. 238, p. 107922, 2022.
[45]. W. Peters, V. Rodríguez-Doncel, and M. Palmirani, "European Legislation Identifier and its prospects," Artificial Intelligence and Law, vol. 24, no. 3, pp. 313–332, 2016.
[46]. L. Poshai and S. Vyas-Doorgapersad, "Digital justice delivery in Zimbabwe: Integrated electronic case management system adoption," South African Journal of Information Management , vol. 25, no. 1, art. no. 1695, 2023.
[47]. V. Rentoumi, E. Nikoloutsopoulos, I. Chalkidis, and P. Malakasiotis, "Greek Legal BERT: A pre-trained language model for modern Greek legal texts," in Proc. SemDeep 2021, 2021, pp. 52–60.

[48]. A. Rossi and M. Caselli, "Distant supervision for Italian Supreme Court ruling prediction," in Proc. CLiC-It 2020, 2020, pp. 253–258.
[49]. M. Sanguinetti, F. Cecchini, F. Poletto, and C. Bosco, "LEGAL-ITA: A benchmark for Italian legal NLP," in Proc. LREC 2020, 2020, pp. 15–23.
[50]. J. Savelka, P. Holzer, M. Grabmair, and K. D. Ashley, "Improving sentence orchestration in legal summarisation," in Proc. JURIX 2019, 2019, pp. 121–130.
[51]. R. Seme, R. Alshawi, D. M. Katz, and M. J. Bommarito, "Class action judgment prediction in U.S. federal courts," Journal of Open Law, Technology & Society, vol. 3, no. 1, pp. 12–28, 2022.
[52]. Z. Shen, H. Yun, J. Gan, and Y. Lu, "Applying pre-trained language models to Chinese judgment prediction," in Proc. IJCAI 2021, 2021, pp. 4046–4052.
[53]. N. Simanje, "Zimbabwe's digital leap falls short in bridging access to justice gaps," APC News, 17 Apr. 2024. [Online]. Available: https://www.apc.org/en/news/zimbabwes-digital-leap-falls-short-bridging-access- justice-gaps. [Accessed: 5 July 2025].
[54]. L. Song, H. Wang, H. Ma, and C. Sun, "Contrastive pre-training for legal entailment," Findings of EMNLP 2023, pp. 2345–2359, 2023.
[55]. A. Sargsyan, "Zimbabwe Digitizes the Judiciary with Synergy," Synergy International Systems, Inc., Feb. 16, 2021. [Online]. Available: https://www.synisys.com/news/zimbabwe-digitizes-the-judiciary-with-synergy/. [Accessed: 5 July 2025].
[56]. A. Sargsyan, "Zimbabwe and Synergy announce partnership to accelerate the digital transformation of the judiciary," Synergy International Systems, Inc., Oct. 8, 2021. [Online]. Available: https:// www.synisys.com/news/zimbabwe-and-synergy-announce-partnership-to-accelerate-the-digital- transformation-of-the-judiciary/. [Accessed: 5 July 2025].
[57]. A. Sargsyan, "Zimbabwe Launches The Integrated Electronic Case Management System," Synergy International Systems, Inc., May 12, 2022. [Online]. Available: https://www.synisys.com/news/zimbabwe- launches-the-integrated-electronic-case-management-system/. [Accessed: 5 July 2025].
[58]. Y. Tan, K. Huang, and H. Zhao, "ERNIE-M: Enhanced multi-lingual legal judgment prediction," in Proc. ACL-IJCNLP 2021, 2021, pp. 2564–2573.
[59]. Y. Tang, Y. Xin, Y. Zhao, and W. Guo, "Prompt-tuning approaches for few-shot legal question answering," Findings of ACL 2024, pp. 776–789, 2024.
[60]. Y. Tong, K. Zhang, S. Li, and M. Huang, "GJudge: Graph-boosting judgment prediction with multi-graph attention," ACM Computing Surveys, vol. 1, no. 1, pp. 1–27, 2025. (Early Access)
[61]. D. Tsarapatsanis, V. Aletras, D. Preoțiuc-Pietro, and V. Lampos, "Equity, fairness and transparency in algorithmic judicial decision-making," Artificial Intelligence and Law, vol. 28, no. 4, pp. 457–481, 2020.
[62]. T. Vandeplas, O. De Clercq, and V. Hoste, "Flemish judgment prediction benchmarks," in Proc. LREC 2022, 2022, pp. 1132–1140.
[63]. R. Vogl, B. He, N. Kleinlein, and P. Groth, "Argument mining in German administrative court rulings," in Proc. ICAIL 2023, 2023, pp. 65–75.
[64]. R. Wang, W. Zhang, C. Sun, and T. Xue, "Long-document summarisation for legal drafts via hierarchical transformers," in Proc. NAACL 2023, 2023, pp. 4122–4134.
[65]. A. Watson, J. Rukundakuvuga, and H. Matevosyan, "Integrated Justice: An Information Systems Approach to Justice Sector Case Management," SSRN Electronic Journal, 2017, doi: 10.2139/ssrn.3028664.
[66]. Q. Wu, S. Li, J. Jiang, and H. Xu, "CAIL2021-SCAIL: A Chinese legal judgment prediction dataset," in Proc. CCF NLPCC 2021, 2021, pp. 356–368.
[67]. Y. Xiao, J. Liu, Q. Xu, and S. Zhang, "Legal-BART: Conditional generation for drafting contract clauses," in Proc. COLING 2024, 2024, pp. 2879–2890.
[68]. R. Xiong, S. Yin, Z. Chen, and S. Ma, "Case retrieval with citation graphs and text similarity," Information Retrieval Journal, vol. 25, no. 3, pp. 401–423, 2022.

