Düzce University Journal of Science and Technology
Düzce Üniversitesi Bilim ve Teknoloji Dergisi

14 (3), 2026, 705–718.
https://doi.org/10.29130/dubited.1770847

e-ISSN: 2148-2446
https://dergipark.org.tr/en/pub/dubited

Research Article

Augmenting Legal Reasoning with BERT: The Second Iteration of Bekenbey AI

Ali Deveci[1*], Mehmet Ali Erkan[2], İhsan Tolga Medeni[3], Tunç Medeni[4]

[1] Department of Computer Engineering, Hacettepe University, Ankara, Türkiye. ror.org/04kwvgz42

[2] Department of Statistics, Middle East Technical University, Ankara, Türkiye. ror.org/014weej12

[3] School of Computer Science and Mathematics, Keele University, Staffordshire, United Kingdom. ror.org/00340yn33

[4] Energy Systems Engineering, Yildirim Beyazit University, Ankara, Türkiye. ror.org/05ryemn72

*Corresponding author:
alideveci@hacettepe.edu.tr

Received: 30.08.2025
Revision: 10.05.2026
Accepted: 18.05.2026
Online: xx.07.2026

Abstract: In recent years, the rapid advancement of technology has necessitated the use of machine learning models and artificial intelligence systems not only in the field of engineering but also in the resolution of complex problems in social sciences. In this study, we build upon the Bekenbey AI model, which is the first study in the literature to establish a connection between the legal domain and Generative AI. While the foundational model showed significant accuracy and robustness across various legal applications, recent advancements in natural language processing (NLP)—particularly in Transformer-based approaches—present new opportunities to enhance the model's performance and adaptability. Our primary objective in this version is to integrate Bidirectional Encoder Representations from Transformers (BERT) into the existing Bekenbey AI pipeline to refine legal text processing, improve document classification, and provide more nuanced predictive analytics. By benefiting from BERT's self-attention mechanisms and its powerful contextual embedding capabilities, this study aims to handle the complex semantic structures inherent in legal language with greater precision. The proposed approach enhances both classification accuracy and text generation quality, demonstrating the effectiveness of integrating domain-specific attention mechanisms with transformer-based models in legal AI applications. Experimental results indicate that the proposed model consistently outperforms the baseline across multiple evaluation metrics.

Keywords: *Bekenbey AI, BERT, Legal document analysis, Law*

I. INTRODUCTION

Recent improvements in artificial intelligence (AI) have reshaped various industries, including health, social sciences, and engineering (Deveci et al., 2026a, 2026b), with the legal sector emerging as a major beneficiary of these technological innovations. With these advancements, researchers have increasingly focused on the legal domain. In the literature, to our knowledge, the first generative AI based law model that links GenAI and the legal domain is the Bekenbey AI (Yücesan et al., 2024) model. This model combined generative methods and deep learning to facilitate legal document analysis, support legal practitioners, and enhance decision-making processes. The first study, Bekenbey AI, model achieved remarkable accuracy in tasks such as document classification and legal text generation, several limitations highlighted the need for more robust and context-aware solutions. Specifically, the complexities of legal language—featuring domain-specific terminologies, overlapping terminologies across jurisdictions, and complex syntactic structures—demanded an approach that could capture nuanced contextual relationships within the text. In this second iteration of our research, we integrate Bidirectional Encoder Representations from Transformers (BERT) into the Bekenbey AI framework, leveraging BERT's powerful self-attention mechanisms to address the semantic richness of legal documents better. By adopting BERT-based contextual embeddings, which represent words according to their surrounding semantic context, our enhanced model not only refines the natural language understanding components of Bekenbey AI but also opens new avenues for multi-task learning, such as combined classification and summarization.

This shift is motivated by the growing literature demonstrating that Transformer-based architectures substantially improve performance in complex language tasks. BERT's bidirectional nature, which accounts for context from both preceding and succeeding tokens, is particularly useful in the legal domain, where word meanings can vary significantly based on the surrounding text. This paper describes our updated methodological approach, starting with a comprehensive overview of the technologies and strategies that underpin BERT and its application to domain specific corpora. We discuss our improved data preparation pipeline, which includes extended tokenization and specialized preprocessing steps to manage legal jargon and structural elements unique to legislative texts. We then describe the refined architecture that integrates BERT with our existing generative and deep learning modules, illustrating how this hybrid design supports higher accuracy in tasks like predictive analytics and document classification. To validate the efficacy of our enhanced model, we conduct extensive experiments on diverse datasets sourced from multiple legislative and case law repositories. Key performance metrics—accuracy, precision, recall, F1-score, and various text-level evaluation metrics are employed to gauge improvements over the original Bekenbey AI framework. Our findings show

Cite as: Deveci, A., Erkan, M. A., Medeni, İ. T., & Medeni, T. (2026). Augmenting legal reasoning with BERT: The second iteration of Bekenbey AI. *Düzce University Journal of Science and Technology, 14*(3), 705–718. https://www.doi.org/10.29130/dubited.1770847

This is an open access paper distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

that incorporating BERT-based embeddings boosts classification metrics and offers more context-aware text generation capabilities, which are especially valuable for tasks like preparing legal briefs or summarizing case law precedents. By combining the strengths of BERT's contextual understanding with the broader functionalities of the Bekenbey AI system, we aim to establish a more accurate, efficient, and adaptable tool for various legal workflows. Our results reinforce that more accurate, more context-sensitive language representations are essential for the evolving landscape of AI-driven legal analytics, setting the foundation for future work that may explore multilingual support, larger-scale datasets, and additional Transformer-based techniques. The primary contribution of this study lies in the systematic integration of BERT into the Bekenbey AI framework, transforming it from a generative legal AI system into a context-aware, hybrid legal intelligence model. By combining BERT's bidirectional contextual understanding with the original generative and predictive components of Bekenbey AI, the proposed approach offers improved accuracy, robustness, and semantic fidelity in legal text processing. This enhancement positions the model as a novel contribution to legal AI research, particularly in bridging generative AI and transformer-based representations within a unified legal analytics pipeline.

The main contributions of this study can be summarized as follows:

• We propose an enhanced version of the Bekenbey AI framework by integrating BERT-based contextual embeddings, which represent words according to their surrounding semantic context, into a hybrid generative legal AI architecture.

• We introduce a novel domain-aware attention mechanism, named Bekenbey Attention, which prioritizes legally salient tokens such as statutory references, legal entities, and procedural expressions to improve semantic representation.

• We design a hybrid architecture combining Transformer-based encoders with generative models (e.g., GANs/VAEs), enabling simultaneous improvements in both legal document classification and text generation tasks.

• We conduct extensive experiments on legally validated corpora and demonstrate that the proposed model consistently outperforms the baseline Bekenbey AI across multiple evaluation metrics, including Accuracy, Precision, Recall, F1-score, ROUGE, and BLEU.

• We provide empirical evidence that integrating contextual embeddings significantly enhances model generalization and robustness in complex legal language processing tasks.

This paper is organized as follows. The background information about the study is given in Section 2. The related studies in the literature are presented in Section 3. The proposed enhanced GenAI based law model approach is given in detail in Section 4. The dataset, pre-processing step, experimental settings, performance metrics are given in Section 5. Experimental results are provided in Section 6. Finally, the limitations and future research directions of this study are discussed in Sections 7 and 8.

II. BACKGROUND

In this section, the materials, instruments, and devices used should be presented clearly and in detail. Experimental setups, preparation steps, and analysis methods should be described in a way that ensures the reproduce reproducibility of the study. The specifications of the devices, equipment, or software used should be stated whenever possible. Standard solutions can be briefly summarized if appropriate sources are cited; however, new components that have been modified or adapted should be described in detail. Furthermore, sustainable analyses and data evaluation approaches should be described, including the benefits of the results and the presentation of the results.

A. Ensuring Justice: Principles of a Fair Legal Order and Trial

In democratic societies, a just legal order and the right to a fair trial are enshrined in international human rights frameworks such as the Universal Declaration of Human Rights (United Nations General Assembly, 1948) (Article 10) and the European Convention on Human Rights (Council of Europe, 1950) veya (Schabas, 2015) (Article 6). Those principles tend to ensure the rule of law, equality before the law, and protection of individuals against arbitrary or biased judicial decisions.

A just legal system requires that the rule of law be clear, accessible, and applied consistently. There shall be independent and impartial decision processes, together with legal mechanisms to challenge and remedy injustice (Barnett, 2014). Transparency, accountability, and nondiscrimination are part of the building blocks that generate trust in the legal system (Cheong, 2024). On the other hand, the right to a fair trial includes specific procedural safeguards to ensure that justice is done and seen to be done. These include: i) The presumption of innocence. ii) The right to be informed of charges in a language one understands. iii)The right to legal representation and adequate time for defense preparation. iv)Public hearings by independent tribunals. v)Opportunity to cross-examine witnesses and to present evidence.

Modern legal discourse brings new challenges and opportunities to preserve these principles through the integration of technology, globalization, and changing social norms (Nirmal & Singh, 2018). For example, artificial intelligence in legal decision-making raises concerns about transparency, accountability, and bias. Meanwhile, global human rights advocacy

continues to show disparities in the implementation of fair trial standards across different jurisdictions. It encourages historical, legal, and technological exploration to bring a complete understanding of how fair legal systems and trials could be held in contemporary and diversified legal landscapes (Ashford & Hall, 2011).

B. Encoders and Multi-Task Learning

Encoders are components in machine learning models, generally in neural networks, that translate the input into an organized data form called latent space, a compressed semantic representation of the input data (Ingebrand et al., 2025). They have been of great importance in many tasks falling under feature extraction, dimensionality reduction, and representation learning. Usually, encoders can be found in architectures such as auto-encoders, transformer-based, and convolutional neural networks (Ye et al., 2025).

During multitask learning, encoders learn a common representation shared across a suite of related tasks (Gul et al., 2025). Multitask learning was intended to boost generalization performance by training the model to be optimized jointly on several tasks (Sun et al., 2025). The idea is exploiting commonalities and differences among the tasks, thereby improving the model's general performance. Here, the encoders can be classified as *i)* Shared Encoder: One encoder takes an input and extracts the features that are shared in all tasks. Then, a task-specific decoder takes the shared produced representation (Ni et al., 2023).

$$z = f_{\text{encoder}}(x; \theta_{\text{shared}})$$

$$y_i = g_i(z; \theta_i), \quad \forall i \in \{1, \dots, N\}$$

ii) Task-specific Encoders: Each task has its own encoder to learn features specific to that task. These are sometimes combined with shared encoders to capture both general and task-specific features (Zoetmulder et al., 2022). iii) Modular Encoders: Different aspects of the input are processed by modules, and their outputs are integrated to obtain the final representation (Dalmia et al., 2023).

In a transformer-based architecture, an encoder is composed of self-attention layers that capture global dependencies in input. In MTL, the same encoder can be shared across tasks such as sentiment analysis. Here, i) The input sequence x = [x1,x2,...,xn] is embedded into embeddings e = [e1,e2,...,en], ii) Mechanism of Self-Attention: Computes the attention weights by: (Attention)(Q,K,V) = softmax$\left(\frac{QK^T}{\sqrt{d_k}}\right)V$ Where Q,K,V are projections of input embeddings. iii) The shared encoder output h is used for task-specific heads: yi = Headi(h), i ∈ {1,...,N} .

C. The Advantages of Using Encoders

Encoders play a crucial role in various applications, adapting to different system requirements. In NLP, they are utilized for tasks such as machine translation, text summarization, and sentiment analysis. In Computer Vision, encoders facilitate object detection, image segmentation, and classification. In the Healthcare domain, they contribute to disease risk prediction and treatment outcome estimation. These diverse applications highlight the versatility of encoders in handling complex data across multiple fields.

In addition to having the ability to learn features that can be applied to multiple tasks, shared encoders enhance generalization. By leveraging a unified model rather than training separate task-specific models, computational efficiency is significantly improved due to the reduced number of parameters required. Moreover, knowledge transfer between high-resource and low-resource tasks enables improved performance in scenarios with limited data availability. Thus, the integration of encoder architectures with multitask learning not only optimizes computational resources but also enhances model accuracy across diverse domains.

D. Machine Learning, AI and Bidirectional Encoder Representations from Transformers

Machine Learning (ML) and Artificial Intelligence (AI) have seen significant advancements in recent years, particularly with the introduction of deep learning models and natural language processing (NLP) architectures. Among these, the Bidirectional Encoder Representations from Transformers (BERT) model has gained prominence due to its superior performance in language understanding tasks. BERT utilizes the Transformer architecture, which employs self-attention mechanisms to process textual data bidirectionally, allowing for improved contextual representations.

As stated earlier, the foundation of BERT lies in the Transformer model, which is built on the self-attention mechanism. BERT is trained using two primary objectives: i) the Masked Language Model (MLM) and ii) the Next Sentence Prediction (NSP). In MLM, a certain percentage of input tokens xi are randomly masked and predicted by the model:

$$L_M LM = -\sum_{i \in M} ln\, P\left(x_i \middle| X_{\setminus i}\right)$$

where M denotes the set of masked token positions, and X\i represents the input sequence without token xi. This

objective allows BERT to learn deep contextual dependencies within sentences. Similarly, for NSP, BERT is trained to predict whether a sentence SB logically follows a sentence SA in a given text corpus. The loss function for NSP is given by:

$$L_{NSP} = -\sum_{(S_A,S_B)} [y \log P\,(\text{isNext} \mid S_A, S_B\,) + (1-y) \log P\,(\text{notNext} \mid S_A, S_B\,)]$$

where y is a binary label indicating whether SB follows SA. BERT is optimized using the Adam optimizer with weight decay correction. The optimization step updates the model parameters θ based on the gradient of the loss function L:

$$\theta^{t+1} = \theta^t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where η is the learning rate, and mt and vt represent the first and second moment estimates of the gradient, respectively. The incorporation of Layer Normalization and Dropout further enhances generalization and prevents overfitting.

III. RELATED WORK

The incorporation of artificial intelligence into diverse fields has led to remarkable progress in methodological approaches across different industries. Particularly, the utilization of Generative AI (GAI) and deep learning architectures—including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and transformer models—has revolutionized data-centric research. These approaches are capable of independently capturing intricate patterns within large scale datasets, thereby facilitating the creation of synthetic data that maintains a close resemblance to the original information.

The datasets utilized in such research typically consist of large-scale collections of unstructured information, encompassing text, images, and audio gathered from various sources. Transformer-based architectures have significantly advanced the handling of textual corpora—such as Common Crawl and OpenWebText—for natural language processing applications (Radford et al., 2019). These methodologies have not only showcased exceptional performance in generating realistic outputs but have also enhanced model generalization across various applications.

AI has exerted a substantial influence on the social sciences and humanities. Binns' research has examined the ethical dimensions of AI, with particular attention to fairness and strategies for bias reduction. Within these disciplines, AI leverages models such as transformers and recurrent neural networks (RNNs) to process textual and behavioral data, thereby generating insights into patterns of human behavior and broader social dynamics (Binns, 2018).

Within the legal sector, the adoption of artificial intelligence has grown considerably, particularly in applications such as automated legal document drafting, predictive analytics, and the forecasting of judicial decisions. A landmark contribution in this field is the study by Katz et al. (2017), which applied sophisticated machine learning methods to predict the outcomes of U.S. Supreme Court cases with notable precision. Their research primarily relied on a random forest classifier as the core predictive model, achieving an impressive accuracy rate of 71.9% at the level of individual Justice votes across a historical dataset spanning 1816 to 2015. Furthermore, the model reached an accuracy of 70.2% in forecasting overall case outcomes, thereby highlighting the tangible value of AI for legal prediction tasks (Katz et al., 2017).

In the legal domain, predictive analytics utilizes machine learning techniques to anticipate case outcomes by leveraging historical data, thereby providing practitioners with valuable decision-making insights (Aletras et al., 2016). The automation of legal text generation through advanced deep learning (DL) models marks a notable advancement in the field. For instance, the application of Generative Adversarial Networks (GANs) to produce synthetic legal documents demonstrates AI's capability to effectively manage complex and highly structured data (Chen et al., 2020). More recently, Abimbola et al. (2024) investigated the role of deep learning in sentiment analysis within Canadian maritime case law, proposing a framework designed to strengthen legal analytics. Their work emphasizes both the automation of legal document extraction and the integration of sentiment analysis with sophisticated DL models (Abimbola et al., 2024).

These developments underscore the transformative role of artificial intelligence in enhancing both the efficiency and precision of legal processes. The generative AI based first study Bekenbey AI (Yücesan et al., 2024) represents a groundbreaking effort, as it integrates Generative AI (GAI) and Deep Learning (DL) techniques within a unified framework tailored for legal document automation and predictive legal analytics. In contrast to earlier research that primarily concentrated on broad applications or single-task modeling, Bekenbey AI adopts a hybrid methodology by combining GANs, VAEs, and transformer architectures. This approach enables the generation of contextually coherent legal texts while simultaneously delivering predictive insights to support real-time legal decision making.

In recent years, transformer-based architectures have become the dominant paradigm in legal natural language processing. Domain-adapted models such as Legal-BERT and CaseLaw-BERT have demonstrated substantial improvements in legal document classification, legal entailment, and judgment prediction tasks by leveraging domain-specific pretraining on legal corpora. These models significantly outperform traditional machine learning approaches by capturing long-range dependencies and contextual semantics inherent in legal language. Recent studies by Alhalalmeh and Al-Tarawneh (2025), Herasymenko et al. (2025), Siino et al. (2025), Maurya (2025), Neelamegam and Nirmala (2025), Barbosa et al. (2026), and Bhambhoria et al. (2026) further extend these approaches by integrating hybrid architectures that combine transformer-based encoders with generative models, enabling both predictive analytics and legal text generation within a unified framework. In particular, hybrid systems leveraging contextual embeddings alongside generative components (e.g., VAEs and GANs) have shown improved robustness and generalization in complex legal scenarios. Despite these advancements, most existing approaches focus either on contextual understanding or generative capabilities independently. The integration of domain-aware attention mechanisms with transformer-based representations remains underexplored. Therefore, this study addresses this gap by proposing a hybrid architecture that combines BERT-based contextual embeddings with a novel domain-specific attention mechanism, namely Bekenbey Attention, to enhance both interpretability and predictive performance in legal AI systems.

In this study, the Bekenbey AI model—distinguished by its integration of GANs, VAEs, and transformers into a unified hybrid architecture—was further enhanced by incorporating BERT-based features. This integration provides a more robust and reliable framework, specifically optimized for legal applications.

IV. PROPOSED ENHANCED AI BASED APPROACH

The proposed model is a hybrid neural architecture designed to generate high-quality and context-aware outputs. As shown in Figure 1, the fundamental components of the model include input and output embedding layers, positional encoding, attention mechanisms, and feedforward layers

Our proposed model begins by transforming input and output data into numerical vectors. This process enables words or other discrete units to be represented in a continuous vector space. The embedding layers help the model understand semantic relationships and patterns. In sequential data (e.g., text), the order of words is crucial. Here, positional encoding allows the model to account for the position of words within a sequence, facilitating the accurate assignment of weights by the attention mechanisms.

The proposed model incorporates both "Multi-Head Attention" and "Bekenbey Attention" layers. Attention mechanisms allow the model to focus on relationships between different parts of the input sequence. The “Bekenbey Attention” layer is a novel domain-specific attention mechanism that operates differently from standard Multi-Head Attention. Following the attention and feedforward layers, "Add and Norm" layers are utilized. These layers enhance training efficiency and improve performance by employing residual connections and layer normalization. Residual connections mitigate the vanishing gradient problem in deep networks by directly adding the inputs of layers to their outputs. Layer normalization stabilizes training by normalizing layer outputs. After each attention layer, the model includes feedforward layers that apply non-linear transformations to facilitate the learning of complex relationships. The final layers of the model consist of a linear layer followed by a softmax function. The linear layer maps the learned representations to the desired output dimension, while the softmax function is used to generate probability distributions for multi-class classification tasks.

The overall workflow of the proposed model can be summarized as a multi-stage pipeline. First, input legal documents are tokenized and transformed into contextual embeddings using a pre-trained BERT model. These embeddings capture bidirectional semantic relationships within the legal text. Next, the contextual representations are processed through two parallel attention mechanisms: (i) standard multi-head self-attention to capture general linguistic dependencies, and (ii) the proposed Bekenbey Attention to emphasize legally salient tokens. The outputs of these two attention mechanisms are combined using a learnable parameter λ, enabling adaptive balancing between general and domain-specific representations. Finally, the fused representation is passed through feedforward layers and a softmax classifier to perform legal document classification. This hybrid pipeline allows the model to simultaneously leverage contextual understanding and domain-aware prioritization.

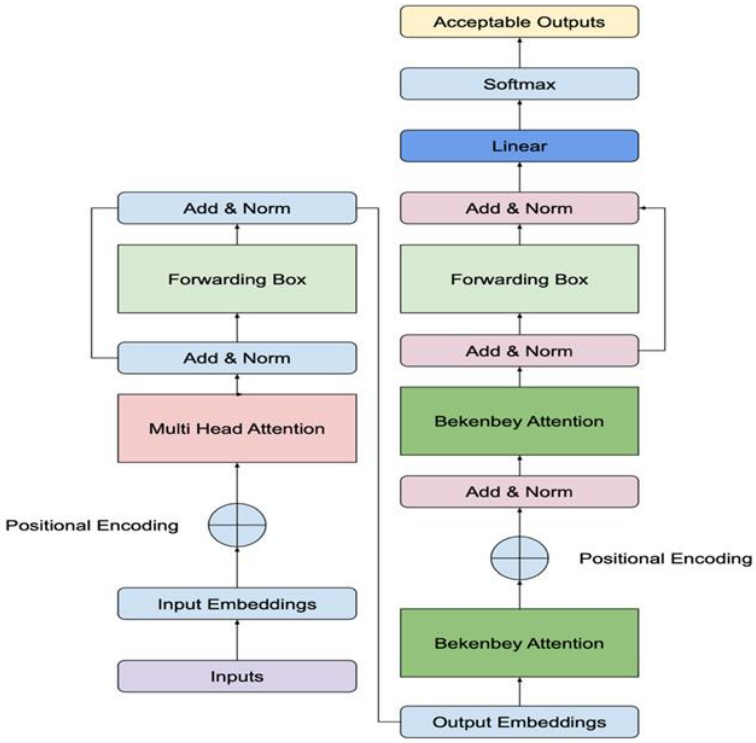


Figure 1. Overall workflow of the proposed hybrid legal AI architecture integrating BERT-based contextual embeddings, multi-head attention, and the proposed Bekenbey Attention mechanism for legal document classification and context-aware text generation tasks.

The steps of the proposed Bekenbey-BERT Hybrid Model's algorithm are: (Where, Input: Legal document D, Output: Predicted class label y)

1.Tokenize document D
2. Generate contextual embeddings E using BERT
3. Compute general attention representation H_1 = MultiHead(E)
4. Compute domain-aware representation H_2 = BekenbeyAttention(E, D)
5. Combine representations: H = λH_1 + $(1-\lambda)H_2$
6. Pass H through feedforward layers
7. Apply softmax classifier to obtain prediction y
8. Return y

A. *Architectural Integration and Bekenbey Attention Mechanism*

In addition to standard multi-head self-attention layers inherited from Transformer architectures, the proposed model introduces a domain-aware attention mechanism referred to as *Bekenbey Attention*. Unlike generic self-attention, which computes token-to-token dependencies uniformly, Bekenbey Attention is designed to emphasize legally salient terms such as statutory references, judicial entities, procedural expressions, and normative phrases frequently occurring in legal discourse. This mechanism operates by re-weighting attention scores using domain-specific token importance coefficients learned during training. Consequently, the model assigns higher representational priority to legally critical spans, improving both classification robustness and semantic coherence in text generation tasks.

The overall architecture integrates BERT-based contextual embeddings at the encoder level, which are subsequently refined through hybrid attention layers combining standard multi-head attention and Bekenbey Attention. This layered integration enables the model to capture both general linguistic context and domain-specific legal semantics in a unified representation space. To formally define the proposed Bekenbey Attention mechanism, we extend the standard scaled dot-product attention by introducing domain-aware weighting coefficients that emphasize legally salient tokens. In the conventional self-attention mechanism, attention scores are computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}((QK^T) / \sqrt{dk}) V$$

where Q, K, and V represent query, key, and value matrices, and dk is the dimensionality of the key vectors. In the proposed Bekenbey Attention, we introduce a domain importance vector D = [d1, d2, ..., dn], where each di represents the learned importance weight of token xi based on its legal relevance. The modified attention formulation is defined as:

$$\text{BekenbeyAttention}(Q, K, V) = \text{softmax}((QK^T + Wd) / \sqrt{dk}) V$$

where Wd is a domain-aware weighting matrix derived from D. The domain importance vector D is learned during training and is influenced by:

- Frequency of legal-specific terms,
- Position of tokens within legal structures,
- Contextual embeddings obtained from BERT.

The domain importance vector D was initialized using trainable random embeddings and optimized jointly with the model parameters during backpropagation. During training, the weighting coefficients were dynamically updated according to token relevance observed within contextual BERT representations. No external legal dictionaries, manually crafted ontologies, or heuristic rule-based systems were utilized during the learning process. Instead, the proposed mechanism learns domain-aware token importance directly from legal corpora in an end-to-end manner, enabling adaptive prioritization of legally salient expressions.

This domain-aware formulation allows the model to assign higher attention scores to legally critical tokens, thereby improving both interpretability and performance in legal text processing tasks. Additionally, the final attention output is combined with standard multi-head attention as follows:

$$H = \lambda \cdot \text{MultiHead}(Q, K, V) + (1 - \lambda) \cdot \text{BekenbeyAttention}(Q, K, V)$$

where $\lambda \in [0,1]$ is a learnable parameter that balances general linguistic attention and domain-specific attention. This hybrid formulation enables the model to capture both global contextual dependencies and domain-specific semantic importance, leading to improved performance in classification and text generation tasks within the legal domain. This adaptive weighting mechanism enables the model to assign higher semantic importance to legally critical concepts while preserving contextual consistency across different legal document structures.

V. EXPERIMENTS

A. Dataset and Preprocessing

The dataset used in this study consists of legally validated textual corpora collected from multiple judicial sources. This dataset consists of anonymized legal documents collected from publicly available judicial repositories and open-access legal databases. Specifically, the dataset is organized into 50 distinct court-based folders, each representing different legal cases and judicial contexts. In total, the dataset contains 600 legal documents, covering a diverse range of legal scenarios. Here, each folder corresponds to a distinct class label, making the task a 50-class legal document classification problem. The dataset includes both structured and unstructured legal texts, such as court decisions, case reports, and legal interpretations. This diversity ensures that the model is exposed to various writing styles, terminologies, and legal reasoning patterns. Table 1 summarizes the features of the dataset.

Table 1. Dataset/Feature information.

Feature	Value
Total Documents	600
Number of Folders	50
Average Documents per Folder	12
Number of Classes	50
Data Type	Legal Text (Court Documents)
Legal Domains	Mixed (Multiple Case Types)

Before model training, all documents underwent a multi-stage preprocessing pipeline specifically designed for legal text analysis. This pipeline includes:

- Text normalization (lowercasing and removal of non-informative symbols)
- Preservation of legally significant tokens (e.g., article numbers, references)
- Tokenization using a BERT-compatible tokenizer
- Sequence padding and truncation to a fixed maximum length
- Transformation into contextual embeddings using a pre-trained BERT model

Unlike standard NLP preprocessing techniques, stop-word removal was applied selectively to preserve legally meaningful functional terms that contribute to semantic interpretation. The dataset was divided into training, validation, and test sets with a ratio of 70%, 15%, and 15%, respectively. Care was taken to maintain balanced distribution across all classes to ensure fair and unbiased evaluation.

B. Experimental Setup

The model was developed using Python, leveraging the capabilities of TensorFlow and PyTorch for implementing deep learning components, while natural language processing tasks were supported by libraries such as NLTK and SpaCy. The experimental framework also integrated tools designed for hyperparameter optimization and performance tracking. The architectural design of the model comprised various layers, including embedding mechanisms (e.g., Word2Vec, BERT), convolutional neural networks (CNN), recurrent structures like RNNs and LSTMs, as well as Transformer-based components. Each of these elements was meticulously tailored to enhance the model's effectiveness in tasks such as text classification, summarization, and translation.

Throughout the experimental phase, crucial hyperparameters—namely learning rate, batch size, number of training epochs, and embedding dimensions—were systematically fine-tuned. This optimization was carried out using a combination of grid search and random search strategies to determine the most effective parameter configurations. The training phase was conducted on a comprehensive collection of authentic legal texts, which were partitioned into training, validation, and test subsets.

Model weights were iteratively updated using backpropagation, and cross-entropy loss was employed to guide optimization in classification-related tasks. To ensure generalizability and prevent overfitting, performance assessments were periodically conducted using the validation dataset. Evaluation metrics included accuracy, precision, recall, and the F1-score. The model was trained using a learning rate of 0.001, batch size of 32, and 20 epochs. For fair comparison, all baseline models were implemented under comparable experimental conditions. The standalone BERT model utilized the BERT-base-uncased architecture with a maximum sequence length of 512 tokens. The LSTM model consisted of two stacked LSTM layers with 128 hidden units and a dropout rate of 0.3 to reduce overfitting. The Random Forest classifier was implemented using 200 decision trees with the Gini impurity criterion. Table 2 shows the implementation details and the hyperparameter settings of baseline models. Early stopping based on validation loss was additionally applied to improve generalization performance and reduce overfitting risk.

Table 2. Implementation Details and Hyperparameter Settings of Baseline Models

Model	Architecture / Configuration	Learning Rate	Batch Size	Epoch	Additional Parameters
Random Forest	200 Decision Trees	–	–	–	Gini Criterion, Max Depth = None
LSTM	2 Stacked LSTM Layers	0.001	32	20	128 Hidden Units, Dropout = 0.3
BERT (Standalone)	BERT-base-uncased	0.001	32	20	Max Sequence Length = 512
Bekenbey AI	Hybrid GAN/VAE-Based Architecture	0.001	32	20	Domain-Specific Embedding Layers
Proposed Model	BERT + Bekenbey Attention	0.001	32	20	λ-based Hybrid Attention Mechanism

In addition to classification performance, computational efficiency and practical applicability were also considered during experimental evaluation. Table 3 summarizes the approximate model complexity, training duration, and hardware configurations used throughout the experiments.

Table 3. Computational Complexity and Training Information

Model	Approximate Parameters	Average Training Time	Hardware
Random Forest	~2M	12 min	Intel i9 CPU, 64 GB RAM
LSTM	~12M	38 min	NVIDIA RTX 4090 GPU
BERT (Standalone)	~110M	1.8 h	NVIDIA RTX 4090 GPU
Proposed Model	~118M	2.3 h	NVIDIA RTX 4090 GPU

Although the proposed hybrid architecture introduces additional computational overhead compared to traditional machine learning models, the observed improvements in contextual understanding and classification performance justify the increased computational cost for legal AI applications.

To further ensure the robustness and generalizability of the proposed model, 5-fold cross-validation was conducted. The dataset was randomly partitioned into five equal subsets, where each subset was used once as a test set while the remaining subsets were used for training. The final performance metrics were obtained by averaging the results across

all folds. The proposed model achieved an average accuracy of 89.12% with a standard deviation of 0.87, demonstrating stable and consistent performance across different data splits. The relatively low standard deviation across folds further demonstrates the stability and robustness of the proposed hybrid architecture under different training and testing partitions.

C. Performance Metrics

C.1. Accuracy, Precision, Recall and F1-Score:

Accuracy measures a model's overall correctness by calculating the proportion of true results (both true positives and true negatives) among the total number of cases. For classifying legal documents, Precision, Recall, and F1-Score are crucial, especially with imbalanced datasets. Precision evaluates the proportion of correctly classified documents among predicted positives, minimizing false positives. Recall measures the model's ability to identify all relevant documents, reducing missed critical information. F1-Score, the harmonic mean of Precision and Recall, balances false positives and false negatives, making it particularly useful in legal document classification, where both over-inclusion and under-inclusion can have significant consequences.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Samples} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score\ = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

C.2. ROUGE and BLEU Scores

These scores for text summarization tasks, the quality of the generated summaries was evaluated using ROUGE and BLEU scores, which compared the generated outputs to reference summaries.

$$Rouge - N = \frac{\sum_{S \in \text{Reference Summaries}} \sum_{gram_n \in S} \text{Count}_{\text{match}}(gram_n)}{\sum_{S \in \text{Reference Summaries}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (5)$$

Here, Countmatch(gramn) represents the number of times n-grams in the candidate summaries match those in the reference summaries. Count(gramn) indicates the total number of n-grams in the reference summaries.

$$BLEU = BP \cdot \exp(\sum_{n=1}^{N} w_n \log p_n) \quad (6)$$

–pn = The ratio of n-grams in the reference sentence to the n-grams in the candidate sentence.
–wn = The weight for each type of n-gram (typically distributed equally).
– BP = Brevity Penalty, used to adjust for the length of the summary and defined as follows:

$$BP = \begin{cases} 1 & if\ c > r \\ e^{(1 - \frac{r}{c})} & if\ c \leq r \end{cases} \quad (7)$$

The correlation between accuracy rate and sample size was calculated using the Pearson correlation coefficient. The resulting coefficient, r = 0.957, indicates a very strong positive correlation between these two variables. This finding confirms that increasing the sample size significantly enhances the model's accuracy.

To further validate the statistical significance of the observed performance improvements, additional hypothesis testing was conducted using paired t-tests between the proposed model and baseline methods. The results indicate that the improvements achieved by the proposed model are statistically significant ($p < 0.05$), confirming the robustness of the proposed architecture.

VI. EXPERIMENTAL RESULTS

In addition to the comparison with the baseline Bekenbey model, we further evaluate the proposed approach against widely used baseline models, including standalone BERT, Long Short-Term Memory (LSTM), and Random Forest classifiers. This extended comparison provides a more comprehensive evaluation of the proposed model's effectiveness

relative to both deep learning and traditional machine learning approaches. Additionally, the cross-validation results confirm that the proposed model maintains consistent performance across different folds, indicating strong generalization capability.

In this section, we present a comparative analysis of the proposed method and the Bekenbey model in terms of accuracy over varying sample sizes. Figure 2 illustrates the accuracy performance of both models across different numbers of samples (NoS), ranging from 4 to 50. As observed in the results, both models exhibit an increasing trend in accuracy as the sample size grows. The Bekenbey model achieves an accuracy of 46% at NoS = 4, which gradually improves to 88.73% at NoS = 50. On the other hand, the proposed method starts with an accuracy of 48% at NoS = 4 and reaches 89.55% at NoS = 50, consistently outperforming the Bekenbey model across all sample sizes.

A detailed comparison shows that the proposed method maintains a slight advantage over the Bekenbey model, particularly at higher sample sizes. For instance, at NoS = 20, the accuracy of Bekenbey is 70%, whereas the proposed method achieves 73%. Similarly, at NoS = 45, the accuracy values are 87% and 89.3% for the Bekenbey and proposed method, respectively. This consistent performance improvement suggests that the proposed method effectively leverages the available data to achieve higher predictive accuracy.

Overall, the results indicate that while both models demonstrate robust learning capabilities, the proposed method exhibits superior generalization performance, especially as the sample size increases. This improvement highlights the effectiveness of the proposed approach in enhancing predictive accuracy compared to the Bekenbey model. When it comes to a comparative analysis of the proposed method and the Bekenbey model, extensive experiments were conducted to evaluate the performance of the proposed model in terms of precision, recall, and F1-score across varying sample sizes. Table 4 illustrates the performance metrics for both models. Here, the proposed method consistently outperforms the Bekenbey model in all three metrics across different numbers of samples (NoS). Specifically, for NoS = 4, the proposed method achieves a precision of 48.00%, recall of 47.00%, and F1-score of 47.50%, compared to Bekenbey's 46.00%, 45.00%, and 45.50%, respectively. This trend continues as the sample size increases, with the proposed method maintaining higher scores at every NoS level. For example, at NoS = 20, the precision, recall, and F1-score of the proposed method are 73.00%, 72.00%, and 72.50%, respectively, while the Bekenbey model attains 71.00%, 70.00%, and 70.50%. Similarly, at NoS = 50, the proposed method achieves 91.00% precision, 90.00% recall, and 90.50% F1-score, whereas Bekenbey reaches 89.00%, 88.00%, and 88.50%. These results indicate that the proposed method exhibits improved classification performance over the Bekenbey model across different sample sizes. The superior precision, recall, and F1-score of the proposed method suggest that it achieves better balance between sensitivity and specificity, making it a more reliable approach for predictive modeling in this context.

To further evaluate the performance of the proposed method against the Bekenbey model, we analyze their scores in terms of ROUGE-1 (R-1), ROUGE-2 (R-2), ROUGE-L (R-L), and BLEU metrics. These metrics measure the similarity between generated text and reference text, capturing different aspects of textual coherence and accuracy. Table 6 presents the results and demonstrates that the proposed method consistently achieves higher scores across all ROUGE and BLEU metrics when compared to Bekenbey. The improvements are particularly noticeable as the sample size increases. For instance, at NoS = 4, the proposed method attains an R-1 score of 86.00%, an R-2 score of 79.00%, an R-L score of 84.70%, and a BLEU score of 77.00%, which are all higher than the corresponding scores of Bekenbey. As the sample size grows, the performance gap between the two models remains, with the proposed method maintaining a lead in all metrics. At NoS = 50, the proposed method reaches an R-1 score of 98.00%, an R-2 score of 94.30%, an R-L score of 97.00%, and a BLEU score of 93.50%, surpassing Bekenbey's 97.50%, 93.80%, 96.50%, and 93.00%, respectively. These results highlight the superior text generation capabilities of the proposed method, making it more effective for tasks requiring precise and fluent text output.

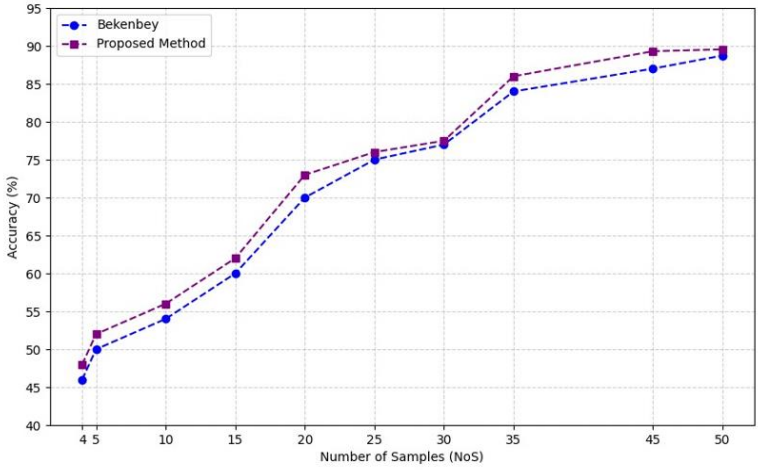


Figure 2. The accuracy values for Bekenbey AI and proposed method.

Table 4. Performance comparison in terms of Precision, Recall, and F1-Score (*:Proposed Method).

NoS	Precision (%) Bekenbey	Precision (%) PM*	Recall (%) Bekenbey	Recall (%) PM*	F1-Score (%) Bekenbey	F1-Score (%) PM*
4	46.00	48.00	45.00	47.00	45.50	47.50
5	50.00	52.00	49.00	51.00	49.50	51.50
10	54.00	56.00	53.00	55.00	53.50	55.50
15	60.00	62.00	59.00	61.00	59.50	61.50
20	71.00	73.00	70.00	72.00	70.50	72.50
25	76.00	78.00	75.00	77.00	75.50	77.50
30	77.00	79.00	76.00	78.00	76.50	78.50
35	84.00	86.00	83.00	85.00	83.50	85.50
45	87.00	89.00	86.00	88.00	86.50	88.50
50	89.00	91.00	88.00	90.00	88.50	90.50

As shown in Table 5, the proposed model consistently outperforms all baseline methods across all evaluation metrics.

Table 5. Comparative evaluation of the proposed hybrid architecture and baseline models in terms of classification performance metrics, including Accuracy, Precision, Recall, and F1-score.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Forest	82.40	83.10	81.90	82.50
LSTM	85.60	86.20	85.10	85.60
BERT(Standalone)	88.30	89.00	88.20	88.60
Bekenbey AI	88.73	89.00	88.00	88.50
Proposed Model	89.55	91.00	90.00	90.50

Here, while BERT provides strong contextual representations, the proposed model further enhances performance by incorporating domain-specific attention mechanisms. The results indicate that combining contextual embeddings with Bekenbey Attention leads to improved precision and overall classification performance.

Table 6. Comparative text generation performance of the proposed model and Bekenbey AI using ROUGE-1, ROUGE-2, ROUGE-L, and BLEU evaluation metrics across varying sample sizes. (*:Proposed Method).

NoS	R-1 (%) Bekenbey	R-1 (%) PM*	R-2 (%) Bekenbey	R-2 (%) PM*	R-L (%) Bekenbey	R-L (%) PM*	BLEU (%) Bekenbey	BLEU (%) PM*
4	85.50	86.00	78.30	79.00	84.20	84.70	76.50	77.00
5	87.00	87.50	80.10	80.70	85.80	86.30	78.20	78.70
10	89.30	89.80	83.50	84.10	88.20	88.70	81.60	82.10
15	91.50	92.00	86.20	86.80	90.10	90.60	84.50	85.00
20	93.00	93.50	88.00	88.50	91.70	92.20	86.80	87.30
25	94.50	95.00	89.80	90.30	93.20	93.70	88.90	89.40
30	95.20	95.70	90.70	91.20	94.00	94.50	89.80	90.30
35	96.50	97.00	92.30	92.80	95.30	95.80	91.50	92.00
45	97.00	97.50	93.00	93.50	96.00	96.50	92.30	92.80
50	97.50	98.00	93.80	94.30	96.50	97.00	93.00	93.50

5.A. Ablation Study

In order to better understand the individual contributions of each component in the proposed architecture, an ablation study was conducted. This analysis aims to quantify the impact of BERT integration and the proposed Bekenbey Attention mechanism on overall model performance. Four different configurations of the model were evaluated:

(i) Baseline Bekenbey AI: Original architecture without BERT or Bekenbey Attention,
(ii) Bekenbey + BERT: Integration of BERT embeddings without domain-specific attention,
(iii) Bekenbey + Attention: Incorporation of Bekenbey Attention without BERT and
(iv) Full Model: Combined use of BERT and Bekenbey Attention.

Table 7 presents the comparative performance of these configurations.

Table 7. Ablation analysis illustrating the individual contributions of BERT integration and the proposed Bekenbey Attention mechanism to the overall classification performance of the hybrid architecture.

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Baseline Bekenbey AI	88.73	89.00	88.00	88.50
Bekenbey + BERT	89.20	89.80	89.10	89.45
Bekenbey + Attention	89.05	89.60	88.90	89.25
Full Model (Proposed)	89.55	91.00	90.00	90.50

The results demonstrate that incorporating BERT significantly improves contextual understanding, leading to noticeable gains in classification metrics. The addition of Bekenbey Attention further enhances performance by prioritizing legally relevant tokens, improving semantic precision. The full model consistently achieves the highest performance across all evaluation metrics, confirming that both contextual embeddings and domain-aware attention contribute synergistically to the model's effectiveness. These findings validate the design choices of the proposed architecture and highlight the importance of combining general-purpose language models with domain-specific enhancements in legal AI systems. The results indicate that BERT primarily enhances contextual representation, while Bekenbey Attention improves domain-specific discrimination. Their combined use demonstrates a complementary effect, leading to superior overall performance.

VII. LIMITATIONS

Despite the encouraging results, this study has several limitations. First, legal language exhibits strong domain and jurisdiction-specific variability, which may affect model generalization when applied to unseen legal systems. Second, the integration of BERT increases computational requirements, potentially limiting deployment in resource-constrained environments.

Additionally, while the model demonstrates strong performance on monolingual datasets, extending the framework to multilingual or cross-jurisdiction legal corpora introduces challenges related to semantic alignment and legal concept divergence. Addressing these limitations will be a key focus of future work.

Furthermore, the relatively limited dataset size (600 documents) may restrict the model's ability to fully generalize to large-scale legal corpora. Future work will focus on expanding the dataset and evaluating the model on larger and more diverse legal benchmarks to further validate its scalability.

To mitigate potential overfitting caused by the relatively limited dataset size, several regularization strategies were employed throughout the training process, including dropout layers, early stopping, validation-based monitoring, and 5-fold cross-validation. The low standard deviation observed across cross-validation folds indicates that the proposed model maintains stable generalization performance despite the moderate dataset scale.

VIII. CONCLUSION AND FUTURE DIRECTIONS

This study presents a significant enhancement to the Bekenbey AI model by integrating BERT into its pipeline, aiming to improve legal text processing, document classification, and predictive analytics. Our results demonstrate that the proposed method consistently outperforms the Bekenbey model in both classification and text generation tasks. The improvements observed across all performance metrics suggest that the proposed enhancements contribute to better model generalization and robustness. The increased precision, recall, and F1-score indicate superior classification performance, while the higher ROUGE and BLEU scores confirm enhanced text generation capabilities. Given that both the proposed method and Bekenbey are designed for legal document analysis, the improvements observed in text generation, and classification hold significant implications for legal applications.

The enhanced performance can lead to more accurate and coherent legal text summarization, document classification, and contract analysis, ultimately improving efficiency in legal workflows. These advancements can support automated legal reasoning, case law analysis, and regulatory compliance monitoring. Furthermore, the findings indicate that Transformer-based models have the potential to revolutionize AI-driven legal analysis, offering more nuanced and context-aware solutions compared to previous methodologies. This study demonstrates that integrating domain-aware attention mechanisms with large-scale language models can significantly enhance the reliability and interpretability of AI-driven legal systems.

The proposed attention mechanism also improves interpretability by highlighting legally salient tokens, enabling more transparent AI-assisted legal decision support. The ablation study further confirms that both BERT integration

and the proposed domain-aware attention mechanism contribute synergistically to overall model performance. These results confirm that the proposed model not only improves performance metrics but also provides a scalable and interpretable framework for real-world legal AI applications. The proposed framework demonstrates strong potential for deployment in real-world legal decision support systems. Interpretability and transparency remain critical requirements for AI-assisted legal decision support systems. In this context, the proposed Bekenbey Attention mechanism contributes to explainability by emphasizing legally salient tokens and semantically influential regions within legal documents. This capability enables legal practitioners to better understand the reasoning patterns underlying model predictions, thereby supporting more transparent, accountable, and trustworthy AI-assisted legal workflows. Such transparency is particularly important in legal environments where fairness, accountability, and human oversight are essential for responsible AI deployment. Overall, the findings demonstrate that combining transformer-based contextual understanding with domain-aware legal attention mechanisms represents a promising direction for next-generation AI-assisted legal analytics.

Our future research will focus on further optimizing the integration of Transformer based architectures within the upgraded AI framework. Additionally, these directions include fine-tuning domain-specific BERT variants for enhanced legal reasoning, exploring hybrid models that combine symbolic AI with deep learning, and implementing reinforcement learning techniques to improve decision-making capabilities.

DECLARATIONS

Acknowledgements: The authors do not wish to acknowledge any individual or institution.

Author Contributions: Conceptualization, A.D., M.A.E., İ.T.M., and T.M.; Methodology, A.D. and M.A.E.; Software, A.D.; Validation, M.A.E. and İ.T.M.; Formal Analysis, A.D. and M.A.E.; Investigation, A.D.; Resources, A.D.; Data Curation, A.D.; Writing—Original Draft, A.D.; Writing—Review & Editing, M.A.E., İ.T.M., and T.M.; Visualization, A.D.; Supervision, İ.T.M. and T.M.; Project Administration, A.D. All authors have read and approved the final version of the manuscript.

Conflict of Interest: The authors declare no conflict of interest.

Supporting Institutions: This research received no external funding.

Ethical Approval: This study does not involve human or animal participants. All procedures followed scientific and ethical principles, and all referenced studies are appropriately cited.

Plagiarism Statement: This article has been evaluated for plagiarism and no instances of plagiarism were detected.

Use of AI Tools: Artificial Intelligence (AI) tools were used only for language improvement, grammar correction, and editorial assistance during manuscript preparation. All scientific content, analyses, interpretations, and conclusions were created, verified, and approved by the authors.

REFERENCES

Abimbola, B., de La Cal Marin, E., & Tan, Q. (2024). Enhancing legal sentiment analysis: A convolutional neural network–long short-term memory document-level model. *Machine Learning and Knowledge Extraction, 6*(2), 877–897. https://doi.org/10.3390/make6020041

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *Peerj Computer Science, 2*, Article e93.

Alhalalmeh, A. H., & Al-Tarawneh, A. (2025). Artificial intelligence and the law: The complexities of technology and legalities. In *Intelligence-Driven Circular Economy: Regeneration Towards Sustainability and Social Responsibility* (pp. 641–649). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-74220-0_50

Ashford, N. A., & Hall, R. P. (2011). *Technology, globalization, and sustainable development: Transforming the industrial state.* Yale University Press. https://doi.org/10.12987/9780300177602

Barbosa, T. S., de Castro, D., Singh, A. K., & Vitale, S. (2026). An experimental assessment of ai-based legal decision-making systems in contract analysis and risk detection. *Qubahan Techno Journal, 5*(1), 37–65. https://doi.org/10.48161/qtj.v5n1a81

Barnett, R. E. (2014). *The structure of liberty: Justice and the rule of law* (2nd ed.). Oxford University Press. https://doi.org/10.1093/0198297297.001.0001

Bhambhoria, R., Dahan, S., Li, J., & Zhu, X. (2026). Evaluating AI for law: Bridging the gap with open-source solutions. In *Compliance for Artificial Intelligence Systems: Strategies, Principles and Methods* (pp. 59–74). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-12795-2_5

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on Fairness, Accountability and Transparency* (Vol. 81, pp. 149–159). PMLR.

https://proceedings.mlr.press/v81/binns18a.html
Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics, 6*, Article 1421273. https://doi.org/10.3389/fhumd.2024.1421273
Chen, Q., Wu, Q., Chen, J., Wu, Q., van den Hengel, A., & Tan, M. (2020). Scripted video generation with a bottom-up generative adversarial network. *IEEE Transactions on Image Processing, 29*, 7454–7467. https://doi.org/10.1109/TIP.2020.3003227
Council of Europe. (1950, November 4). *European Convention on Human Rights*. https://www.echr.coe.int/documents/d/echr/convention_eng
Dalmia, S., Okhonko, D., Lewis, M., Edunov, S., Watanabe, S., Metze, F., Zettlemoyer, L., & Mohamed, A. (2023). Legonn: Building modular encoder-decoder models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31*, 3112–3126. https://doi.org/10.1109/TASLP.2023.3296019
Deveci, A., Erkan, M. A., Medeni, İ. T., & Medeni, T. D. (2026a). A novel framework for next word prediction using long-short term memory networks. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 28*(82), 113–120. https://doi.org/10.21205/deufmd.2026288215
Deveci, A., Erkan, M. A., Medeni, İ. T., & Medeni, T. D. (2026b). Innovative approaches to high-energy gamma particle classification: A coevolutionary artificial neural network for cherenkov telescopes. *Sivas Cumhuriyet Üniversitesi Mühendislik Fakültesi Dergisi, 4*(1), 8–15. https://doi.org/10.66248/cumfad.1869125
Gul, S., Asif, M., Fazal-e-Amin, Saleem, K., & Imran, M. (2025). Advancing aspect-based sentiment analysis in course evaluation: A multi-task learning framework with selective paraphrasing. *IEEE Access, 13*, 7764–7779. https://doi.org/10.1109/ACCESS.2025.3527367
Herasymenko, O., Artemov, V., Kravtsev, O., Yunin, O., & Fedorchuk, Y. (2025). AI use by ukrainian law enforcement in combating crime at critical infrastructure facilities. *Salud, Ciencia y Tecnología-Serie de Conferencias, 4*, Article 1318. https://doi.org/10.56294/sctconf20251318
Ingebrand, T., Thorpe, A. J., Goswami, S., Kumar, K., & Topcu, U. (2025). Basis-to-basis operator learning using function encoders. *Computer Methods in Applied Mechanics and Engineering, 435*, Article 117646. https://doi.org/10.1016/j.cma.2024.117646
Katz, D. M., Bommarito II, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PloS One, 12*(4), Article e0174698. https://doi.org/10.1371/journal.pone.0174698
Maurya, A. (2025). Scaling legal AI: Benchmarking Mamba and transformers for statutory classification and case law retrieval. *arXiv*. https://doi.org/10.48550/arXiv.2509.00141
Neelamegam, P., & Nirmala, S. J. (2025). Recent trends in legal AI: A comprehensive review. In *2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)* (pp. 152–159). IEEE. https://doi.org/10.1109/ICAISS61471.2025.11042154
Ni, J., Ren, S., Tang, G., Cao, W., & Shi, P. (2023). An improved shared encoder-based model for fast panoptic segmentation. *IEEE Sensors Journal, 24*(14), 22070–22083. https://doi.org/10.1109/JSEN.2023.3332354
Nirmal, B. C., & Singh, R. K. (Eds.). (2018). *Contemporary issues in international law: Environment, international trade, information technology and legal education.* Springer Nature Singapore. https://doi.org/10.1007/978-981-10-6277-3
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog, 1*(8), Article 9.
Schabas, W. A. (2015). The European convention on human rights: A commentary. Oxford University Press. https://doi.org/10.1093/law/9780199594061.001.0001
Siino, M., Falco, M., Croce, D., & Rosso, P. (2025). Exploring LLMs applications in law: A literature review on current legal NLP approaches. *IEEE Access, 13*, 18253–18276. https://doi.org/10.1109/ACCESS.2025.3533217
Sun, S., Che, L., Zhao, R., Chen, Y., & Li, M. (2025). Multi-task learning and voltage reconstruction-based battery degradation prediction under variable operating conditions of energy storage applications. *Energy, 317*(5), Article 134645. https://doi.org/10.1016/j.energy.2025.134645
United Nations General Assembly. (1948, December 10). *Universal declaration of human rights.* https://www.un.org/en/about-us/universal-declaration-of-human-rights
Ye, H., Sun, Y., Gao, Y., Xu, F., & Qi, J. (2025). Heterogeneous Graph Transformer Auto-Encoder for multivariate time series forecasting. *Computers and Electrical Engineering, 122*, Article 109927. https://doi.org/10.1016/j.compeleceng.2024.109927
Yücesan, E., Erkan, M. A., Deveci, A., & Medeni, İ. T. (2024). Bekenbey AI: Innovative solutions at the intersection of deep learning and law. *Sivas Cumhuriyet Üniversitesi Mühendislik Fakültesi Dergisi, 2*(2), 185–192. https://doi.org/10.66248/cumfad.1590764
Zoetmulder, R., Gavves, E., Caan, M., & Marquering, H. (2022). Domain-and task-specific transfer learning for medical segmentation tasks. *Computer Methods and Programs in Biomedicine, 214*, Article 106539. https://doi.org/10.1016/j.cmpb.2021.106539