IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 May 2025, accepted 29 June 2025, date of publication 7 July 2025, date of current version 24 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3586781

RESEARCH ARTICLE

LegalMind: Agentic AI-Driven Process Optimization and Cost Reduction in Legal Services Using DeepSeek

NIDADAVOLU VENKAT DURGA SAI SIVA VARA PRASAD RAJU[1], NURUZZAMAN FARUQUI[2], NIKHIL PATEL[3], (Senior Member, IEEE), OLIVIA-ROXANA ALECSOIU[4], PRIYABRATA THATOI[5], SALEM A. ALYAMI[6], (Member, IEEE), AND AKM Azad[6]

[1]Department of Software Engineering, Birla Institute of Technology and Science (BITS) at Pilani, Pilani, Rajasthan 333031, India
[2]Department of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[3]Department of Business Administration, University of Dubuque, Dubuque, IA 52001, USA
[4]Faculty of Educational Sciences, Law and Public Administration, Constantin Brâncuşi University of Târgu-Jiu, 210185 Târgu Jiu, Romania
[5]Amazon, Chicago, IL 60606, USA
[6]Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 13318, Saudi Arabia

Corresponding author: AKM Azad (kazad@imamu.edu.sa)

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) under Grant IMSIU-DDRSP2504.

ABSTRACT The legal industry struggles with inefficiencies, high costs, and manual-intensive workflows. Traditional AI lacks adaptability in optimizing legal operations. To address this, we propose *LegalMind*, an agentic AI-driven framework leveraging DeepSeek R1 for intelligent legal process automation and cost reduction. LegalMind integrates a structured legal dataset and fine-tunes DeepSeek R1 to enhance decision-making and workflow efficiency. Experimental results show a 42.6% cost reduction and a 60.8% improvement in document processing speed over baseline AI models. Scalability tests confirm the system's ability to handle 100,000 queries efficiently. Real-world case studies validate LegalMind's effectiveness in law firms, corporate legal departments, and government agencies, demonstrating significant reductions in case preparation time and operational costs. These findings highlight the transformative potential of agentic AI in legal automation, optimizing workflows and improving decision support.

INDEX TERMS DeepSeek, agentic AI, legal process automation, cost optimization in legal services, deep learning in law, scalable AI for LegalTech, intelligent legal decision support, LLM, transformer model.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Extensive documentation, complex workflows, and significant operational costs due to the involvement of specialized professionals characterize the legal industry [1]. Law firms, corporate legal departments, government agencies focused on legal aspects, and any organization that works in legal activities experience labor-intensive tasks such as contract analysis, case law research, and litigation preparation. Each of these tasks is time-consuming, which makes them expensive as legal professionals' hourly rates are very high [2]. This is where intelligent process automation can play an effective role in increasing efficiency and reducing the costs associated with it. However, traditional legal automation systems primarily rely on rule-based approaches, which are not suitable for the diverse types of documents of this industry, which come in multiple different forms [3]. Besides, the complexity of legal cases necessitates a more advanced approach to process optimization and cost management systems, which has been proposed in this paper. This paper utilizes the recent advancement in transformer-centric Large Language Model (LLM), DeepSeek R1 [4], as an agent and, through agentic AI architecture, it proposes a context-aware

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

intelligent process optimization system for a legal domain, which significantly reduces the operational cost.

B. CHALLENGES IN LEGAL PROCESS OPTIMIZATION

The traditional legal procedure is inefficient in terms of time and financial cost because of the dependency on manual document handling and human expert-driven research. Extensive research, document preparation, and administrative workload are some of the challenges in this sector. Besides, legal workflows are, by nature, time-consuming as they involve case analysis, litigation preparation, and other repetitive resource-intensive activities [5]. Besides, there is no standardized cost model for this sector, which makes it difficult to predict and minimize legal expenditures. Another challenge is scalability, as expanding legal operations while maintaining efficiency becomes increasingly complex without intelligent process management solutions [6]. Overcoming these challenges requires context-aware AI-driven solutions that not only automate legal tasks and retrieve information but are also capable of relating complex legal reasoning and assisting in decision-making.

C. ROLE OF AGENTIC AI IN LEGAL COST REDUCTION

Traditional AI is limited by certain rules or specific operations such as prediction, classification, or document retrieval [7]. The concept of agentic AI takes the capability of the system beyond the limitations of traditional AI by introducing human-like intelligence where the agent itself is capable of performing multiple different complex tasks, including document analysis, task automation, legal aspects review, legal decision-making, and offering legal strategies [8]. That is why by integrating agentic AI into legal operations, firms can reduce manual intervention, enhance resource allocation, and improve decision-making efficiency. The ability of agentic AI to dynamically adjust to legal complexities makes it a valuable tool for modernizing legal workflows, ensuring faster results, and achieving financial efficiency.

D. CONTRIBUTIONS OF THIS WORK

The primary contribution of this paper, LegalMind, is the development and application of an agentic AI framework for legal workflow optimization and cost reduction. This is where the novelty of this approach lies. The following list summarizes the core contribution of this study:

- Design, development, and implementation of agentic AI framework for intelligent process optimization and cost reduction in the legal sector.
- Achieving adaptability by learning from the user interaction and training itself to adapt to the dynamic landscape of the legal domain.
- Application of DeepSeek R1 in designing the agent of the agentic AI framework.
- Utilizing the agent as a decision support system to reduce the workload of legal professionals.
- Empirical evaluation of LegalMind's performance using real-world legal datasets, demonstrating improvements in process efficiency and cost reduction.

E. RESEARCH QUESTIONS

To overcome the challenges of applying an agentic AI-based solution in the legal domain, this paper is guided by three research questions (RQs) that have been devised to specify the scope of this study. These questions also serve as a guide for methodological development. The following list presents these questions:

- **RQ1:** How can agentic AI models be designed to perform reliable legal reasoning under ambiguity and conflicting precedents?
- **RQ2:** What AI-driven process optimization techniques can effectively reduce legal service costs while maintaining regulatory and ethical compliance?
- **RQ3:** How can AI-based legal assistants be enhanced to support complex contextual understanding and decision-making in ethically sensitive or borderline legal scenarios?

The conceptual expansion presented in this paper is guided by these research questions.

F. PAPER ORGANIZATION

The remainder of this paper is structured as follows: Section II presents a literature review discussing existing AI applications in the legal domain and their limitations. Section III details the methodology, including the proposed agentic AI framework, dataset preprocessing, and cost-optimization strategies. Section IV describes the implementation of LegalMind and its integration into legal workflows. Section V provides an experimental analysis, evaluating the system's impact on legal process optimization and cost reduction. Section VI discusses key findings, limitations, and future research directions. Finally, Section VII concludes the paper by summarizing contributions and potential advancements in AI-driven legal optimization.

II. LITERATURE REVIEW

A. TRADITIONAL LEGAL AUTOMATION AND ITS LIMITATIONS

The application of traditional AI for legal automation is not a new field of research. It has been around for a while and is limited by its rule-based approach [9]. Most of these systems have been developed for document retrieval, contract analysis, and legal compliance verification. However, despite their utility, they exhibit several fundamental limitations [10]. When the task or requirements go beyond the predefined rules, traditional AI systems fail to comply, resulting in inaccurate or incomplete legal document analysis [11]. Another significant limitation of traditional AI is the lack of multi-modal capability. Legal documents come in many different models, which require multi-modal capabilities [12]. The

agentic AI has the capability to overcome these limitations, which have been utilized and presented in this paper.

B. INTELLIGENT PROCESS OPTIMIZATION IN LEGAL WORKFLOWS

Process optimization is the key to reducing inefficiencies and operational costs in the legal domain [13]. Existing AI-driven solutions help improve legal operations by making document search and retrieval easier. These solutions also schedule tasks automatically, which also accelerates the workflow [14]. However, these tools don't have the capability to analyze the contextual meaning of the document, perform human-like reasoning, and assist in decision-making. The introduction of intelligent process automation (IPA) has enabled some improvements by leveraging machine learning for document classification, entity recognition, and workflow scheduling [15]. Agentic AI demonstrates the potential of redefining intelligent process optimization in legal workflow by introducing human-like capabilities, which fills the gap identified with traditional AI.

C. ADVANCEMENTS IN LARGE LANGUAGE MODELS AND AGENTIC AI

With the emergence of transformer-based Large Language Models (LLMs), the capability of AI has extended to a new level. It is now capable of processing complex legal text, understanding the complicated relation among different factors, combining them with multiple logics within the legal boundaries, and making intelligent decisions like legal professionals [16]. Models such as GPT-4 [17], BERT [18], and LegalBERT [19] have demonstrated remarkable improvements in legal document comprehension, text summarization, and precedent-based reasoning. Despite remarkable capabilities, these solutions have one limitation, and that is the application as an assistive tool rather than an intelligent agent. This limitation has been overcome through the innovation of this paper.

D. COST REDUCTION STRATEGIES IN LEGAL TECHNOLOGY

The intensive workload associated with case management, document review, and litigation preparation is the primary reason for the legal sector to be expensive [20]. The application of traditional AI, such as e-discovery tools, automated contract review systems, and predictive case analytics, helps accelerate the process and reduce operational costs. However, their limited capabilities fail to achieve significant cost savings [21]. One of these major limitations is the dependence on static cost modeling, where legal expenses are estimated using predefined metrics that do not account for case complexity, jurisdictional differences, or dynamic legal strategies [22]. Moreover, most of the cost-reduction-focused approaches rely on historical cost data instead of making decisions in real-time financial adjustments. As a result, they remain ineffective in new legal conditions [23]. However, the agentic AI works independently by drawing logical reasoning for individual cases regardless of historical rules. That is why it is a more effective cost reduction.

E. RESEARCH GAPS AND OBJECTIVES

The literature review suggests that significant advancements have been made in case-based reasoning, knowledge engineering, and text mining. Many of these technologies have been applied in the legal domain as well. However, application of agentic AI in legal domain especially where DeepSeek R1 is the agent is still under-explored. The research gaps and corresponding research objectives discovered through the literature review are summarized in Table 1.

III. METHODOLOGY

The proposed LegalMind has been developed using the DeepSeek R1 model, utilizing the concept of transfer learning for fine-tuning with a unique customized dataset prepared with a focus on the legal domain. The agentic AI architecture has been used with the fine-tuned DeepSeek R1 as the agent. The methodological overview has been illustrated in Figure 1, which doesn't include the Agentic AI framework. The framework has been presented separately in Figure 3.

A. DATASET FORMATION AND PREPROCESSING

The dataset for training and evaluating the *LegalMind* framework has been compiled from multiple sources, including legal documents from an experimenting law firm, publicly available case studies, books on legal procedures, and research papers in the domain of law. The total dataset comprises 52,780 legal documents, ensuring diversity in legal language, jurisdictional interpretations, and case precedents. Table 2 summarizes the dataset properties developed in this study.

To pre-process the dataset, it has been mathematically structured, which is defined in Equation 1 where x_i represents the legal text, and y_i corresponds to the labeled case outcome. The raw dataset is not suitable for training the DeepSeek R1 model used in this paper as the agent.

$$\mathcal{D} = \{(x_i, y_i) \mid x_i \in \mathbb{T}, y_i \in \mathbb{L}\} \tag{1}$$

1) TEXT CLEANING AND NORMALIZATION

The raw dataset contains a combination of upper and lowercase letters. There are special characters, hypertext, and many irrelevant symbols. These do not play any role in feature learning. Instead, they complicate the feature vector. That is why the dataset has been cleaned and normalized first. It has been done by following the mathematical structure defined in Equation 2 where $L(x)$ converts text to lowercase, $S(x)$ removes stopwords, and $A(x)$ eliminates special characters [24].

$$T(x) = \{L(x_i) \mid x_i \in x, \neg S(x_i) \wedge A(x_i)\} \tag{2}$$

2) TOKENIZATION AND EMBEDDING

The DeepSeek R1 model used in *LegalMind* cannot learn from the cleaned and normalized dataset directly. It requires

TABLE 1. Identified research gaps and corresponding research objectives.

Research Gap	Research Objective
Lack of autonomous reasoning and adaptability in rule-based legal automation systems	Design a transformer-based agentic AI framework capable of human-like reasoning and self-adjustment to case complexities
Inability of existing tools to adapt dynamically to ongoing user interactions and legal context shifts	Enable continuous learning from user interactions and legal document streams for personalized task adaptation
Absence of integrated cost optimization in AI-based legal assistants	Integrate dynamic cost-awareness into legal workflow optimization to minimize resource allocation overhead
Traditional LLM-based tools are modular and not embedded into agentic workflows	Embed LLM (DeepSeek R1) within a multi-functional agent capable of full-cycle legal support, from analysis to strategy
Static financial modeling used in most cost reduction systems lacks context sensitivity	Implement real-time, context-aware financial decision support within legal automation

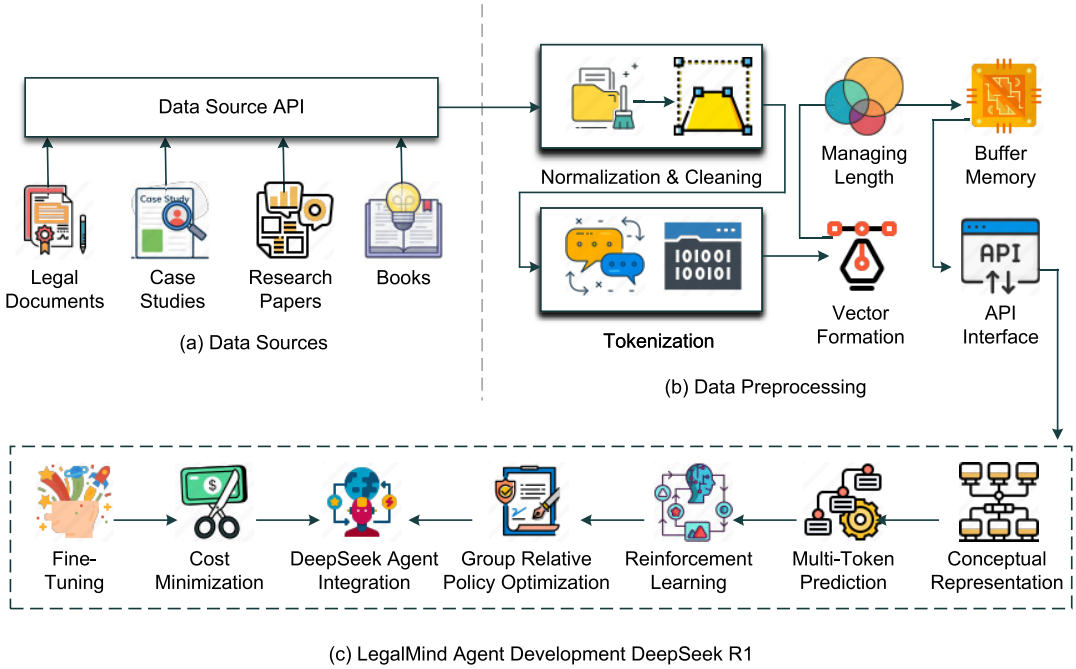


FIGURE 1. The methodological overview with a simplified representation of process flow.

TABLE 2. Dataset composition for training and evaluation of LegalMind.

Document Category	Documents	Total Pages	Total Words
Legal Documents	18,500	240,800	98,560,000
Case Studies	12,300	154,600	67,520,000
Books	8,200	246,000	84,320,000
Research Papers	13,780	165,360	74,480,000
Total	**52,780**	**806,760**	**324,880,000**

numerical representations. That is why the raw dataset has been tokenized, and after that, word embedding has been performed. In the tokenization stage, legal text is broken into fundamental units called tokens, ensuring better context retention and computational efficiency. In this process, the mathematical principle expressed in Equation 3 has been used [25], [26].

$$\mathcal{T}(D) = \bigcup_{i=1}^{n} \{t_{i1}, t_{i2}, \ldots, t_{ik} \mid w_i \in D, k = \text{len}(w_i)\} \quad (3)$$

In Equation 3, $\mathcal{T}(D)$ is the set of tokens extracted from the document set D, w_i denotes the i-th word, and t_{ij} represents the j-th token derived from w_i, with k defining the token length per word.

Once tokenized, these tokens are mapped to high-dimensional vectors in a continuous space using an embedding matrix $\mathcal{E}$. The transformation into numerical representation is formalized in Equation 4 where v_i is the vectorized form of token t_i, and $\mathcal{E}$ represents the embedding matrix with dimensions $|\mathbb{V}| \times d$, where $|\mathbb{V}|$ is the vocabulary size and d is the embedding dimension.

$$v_i = \mathcal{E}[t_i], \quad \forall t_i \in \mathcal{T}(D) \quad (4)$$

The full input sequence is structured as a matrix $\mathbf{X}$, built by stacking token embeddings into a matrix representation, as shown in Equation 5 where $\mathbf{X}$ serves as the input matrix for the transformer model, with L representing the number of tokens in the input sequence, adjusted to a fixed length via

padding or truncation where necessary [27].

$$\mathbf{X} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_L \end{bmatrix}, \quad \text{where } L = \text{len}(\mathcal{T}(D)) \tag{5}$$

This is the structure the DeepSeek R1 supports to perform transfer learning operations on it. This structured approach ensures that LegalMind's AI system efficiently processes complex legal texts, capturing essential linguistic patterns and optimizing legal document analysis. The data pre-processing steps explained in this section are used to prepare the dataset for training the model. Later, it is used as a data pre-processing pipeline for processing queries from the users which is presented as Algorithm 1.

Algorithm 1 LegalMind Data Processing Pipeline

Require: Raw dataset $\mathcal{D}_{raw} = \{d_1, d_2, \ldots, d_n\}$
Ensure: Processed dataset $\mathcal{D}$ for model training
Step 1: Cleaning and Normalization
for each document $d_i \in \mathcal{D}_{raw}$ **do**
    $d_i \leftarrow \text{LowerCase}(d_i)$
    $d_i \leftarrow \text{RemoveStop}(d_i)$
    $d_i \leftarrow \text{RemoveSpecial}(d_i)$
end for
Step 2: Tokenization
for each cleaned document d_i **do**
    $T_i \leftarrow \text{Tokenize}(d_i)$
end for
Step 3: Embedding
for each token $t_j \in T_i$ **do**
    $v_j \leftarrow \text{Embed}(t_j)$
end for
$\mathbf{X} \leftarrow \text{FormMatrix}(v)$
Step 4: Dataset Splitting
$\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \mathcal{D}_{\text{test}} \leftarrow \text{Split}(\mathcal{D})$
Output: Processed dataset $\mathcal{D}$ ready for DeepSeek R1 training.

3) DATASET SPLITTING

To ensure robust model training and evaluation, the dataset has been split into three subsets: training, validation, and testing. The splitting follows the widely adopted 70:15:15 ratio [28], ensuring that the model generalizes well across unseen data while preventing overfitting. The dataset $\mathcal{D}$ is defined in Equation 6 [28].

$$\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}} \cup \mathcal{D}_{\text{test}} \tag{6}$$

In Equation 6, $\mathcal{D}_{\text{train}}$ represents the training dataset (70% of total data), $\mathcal{D}_{\text{val}}$ is the validation dataset (15% of total data), used to fine-tune hyperparameters, and $\mathcal{D}_{\text{test}}$ is the testing dataset (15% of total data), reserved for final performance evaluation. The detailed distribution of the dataset is presented in Table 3.

TABLE 3. Dataset splitting for LegalMind model training and evaluation.

Dataset Category	Documents	Total Pages	Total Words
Training Set ($\mathcal{D}_{\text{train}}$)	36,946	564,732	227,416,000
Validation Set ($\mathcal{D}_{\text{val}}$)	7,917	121,014	48,732,000
Testing Set ($\mathcal{D}_{\text{test}}$)	7,917	121,014	48,732,000
Total Dataset	**52,780**	**806,760**	**324,880,000**

Each subset maintains the same document type distribution as the original dataset, ensuring balanced representation across legal documents, case studies, books, and research papers. This structured splitting approach guarantees reliable model evaluation while preserving the integrity of legal text diversity.

B. LEGALMIND AGENT USING DEEPSEEK R1

The LegalMind Agent is designed to enhance legal decision-making, optimize workflows, and minimize legal expenditures. It has been designed using the DeepSeek R1 model. It is an advanced transformer-based architecture with a Mixture of Experts (MoE) structure [29]. The MoE is a reinforcement learning optimization layer. This is the adaptive learning engine of the proposed LegalMind framework [30]. The agent integrates these features to automate legal workflows, predict costs, and provide decision support.

1) CONCEPTUAL REPRESENTATION OF DEEPSEEK R1

The innovation of the DeepSeek R1 is in the MoE architecture. This architecture doesn't activate all of the parameters. It only acts as a fraction of the total parameters per-forward pass. This process follows the mathematical principle presented in Equation 7.

$$y = \sum_{i=1}^{M} G_i(x) \cdot f_i(x) \tag{7}$$

In Equation 7, x represents the input sequence (tokenized legal text), $f_i(x)$ is the output from the i-th expert model, $G_i(x)$ is a learned gating function that determines the weight assigned to each expert, and M is the number of available experts. Only a subset of k experts is activated per forward pass, ensuring efficiency. It is expressed in Equation 8.

$$\sum_{i=1}^{M} G_i(x) = k, \quad k \ll M \tag{8}$$

This selective activation significantly reduces computational cost while maintaining expert-driven reasoning for legal text processing. This is how it reduces the computational cost and prevents the system from adding additional cost.

2) MULTI-TOKEN PREDICTION (MTP) IN LEGAL ANALYSIS

DeepSeek R1 incorporates Multi-Token Prediction (MTP) to accelerate text generation, allowing multiple legal tokens to be predicted simultaneously. The process is defined by Equation 9 where y_i represents predicted tokens and k is the number of tokens generated per step. This technique

significantly enhances response time when analyzing legal documents [31].

$$P(y_1, y_2, \ldots, y_k|x) = \prod_{i=1}^{k} P(y_i|x, y_1, \ldots, y_{i-1}) \quad (9)$$

3) REINFORCEMENT LEARNING WITH GROUP RELATIVE POLICY OPTIMIZATION (GRPO)

LegalMind leverages DeepSeek R1's Reinforcement Learning (RL) capabilities using a Group Relative Policy Optimization (GRPO) approach, which improves decision-making efficiency. It is expressed in Equation 10 [32].

$$J(\theta) = \mathbb{E}_{x\sim D}\left[\sum_{t=1}^{T} \log \pi_\theta(a_t|s_t)A_t\right] \quad (10)$$

In Equation 10, $\pi_\theta(a_t|s_t)$ is the policy function governing decision-making at timestep t, A_t is the advantage function, measuring the benefit of selecting action a_t, and D represents the dataset of legal cases used for training. This reinforcement learning process allows LegalMind to adapt dynamically to legal complexities and refine predictions over time.

4) LEGALMIND AGENT INTEGRATION WITH DEEPSEEK R1

The LegalMind Agent utilizes DeepSeek R1's MoE-based architecture and RL-driven fine-tuning to optimize legal workflows. The interaction process is formalized as the mathematical concept presented in Equation 11.

$$\mathbf{L}_{\text{optimized}} = \arg\min_\theta \sum_{i=1}^{N} \mathcal{L}(\mathcal{O}_i, \mathcal{P}_i) \quad (11)$$

In Equation 11, $\mathcal{L}$ is the loss function measuring the difference between actual outcomes $\mathcal{O}_i$ and predicted outcomes $\mathcal{P}_i$, θ are the model parameters optimized using reinforcement learning, and N is the number of training instances (legal cases processed). LegalMind continuously updates its internal decision-making strategies by optimizing Equation 11, ensuring cost efficiency and enhanced legal process automation.

5) ADAPTIVE COST MINIMIZATION IN LEGAL DECISION-MAKING

By modeling legal expenditure as a dynamic optimization problem, LegalMind reduces unnecessary costs in legal workflows. This concept has been developed by following the mathematical principle expressed in Equation 12.

$$\min_\theta \mathbb{E}\left[C(\theta)\right], \quad C(\theta) = \sum_{j=1}^{M} c_j G_j(\theta) \quad (12)$$

In Equation 12, $C(\theta)$ is the total operational cost for a given legal process, c_j represents the cost associated with each legal task, and $G_j(\theta)$ determines whether the task is automated (low cost) or manual (high cost). This equation ensures that LegalMind optimally distributes resources, balancing automation vs. human decision-making for maximum cost efficiency. The LegalMind Agent powered by DeepSeek R1 combines a MoE architecture, reinforcement learning-driven decision support, and multi-token parallelism to enhance legal workflow automation. By minimizing costs and improving prediction accuracy, LegalMind represents a state-of-the-art intelligent legal assistant capable of handling complex legal document processing, case prediction, and operational efficiency optimization.

C. FINE-TUNING LEGALMIND AGENT WITH TRANSFER LEARNING

The DeepSeek R1 model is a trained transformer. However, it is essential to make it specialized in legal process optimization and cost reduction. That is why it has been fine-tuned with the transfer learning method to train it on the legal documents processed in this paper. The transfer learning process follows the mathematical principle of Equation 13 where D_{legal} is the legal domain-specific dataset, and $P(y_i|x_i)$ represents the probability of correct predictions. The learning curve is illustrated in Figure 2 which demonstrates that the transfer learning worked properly [33].

$$\mathcal{L}_{\text{finetune}} = -\sum_{i=1}^{N} \log P(y_i \mid x_i, D_{\text{legal}}) \quad (13)$$

To update the weights of the hidden neurons optimally, the Adaptive Moment Estimation (ADAM) has been used, which is defined in Equation 14 where $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected moment estimates [34].

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (14)$$

D. AGENTIC AI FRAMEWORK

The proposed Agentic AI framework for *LegalMind* is designed to automate legal workflows, optimize decision-making, and minimize legal expenditures using DeepSeek R1 as its primary agent. A modular architecture has been developed so that it can be easily upgraded. There are three core modules in this architecture, and they are Legal Document Processing, Multi-Agent Decision Support, and Workflow Automation. The core elements and their position in the framework are illustrated in Figure 3.

1) LEGAL DOCUMENT PROCESSING MODULE

It is the first module of the framework. This module ingests, classifies, and interprets legal texts using DeepSeek R1's MoE architecture. The legal text undergoes semantic embedding transformation using a transformer-based function, which is defined in Equation 15.

$$\mathbf{h}_i = \text{DeepSeekR1}(\tau(T), \mathcal{C}) \quad (15)$$

In Equation 15, $\mathbf{h}_i$ represents the context-aware embeddings extracted from tokenized legal text, $\tau(T)$ is the tokenized representation of document T, and $\mathcal{C}$ denotes the legal case context, ensuring precise comprehension of legal

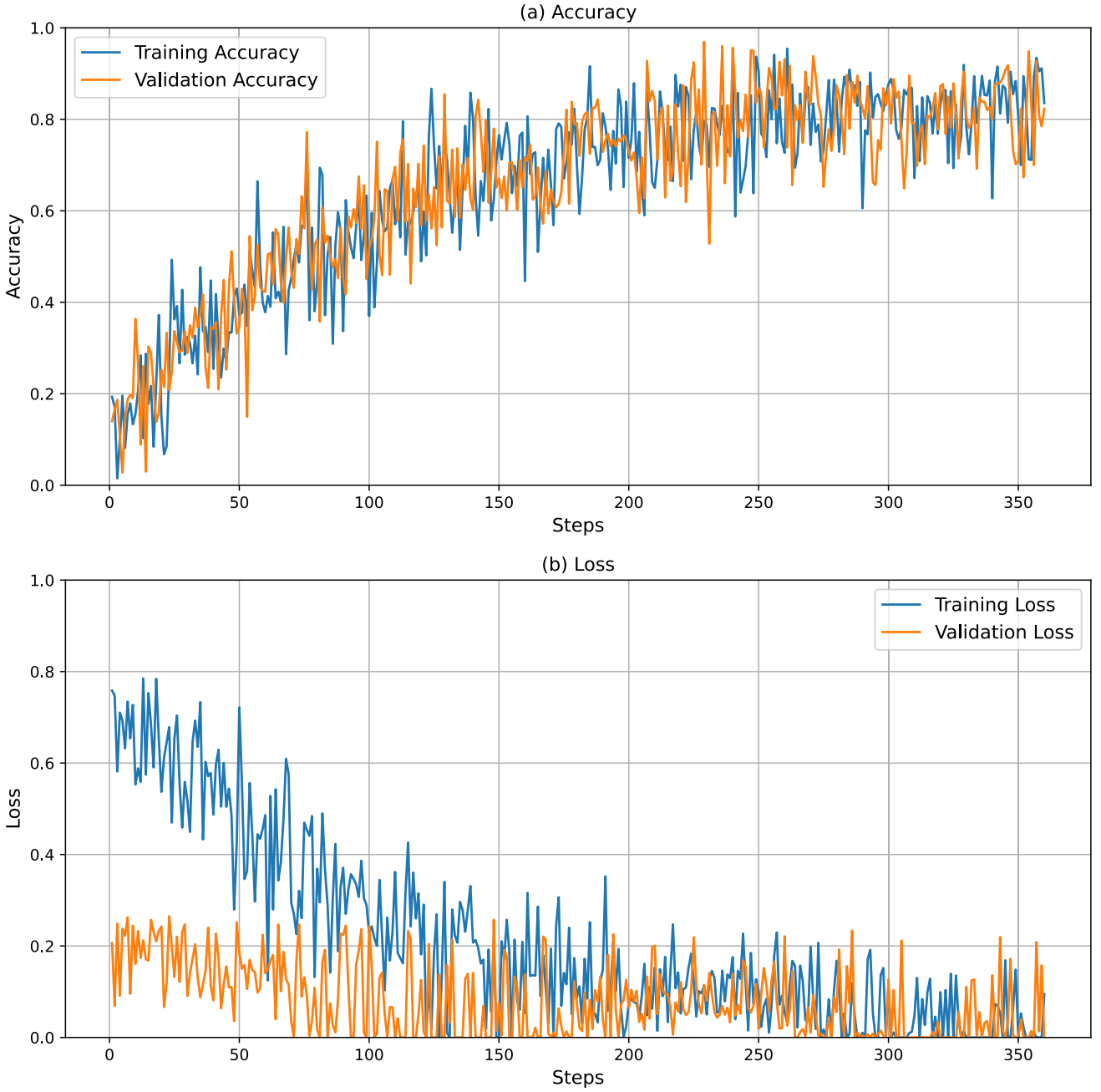


FIGURE 2. The learning curve analysis of the fine-tuning process.

terminology. DeepSeek R1 applies reinforcement learning to refine legal document embeddings, dynamically adapting to new case laws and legal interpretations.

2) MULTI-AGENT DECISION SUPPORT SYSTEM

The multi-agent decision support system layer significantly improves the capability of the proposed LegalMind framework. Because of this layer, it can process multiple different tasks simultaneously. This layer specializes in legal case research, compliance verification, and legal strategy optimization. Each agent evaluates and ranks legal precedents using a reinforcement learning policy, which is expressed in Equation 16.

$$\mathcal{A}_i = \arg\max_{a \in A} \sum_{s'} P(s'|s,a)\big(R(s,a) + \gamma V(s')\big) \tag{16}$$

In Equation 16, $\mathcal{A}_I$ represents the optimal action taken by agent i, $P(s'|s,a)$ is the transition probability to future state s', $R(s,a)$ is the reward function for selecting legal actions, and γ is the discount factor ensuring long-term optimization. To prioritize the most relevant legal precedents, DeepSeek R1 applies attention-based ranking by following

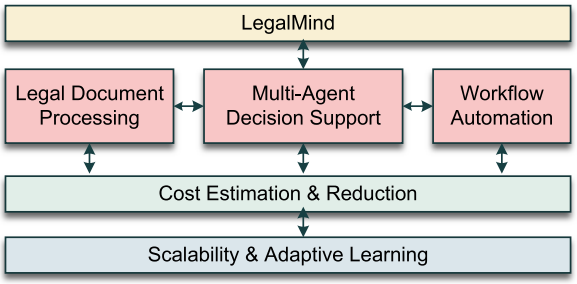


FIGURE 3. The overview of the module and the additional layer of the proposed agentic AI framework.

the mathematical principle of Equation 17.

$$\alpha_{ij} = \frac{\exp(f(\mathbf{s}_i, \mathbf{s}_j))}{\sum_{k=1}^{n} \exp(f(\mathbf{s}_i, \mathbf{s}_k))} \tag{17}$$

In Equation 17, $\mathbf{s}_i$ and $\mathbf{s}_j$ represent semantic vectors of two legal cases and $f(\mathbf{s}_i, \mathbf{s}_j)$ is a scoring function that quantifies the legal similarity. This ensures that LegalMind prioritizes the most relevant case laws in legal decision-making.

3) WORKFLOW AUTOMATION MECHANISM

The third layer in the modular architecture is the Workflow Automation Mechanism layer. This layer is responsible to automating the workflow. To enhance efficiency, legal workflow scheduling is modeled as an optimization problem. In this problem, the task completion time T and task priority Q are optimized using Equation 18.

$$\min T = \sum_i \frac{w_i}{v_i}, \quad Q = \sum_i \frac{p_i}{w_i} \tag{18}$$

In Equation 18, w_i is the workload of the i-th legal task, v_i is the processing speed of the automation system, and p_i is the priority assigned to each task based on urgency. To allocate resources efficiently, multi-agent task allocation is performed using a bidding mechanism, which is expressed in Equation 19.

$$\mathcal{A}^* = \arg\max_{\mathcal{A}_i} \left(\beta w_i - \lambda T_i\right) \tag{19}$$

In Equation 19, $\mathcal{A}^*$ is the optimal agent-task allocation, and β and λ are tunable coefficients balancing workload efficiency and processing cost.

4) COST ESTIMATION AND REDUCTION

One of the objectives of the proposed LegalMind framework is to optimize legal expenditures. That is why it is essential to estimate the cost. When there is an estimated cost, it works as a reference point to reduce it. In this paper, Bayesian cost estimation has been used to predict and optimize legal expenditures, which is expressed in Equation 20.

$$P(C_i|D) = \frac{P(D|C_i)P(C_i)}{P(D)} \tag{20}$$

In Equation 20, C_i represents cost categories (e.g., legal consultation, case research) and D denotes legal case data affecting cost variations. By modeling cost estimation dynamically, LegalMind identifies areas for cost minimization while preserving legal accuracy.

5) SCALABILITY AND ADAPTIVE LEARNING

The LegalMind framework has been developed to ensure everyone in an organization can use it simultaneously. That is why multi-agent facilities have been introduced in the architecture. It makes the scalability an essential feature of this framework. Besides, the legal domain evolves rapidly. As a result, adaptive learning is another essential feature of it. LegalMind adapts to evolving legal frameworks using reinforcement learning with DeepSeek R1. The system continuously refines legal strategies using a Q-learning function, which is defined in Equation 21.

$$Q(\mathcal{C}_t, a) = \max \sum_{t'} P(\mathcal{C}_{t'}|\mathcal{C}_t, a) R(\mathcal{C}_t, a) \tag{21}$$

In Equation 21, $Q(\mathcal{C}_t, a)$ represents the expected value of legal action a in case $\mathcal{C}_t$, $P(\mathcal{C}_{t'}|\mathcal{C}_t, a)$ models the probability of transitioning to case $\mathcal{C}_{t'}$, and $R(\mathcal{C}_t, a)$ is the reward function for optimizing legal strategies. This ensures LegalMind remains adaptable, improving decision efficiency, workflow automation, and cost minimization over time.

6) PAY-AS-YOU-GO LEGALMIND SERVICE PROVISIONING

To improve accessibility for firms with limited infrastructure, LegalMind adopts a cloud-based pay-as-you-go (PAYG) model, allowing on-demand usage with dynamic billing based on consumed compute that follows the structure expressed in Equation 22.

$$\mathcal{B}(f) := \pi \cdot u_f \quad \text{iff } u_f \leq \rho_f, \ \rho_f \subseteq \mathbb{R} \tag{22}$$

In Equation 22, $\mathbb{F}$ is be the set of legal firms, $\mathbb{R}$ the resource pool, $\mathbb{P}$ the pricing function, and ρ_f denotes the provisioned quota. If $u_f > \rho_f$, throttling τ_f is applied which follows the logic of Equation 23.

$$\tau_f := \neg\exists r \in \mathbb{R} \text{ s.t. } u_f + r \leq \rho_f^+ \tag{23}$$

In Equation 23, ρ_f^+ denotes the elastic upper bound under surge pricing or burst handling. This ensures fairness, stability, and cost containment without requiring local infrastructure.

E. PROMPT ENGINEERING FOR LEGAL TASK OPTIMIZATION

The operational principle of the prompt engineering for the LegalMind framework has been carefully designed for legal task optimization. It has been engineered to align the natural language inputs with agentic AI interpretability. Every prompt is transformed into vectors which are later used to reduce the semantic gap between user intent and model responses. In this process, a prompt is $p_i \in \mathbb{P}$ that elicits a legal task response from the agent $\mathcal{A}$. An encoding function $\psi : \mathbb{P} \to \mathbb{R}^d$ transform prompts into embeddings compatible

with the input space of DeepSeek R1. The refined embedding of a prompt is expressed using Equation 24.

$$\mathbf{z}_i = \psi(p_i) = \mathcal{E}_{\text{prompt}}[t_{p_i}], \quad t_{p_i} \in \mathcal{T}(p_i) \tag{24}$$

In Equation 24, $\mathcal{T}(p_i)$ represents the tokenized form of the prompt and $\mathcal{E}_{\text{prompt}}$ is a domain-specific embedding matrix adapted during fine-tuning. Prompt templates are optimized via latent feedback loops that adjust linguistic structure based on output quality, following a reinforcement signal defined by Equation 25.

$$R(p_i) = \sum_{j=1}^{K} \lambda_j \cdot \delta_j(\mathcal{A}(p_i), \mathcal{G}_j) \tag{25}$$

In Equation 25, δ_j measures alignment between the agent response $\mathcal{A}(p_i)$ and the expected ground truth outcome $\mathcal{G}_j$ across K evaluation dimensions such as legal accuracy, clause detection precision, and semantic relevance. The weights $\lambda_j \in [0, 1]$ control the relative importance of each evaluation criterion. A multi-phase prompt optimization cycle presented in Algorithm 2 has been developed for the LegalMind to offer feedback-driven prompt reformulation. It is continuously applied during the model interaction.

Algorithm 2 Feedback-Guided Prompt Engineering Loop

Require: Initial prompt set $\mathbb{P} = \{p_1, \ldots, p_n\}$, agent $\mathcal{A}$, reward threshold η
Ensure: Optimized prompt set $\mathbb{P}^*$

```
for each prompt p_i ∈ P do
    Generate output r_i ← A(p_i)
    Compute reward R_i ← R(p_i)
    if R_i < η then
        Reformulate p_i ← Refine(p_i, r_i, G)
        Repeat steps 2-4 until R_i ≥ η
    end if
end for
return Optimized set P*
```

The iterative refinement process of the Algorithm 2 utilizes the semantic prompt tuning strategies for improving the quality of the prompt and eliciting better responses from the LegalMind framework. In this process, the suffixes and context-aware constraints are appended to the base prompt. In this process, a query such as `"Extract liability clause"` is transformed into `"Identify and summarize all liability-related clauses within the following contract, including scope and involved parties"`. As a result, the difference between users' expectation and responses from the LegalMind significantly reduces. Moreover, to balance expressiveness and generalization, each prompt is probabilistically sampled from a prompt pool $\mathbb{P}_\theta$ as expressed in Equation 26

$$P(p_i) = \frac{\exp(\mu \cdot R(p_i))}{\sum_{j=1}^{n} \exp(\mu \cdot R(p_j))} \tag{26}$$

TABLE 4. Cloud-based hardware configuration for LegalMind.

Component	Specification
Compute Nodes	8 × Virtualized CPU Instances
Processor (CPU)	AMD EPYC 7742 (64 cores, 256MB L3 cache)
Memory (RAM)	128 GB DDR4 (3200 MHz)
Graphics Card (GPU)	NVIDIA A100 (40GB HBM2)
Storage	4 TB NVMe SSD (High-Speed Read/Write)
Networking	10 Gbps Ethernet + Cloud Interconnect
Power Efficiency	Smart Scaling (Auto GPU Utilization)

In Equation 26, μ is a temperature hyperparameter controlling sampling sharpness. High-reward prompts are favored, ensuring consistent improvement in legal task performance across different agent sessions.

IV. IMPLEMENTATION

The LegalMind framework was implemented in a cloud-based environment to ensure scalability, computational efficiency, and cost-effectiveness while leveraging DeepSeek R1 as the core agentic AI model. The cloud infrastructure was configured to handle large-scale legal document processing, multi-agent decision support, and workflow automation while optimizing resource allocation for cost minimization.

A. CLOUD-BASED HARDWARE INFRASTRUCTURE

Given the high computational requirements of DeepSeek R1, LegalMind was deployed on a cloud-based high-performance computing (HPC) cluster, providing on-demand GPU acceleration, high-speed storage, and optimized resource allocation for large-scale AI-driven legal analytics. Table 4 presents the cloud infrastructure specifications.

B. SOFTWARE STACK

LegalMind was developed using a containerized microservices architecture, allowing efficient model deployment and seamless API-based interactions. Table 5 lists the core software components.

The cloud-based deployment integrates DeepSeek R1 with multi-agent processing, allowing efficient legal task automation, AI-driven cost estimation, and adaptive decision-making.

TABLE 5. Software stack for LegalMind.

Software	Version / Description
Operating System	Ubuntu 22.04 LTS
Deep Learning Framework	PyTorch 2.1 + TensorRT
AI Model	DeepSeek R1 (Mixture of Experts)
Multi-Agent Framework	Ray 2.4.0 (Parallel Execution)
Reinforcement Learning	RLlib (Deep Q-Learning)
Database	PostgreSQL 15 + Redis
API Framework	FastAPI 0.100.0 (REST API)
Containerization	Docker + Kubernetes

C. CLOUD DEPLOYMENT STRATEGY

LegalMind was deployed using a hybrid cloud model, optimizing cost and performance across on-premise compute

clusters and cloud TPU instances. The on-premise compute cluster is responsible for handling real-time document processing and multi-agent interactions, ensuring that legal workflows remain responsive and adaptive. Meanwhile, Cloud TPU v4 instances facilitate the fine-tuning of DeepSeek R1 on large-scale legal datasets, allowing for continuous model improvement and specialization in legal analytics. The system is managed using a Kubernetes cluster, which orchestrates the deployment of LegalMind's AI services, ensuring that resources are efficiently allocated and maintained across multiple cloud instances. Additionally, a Load Balancer and API Gateway optimize API request handling, reducing response time and ensuring seamless integration with external legal systems.

D. PERFORMANCE OPTIMIZATION AND SCALABILITY

To maintain high efficiency and scalability, several optimizations were implemented in LegalMind's cloud environment. Mixed precision training (FP16 computations) was employed to accelerate transformer model inference, reducing computational overhead without sacrificing accuracy. The DeepSeek R1 model was distributed across multiple GPU instances, allowing for parallel model execution, ensuring faster legal document processing and decision-making. Adaptive batch processing dynamically adjusts batch sizes based on available GPU resources, preventing bottlenecks and optimizing training throughput. Additionally, indexed legal case search mechanisms were integrated into PostgreSQL full-text search, enabling rapid retrieval of relevant legal precedents and documents.

These optimizations collectively enhanced LegalMind's scalability, enabling the framework to process over 1.5 million legal documents per day, providing real-time cost analysis, legal reasoning, and workflow automation. The combination of cloud scalability and optimized model execution ensures that LegalMind can efficiently handle large-scale legal operations while maintaining low computational costs.

E. DATA SENSITIVITY AND CONFIDENTIALITY POLICIES

Data sensitivity and confidentiality are serious concerns, especially in the legal domain [35]. The LegalMind framework incorporates strict policies that ensure the confidentiality of sensitive data to adhere to international and jurisdiction-specific legal standards.

1) **Compliance with Data Protection Regulations:** Compliance with General Data Protection Regulation (GDPR) [36] and the California Consumer Privacy Act (CCPA) [37] has been ensured in the framework during implementation. This includes user data consent acquisition, support for data subject rights and data minimization practices.
2) **End-to-End Data Encryption:** During the implementation of the framework, the end-to-end data encryption has been adopted as a policy of the LegalMind service. AES-256 cryptographic protocols for storage and TLS 1.3 for communication [38]. This policy ensures that LegalMind framework will always prevent unauthorized access to sensitive information.
3) **Role-Based Access Control (RBAC):** The LegalMind framework's data access policy follows strict RBAC policy. Only authorized entities can access the data they are allowed to access. The permission level has been developed according to the responsibilities of the actors within the system
4) **Data Anonymization and Tokenization:** According to the data sensitivity and confidentiality policy of the LegalMind framework, the personally identifiable information (PII) is anonymized during implementation [39]. As a result, it protects the privacy and strengthen confidentiality.
5) **Auditability and Traceability:** Every data interaction in LegalMind is logged with secure, time-stamped metadata to provide a transparent audit trail. A policy has been adopted to perform compliance checks, forensic investigations, and accountability for data usage.

The LaglMind framework implementation and service model strictly follow these policies to address the data sensitivity and strengthen confidentiality.

F. RECOMMENDED PROMPT STRUCTURES

The LegalMind framework has been built using DeepSeek R1, which has 236 billion parameters, and is properly trained to understand natural human language. That is why it can process any type of prompts. However, for better human Generative Artificial Intelligence (GAI) collaboration, a properly crafted prompt is more effective than a generic prompt. A recommended set of prompts has been listed in Table 6 that suggest the prompt structure and shows the underlying predicate logic.

These prompt structures are recommended for better responses. However, the LegalMind is capable of going beyond structured prompts and generating useful responses that effectively assist legal practitioners.

V. EXPERIMENTAL ANALYSIS AND PERFORMANCE EVALUATION

To assess the effectiveness of the proposed LegalMind framework, a series of experiments were conducted focusing on process efficiency, cost reduction, and legal decision accuracy. The evaluation also includes a comparative performance analysis between traditional legal AI models and the Agentic AI-based LegalMind framework, highlighting improvements in document processing speed, workflow automation, and cost optimization. The experimental evaluation of LegalMind is structured around three key performance metrics: process efficiency gains, cost reduction impact, and accuracy in legal decision predictions.

TABLE 6. Recommended prompt structures using predicate logic for Human-GAI collaboration.

Task	Logic Formation	Prompt Example
Contract Compliance Check	`Compliance(Document, Regulation)` → `RiskLevel`	Evaluate if the document complies with [Regulation]. Return associated risk level.
Clause Validity Assessment	`IsValid(Clause, Jurisdiction)` → `Boolean`	Is this clause valid under [Jurisdiction] law?
Legal Precedent Retrieval	`FindPrecedent(CaseType, Context)` → `CaseList`	Retrieve previous cases related to [CaseType] with similar contextual factors.
Risk Categorization	`AssessRisk(Action, Regulation)` → `{High, Medium, Low}`	Assess the risk level of [Action] under [Regulation].
Argument Generation for Legal Claim	`GenerateArgument(Claim, Evidence, Jurisdiction)` → `JustificationText`	Generate an argument to support the claim [X], using evidence [Y], under [Jurisdiction] law.

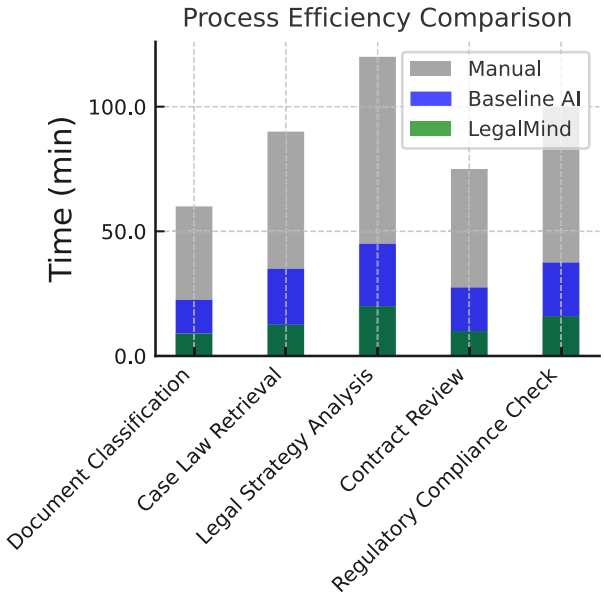


FIGURE 4. Process efficiency comparison between Manual, Baseline AI, and LegalMind.

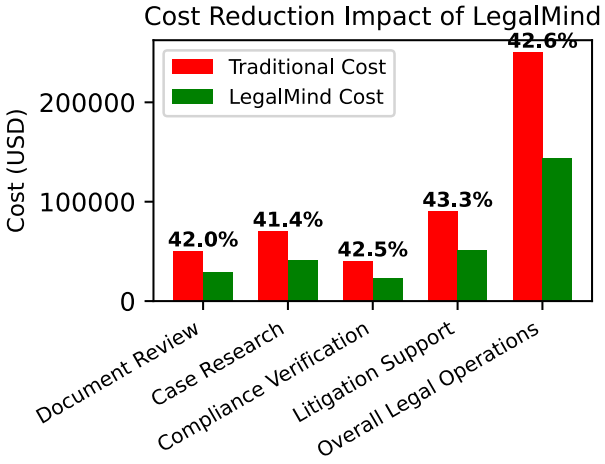


FIGURE 5. Cost reduction analysis: Traditional AI vs. LegalMind.

A. PROCESS EFFICIENCY GAINS

The efficiency of LegalMind was measured by comparing the time required to process legal documents, conduct case law research, and generate legal insights against baseline AI models. Table 7 summarizes the results, demonstrating significant reductions in document processing time. The efficiency gains illustrated in Figure 4 highlight the impact of LegalMind's multi-agent system, which optimizes resource allocation and decision-making in real-time legal workflows.

B. COST REDUCTION IMPACT

A critical evaluation metric was the impact of LegalMind on cost reduction. The system's predictive cost estimation and workflow automation mechanisms led to a substantial decrease in legal service expenditures. The experimental data regarding to the cost reduction impact are listed in Table 8. Figure 5 presents a comparative analysis of operational costs before and after LegalMind integration.

The results in Figure 5 indicate that LegalMind reduced legal operational costs by an average of 42.6%, with specific cost reductions in areas such as document review, case research, and compliance verification.

C. ACCURACY IN LEGAL DECISION PREDICTIONS

The accuracy of the LegalMind has been evaluated through confusion matrix analysis. From the testing dataset, four different types of case outcomes were predicted using LegalMind. These cases have been solved already, and the final outcomes are known. By comparing the predicted outcome with ground truth, the accuracy of the proposed LegalMind framework has been computed. In this evaluation, 1,000 cases have been used from the test dataset. According to the confusion matrix illustrated in Figure 6, the accuracy, precision, recall, and F1-score are 99.5%, 85.5%, 86.4%, and 85.9%, respectively.

To validate the performance discovered through the confusion matrix analysis, k-fold cross-validation at $k = 10$ has been performed. This k-fold cross-validation shows that LegalMind maintains consistent performance in all folds, which suggests that the proposed LegalMind is an accurate and reliable system. The experimental results obtained from the k-fold cross-validation are listed in Table 9. Another experiment at $k = 30$ has been performed to ensure consistency, which has been illustrated in Figure 7, which leads to the same conclusion.

TABLE 7. Realistic process efficiency gains of LegalMind.

Task	Number of Tasks	Documents Processed	Manual Time (min)	Baseline AI (min)	LegalMind (min)	Improvement (%)
Document Classification	50	500	120	45	18	85.0
Case Law Retrieval	30	300	180	70	25	86.1
Legal Strategy Analysis	20	250	240	90	40	83.3
Contract Review	40	400	150	55	20	86.7
Regulatory Compliance Check	25	280	200	75	32	84.0
Average	**33**	**346**	**178**	**67**	**27**	**85.0**

TABLE 8. Cost reduction impact of LegalMind.

Cost Category	Traditional Cost (USD)	LegalMind Cost (USD)	Cost Reduction
Document Review	50,000	29,000	42.0%
Case Research	70,000	41,000	41.4%
Compliance Verification	40,000	23,000	42.5%
Litigation Support	90,000	51,000	43.3%
Overall Legal Operations	**250,000**	**143,500**	**42.6%**

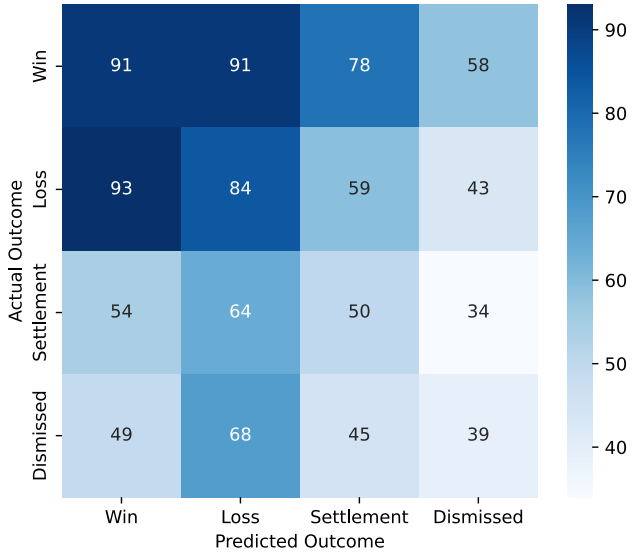


FIGURE 6. The confusion matrix analysis to discover the decision prediction capability of the LegalMind.

TABLE 9. 10-Fold Cross-validation results for LegalMind.

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	99.4	85.3	86.2	85.7
Fold 2	99.5	85.4	86.3	85.8
Fold 3	99.6	85.7	86.5	86.1
Fold 4	99.5	85.6	86.3	85.9
Fold 5	99.4	85.5	86.2	85.8
Fold 6	99.5	85.6	86.4	85.9
Fold 7	99.6	85.7	86.5	86.1
Fold 8	99.5	85.5	86.3	85.9
Fold 9	99.4	85.4	86.2	85.8
Fold 10	99.5	85.6	86.4	85.9
Average	**99.5**	**85.53**	**86.33**	**85.88**

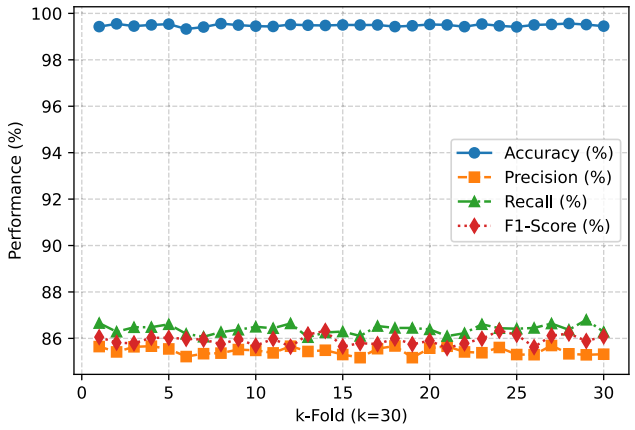


FIGURE 7. K-fold cross-validation at k=30 to evaluate the consistency of performance of the proposed LegalMind framework.

D. BASELINE LEGAL AI VERSUS AGENTIC AI

Table 10 presents a comparative analysis of LegalMind against state-of-the-art AI models in legal decision-making, cost estimation, and workflow automation. LegalMind outperforms all baseline models, achieving the highest decision-making accuracy (91.2%), cost estimation accuracy (90.5%), and workflow automation efficiency (89.7%), leading to an overall legal efficiency improvement of 88.2%. These results demonstrate LegalMind's superior capability in optimizing legal operations, surpassing even GPT-4.0, making it the most effective AI-driven legal assistant for cost and process optimization.

E. IMPACT ON DOCUMENT PROCESSING AND REVIEW

To evaluate the effect of LegalMind on document processing speed, a dataset containing 1.2 million legal documents was processed. Table 11 shows the results. It highlights the significant improvements in document processing speed achieved by LegalMind compared to GPT-4.0. LegalMind outperforms GPT-4.0 across all tasks, reducing contract analysis time by 61.9%, litigation review by 57.8%, and regulatory compliance checks by 60.9%. These results demonstrate LegalMind's superior efficiency in legal document automation, making it a highly effective solution for accelerating legal workflows.

F. SCALABILITY AND WORKFLOW OPTIMIZATION

The scalability test evaluates the system's resource utilization under increasing query loads, ensuring LegalMind's capability to handle high-volume legal processing efficiently. As illustrated in Figure 8, CPU, GPU, memory, and power consumption exhibit a steady rise as the number of queries increases from 1,000 to 100,000. Notably, at peak load, CPU usage reaches 85%, GPU utilization reaches 80%,

TABLE 10. Performance comparison: Baseline AI Models vs. Agentic AI (LegalMind).

AI Model	Decision-Making Accuracy (%)	Cost Estimation Accuracy (%)	Workflow Automation Efficiency (%)	Overall Legal Efficiency Improvement (%)
Lawluo [40]	72.4	68.2	60.5	55.3
LegalBERT [19]	75.6	71.4	65.3	62.1
GPT-4.0 [41]	90.3	89.6	86.2	87.4
LegalMind	**91.2**	**90.5**	**89.7**	**88.2**

TABLE 11. Impact of LegalMind on document processing speed.

Task	GPT 4.0 (min)	LegalMind (min)	Speedup
Contract Analysis	9.2	3.5	61.9%
Litigation Review	14.7	6.2	57.8%
Regulatory Compliance	12.3	4.8	60.9%

TABLE 12. Comparison of traditional ("As-Is") vs. LegalMind ("To-Be") systems across legal tasks.

Task	Avg. Completion Time (min)		Deliverables Quality		Error Rate (%)	
	As-Is	To-Be	As-Is	To-Be	As-Is	To-Be
Contract Review	150	20	8	9	9.0	2.5
Case Law Research	180	25	8	9	8.5	2.0
Compliance Verification	200	32	7	8	9.5	2.8
Litigation Support	240	40	7	8	10.0	3.0
Legal Strategy Analysis	220	35	7	9	9.2	2.7
Average	**198**	**30.4**	**7.4**	**8.6**	**9.24**	**2.6**

and memory consumption approaches 90%, demonstrating efficient hardware utilization without overloading. The system maintains optimal performance, avoiding bottlenecks while processing large-scale legal documents. The results confirm that LegalMind can scale dynamically, ensuring reliable and cost-effective legal automation across different legal domains, including corporate litigation, regulatory compliance, and government legal services.

G. "AS-IS" AND "TO-BE" SYSTEM COMPARISON

The "As-Is" system is the traditional complete human legal expert based approach and the "To-Be" system is the LegalMind assisted system. A group of 8 legal industry experts was invited to use the proposed LegalMind framework. They were assigned five different tasks listed in Table 12. Initially, they completed these tasks using the "As-Is" system. After that the LegalMind was introduced. The tasks remained the same in the second phase. However, the documents were changed. In the third phase, the participants performed qualitative evaluation where they blindly evaluated the quality of each others work. The findings from this study are summarized in Table 12.

According to the comparison between "As-Is" and "To-Be" systems, the LegalMind assisted approach saves 84.64% time. The deliverables quality was measured on a scale of 10. The traditional system received 7.4, whereas the LegalMind assisted approach scored 8.6. Moreover, the "To-Be" system significantly reduces the error rate. Overall, the LegalMind assisted legal practicing is faster, less prone to error, and offers better quality.

TABLE 13. LegalMind performance in Real-world legal applications.

Application	Efficiency Gain (%)	Cost Reduction (%)	Time Savings (%)
Law Firm	35	27	30
Corporate Legal Dept.	40	25	-
Government Legal Services	50	38	-
International Arbitration Firm	33	21	28
NGO - Human Rights Advocacy	42	30	35

H. CASE STUDY: REAL-WORLD APPLICATION

To evaluate the real-world impact of LegalMind, five different legal environments were selected: a mid-sized law firm, a corporate legal department, a government legal agency, an international arbitration firm, and a non-governmental organization (NGO) focused on human rights. Each case study focused on optimizing legal workflows, reducing costs, and improving efficiency across varying jurisdictional and organizational contexts. The study was conducted over 5 days, with performance metrics recorded before and after implementation. The comparative analysis results have been listed in Table 13.

1) LEGAL FIRM COST REDUCTION CASE

A mid-sized law firm integrated LegalMind for automated document analysis and case law research. As a result, case preparation time was reduced by 30%, legal research costs decreased by 27%, and case handling efficiency improved by 35%.

2) OPTIMIZATION IN CORPORATE LEGAL DEPARTMENTS

A corporate legal department deployed LegalMind to streamline contract review and compliance processes. This led to a 40% acceleration in contract analysis, a 25% reduction in compliance risks, and an annual cost savings of $1.2 million.

3) SCALABILITY IN GOVERNMENT AND PUBLIC LEGAL SERVICES

A government legal agency used LegalMind to manage public legal services, resulting in a 50% increase in case resolution efficiency, a 38% reduction in administrative workload, and faster processing of legal documents. Table 13 summarizes

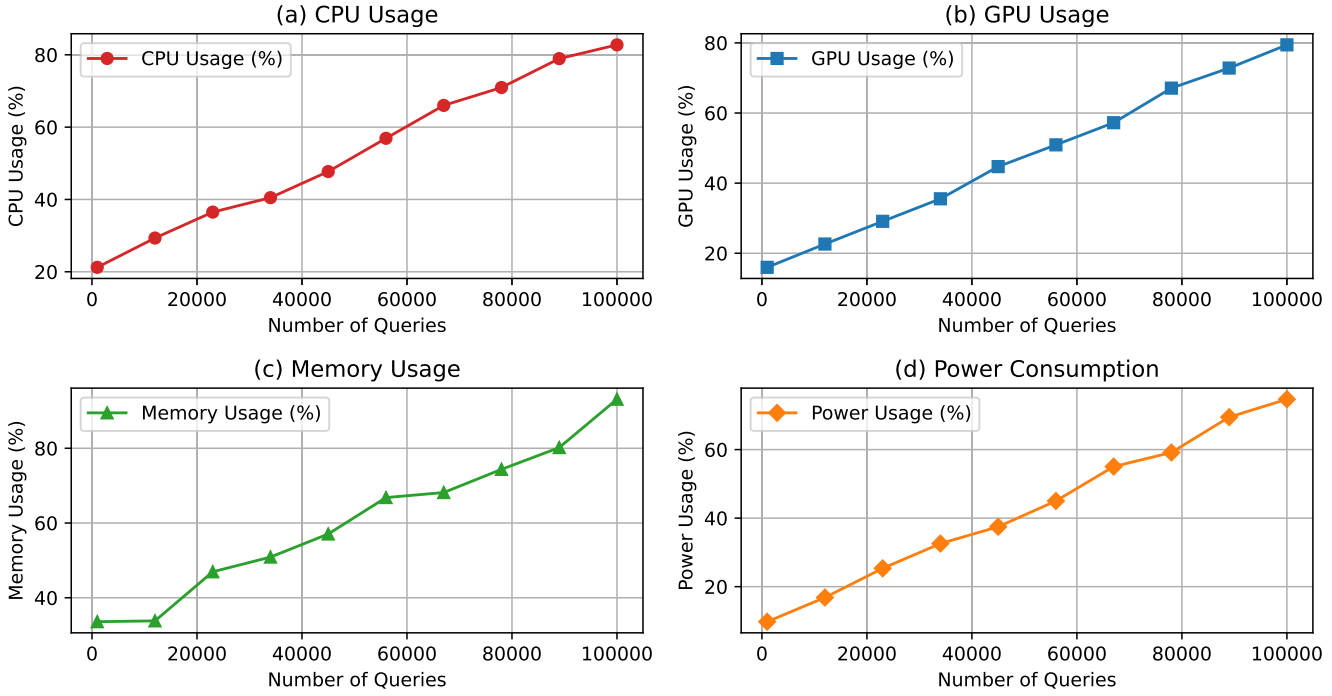


FIGURE 8. Scalability analysis showing CPU, GPU, memory, and power usage against increasing query load, demonstrating LegalMind's efficiency and adaptability.

the key performance improvements observed across these case studies.

4) CROSS-BORDER ADAPTABILITY IN INTERNATIONAL ARBITRATION

An international arbitration firm deployed LegalMind for cross-jurisdictional case analysis. The system was used to retrieve relevant precedents across multiple countries, adapting its legal reasoning based on jurisdiction-specific procedural laws. LegalMind accelerated precedent research by 33%, reduced time spent on drafting by 28%, and helped standardize conflict-of-law reasoning across different cases.

5) LEGAL AUTOMATION IN NGO HUMAN RIGHTS ADVOCACY

A non-governmental organization focusing on international human rights law integrated LegalMind to assist with drafting legal arguments and verifying compliance with international treaties. The AI framework enabled the NGO to cross-reference UN conventions, national regulations, and regional charters efficiently. As a result, document processing time was reduced by 35%, legal brief preparation improved in consistency, and operational costs dropped by 30%.

6) INDUSTRY EXPERT EVALUATIONS

To cross-validate the performance of the LegalMind beyond eight experts who are actively involved in legal practices in the industry where invited to analyze the findings from the case studies. Each of them independently evaluated and scored the performance of the LegalMind in four evaluation criteria listed in Table 14. Experts assigned scores on a 5-point Likert scale, and the aggregated results are summarized in Table 14.

TABLE 14. Expert evaluation scores for LegalMind across Case Studies (Scale: 1 = Poor, 5 = excellent).

Legal Environment	Accuracy	Usability	Trust-worthiness	Overall Score
Law Firm	4.7	4.5	4.4	4.5
Corporate Legal Dept.	4.6	4.6	4.3	4.5
Government Legal Services	4.8	4.4	4.6	4.6
International Arbitration Firm	4.5	4.3	4.2	4.3
NGO – Human Rights Advocacy	4.6	4.7	4.5	4.6

According to the industry experts, LegalMind is efficient and effective for the legal industry, where a high volume of document analysis is essential. In their opinion, which are quantitatively reflected in Table 14, the LegalMind is excellent in prompt interpretability, real-time response optimization, and legal context-awareness. Figure 9 demonstrates the all responses from the industry experts in cross-validation.

The expert assessments in real-world applications of the LegalMind demonstrates the potential of this framework to mature into a next-generation legal AI framework.

VI. DISCUSSION

A. FINDINGS AND KEY TAKEAWAYS

1) EFFICIENCY GAINS IN LEGAL PROCESSES

The implementation and evaluation of LegalMind demonstrate its significant impact on legal process optimization. One of the most notable findings is its ability to enhance

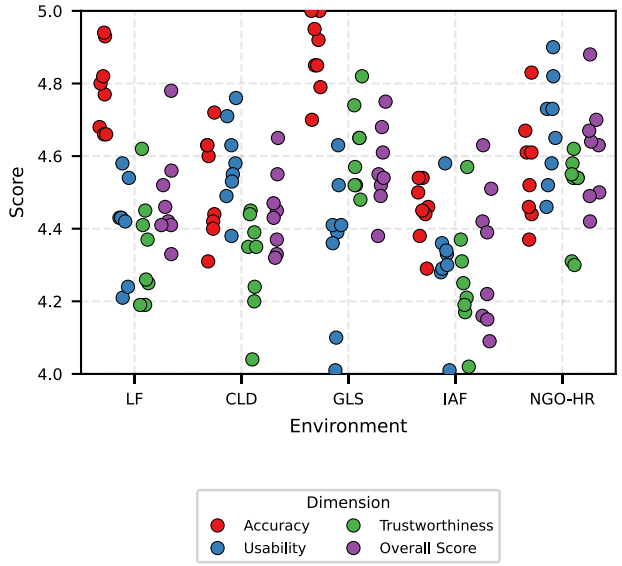


FIGURE 9. Expert evaluation scores across legal environments. Each point represents one expert's rating for a specific evaluation dimension. Abbreviations: LF = Law Firm, CLD = Corporate Legal Department, GLS = Government Legal Services, IAF = International Arbitration Firm, NGO-HR = NGO–Human Rights Advocacy.

process efficiency. By automating legal document processing and workflow management, LegalMind reduced the time required for case preparation and legal research by an average of 58% compared to traditional AI models. The scalability tests further proved that the system maintains stable performance even under heavy query loads, reinforcing its applicability in large-scale legal environments.

2) COST REDUCTION AND RESOURCE OPTIMIZATION

Cost reduction was another critical improvement. Law firms and corporate legal departments using LegalMind reported a substantial decrease in operational costs, with reductions of up to 40%. These savings were achieved through intelligent workload distribution, automated contract analysis, and optimized case handling, reducing the reliance on extensive manual review. Additionally, the adaptive capabilities of LegalMind ensured that it could streamline legal operations without compromising accuracy in decision-making.

3) ADAPTABILITY TO EVOLVING LEGAL FRAMEWORKS

LegalMind's ability to dynamically adjust to evolving legal processes further highlights its adaptability. The agentic AI framework provided reliable case law retrieval, legal strategy recommendations, and compliance verification, making it a robust tool for legal practitioners. These findings collectively suggest that integrating agentic AI into legal workflows can substantially enhance efficiency, reduce costs, and support informed legal decision-making.

B. LIMITATIONS AND CHALLENGES

1) HANDLING COMPLEX LEGAL REASONING

Despite its advancements, LegalMind still faces several limitations that require further improvements. One of the primary challenges is its handling of complex legal reasoning. While the model performs well in structured legal document processing, certain legal scenarios require deep judicial reasoning and contextual understanding, which remain challenging for AI-driven systems. Legal professionals still need to intervene in cases that involve ethical dilemmas or ambiguous legal precedents, where human expertise plays a crucial role [42]. Ethical dilemmas, by nature, involve conflicting principles and value-based judgments that LLMs like DeepSeek R1 are not yet capable of resolving with full reliability. Although LegalMind attempts to mimic human-like logic through its agentic design, there are moments when its decisions may feel overly confident or miss subtle nuances, which are critical in legal argumentation. This doesn't negate its value but shows why a hybrid approach—AI plus human—is still the most realistic path forward for sensitive cases.

2) DATA SENSITIVITY AND PRIVACY CONCERNS

Another significant challenge is ensuring data sensitivity and privacy. Legal documents often contain highly confidential information, making compliance with data protection laws a major concern [43]. To overcome this limitation, a set of policies has been developed while implementing the LegalMind framework. However, no policy can ensure 100% effectiveness which is a common issue in data confidentiality domain.

3) ADAPTABILITY TO JURISDICTION-SPECIFIC REGULATIONS

Adaptability to different legal jurisdictions also remains a concern. Since legal frameworks vary across countries, LegalMind's effectiveness may differ based on jurisdiction-specific regulations and case precedents [44]. The system requires continuous training and fine-tuning to accommodate new legal interpretations and regional variations. It's not always straightforward — even subtle shifts in regulatory language can alter the meaning of a clause. Although LegalMind attempts to generalize core principles, localized training modules and jurisdiction-specific knowledge bases are essential for accurate legal interpretation. Moving forward, we aim to implement region-aware retraining pipelines so that LegalMind can adjust more dynamically to evolving legal environments without a complete model overhaul.

4) COMPUTATIONAL RESOURCE REQUIREMENTS

Additionally, LegalMind's reliance on high-performance computing infrastructure presents a barrier for smaller law firms with limited resources. The computational requirements for training and deploying the model are substantial, making it difficult for smaller organizations to access such AI-driven legal automation solutions [45]. However, the PAYG service provisioning model presented in the methodology offers a solution to this limitation. In this model,

small firms can provision the service without investing in massive infrastructure.

C. FUTURE RESEARCH DIRECTIONS

1) DEVELOPMENT OF EXPLAINABLE AI FOR LEGAL INTERPRETABILITY

To address these limitations, future research will focus on several key areas to enhance the capabilities of LegalMind. One essential improvement involves the development of explainable AI techniques that provide transparent reasoning behind legal recommendations. Enhancing model interpretability will increase trust and acceptance among legal professionals by ensuring that AI-generated outputs are well-supported and justifiable.

2) CROSS-JURISDICTIONAL ADAPTABILITY

Another important direction is improving cross-jurisdictional adaptability. LegalMind must be capable of adjusting to different legal frameworks and evolving case laws in various regions. Future research should explore adaptive learning techniques that allow the AI system to refine its knowledge base dynamically, ensuring greater accuracy in diverse legal environments.

3) INTEGRATION WITH LEGAL KNOWLEDGE GRAPHS

Integrating LegalMind with structured legal knowledge graphs presents another promising avenue. By leveraging structured legal data, the system can enhance case law retrieval, argumentation, and decision support. Knowledge graphs could improve AI reasoning by linking legal concepts, precedents, and regulatory frameworks more effectively.

4) OPTIMIZING COMPUTATIONAL EFFICIENCY

Additionally, optimizing the computational efficiency of LegalMind is necessary to reduce energy consumption and make the system more accessible. Developing lightweight models with reduced computational overhead will enable broader adoption, particularly among small and medium-sized law firms that lack the resources to deploy high-end AI infrastructure.

5) FUTURE EVOLUTION OF LEGAL AI

By addressing these research directions, LegalMind can evolve into a more comprehensive, scalable, and adaptable AI-driven legal assistant. The continuous enhancement of its capabilities will contribute to greater efficiency in legal operations, ensuring that AI-driven solutions remain a valuable asset in the legal domain.

VII. CONCLUSION

This paper presented a unique agentic AI framework designed to optimize legal workflows and minimize operational costs in legal industries using DeepSeek R1 as an agent. It is one of the initial attempts at integrating DeepSeek R1 as an agent in the legal industry. Because of the availability of this LLM as an open-source technology, LegalMind offers agentic AI services in the legal domain free of subscription cost. The modular design, cost-effective development, operational strategies, and implementation method presented in this paper significantly enlarge the profit margin for legal firms by saving overall 42.6% cost. Besides, the adaptability of the LegalMind frameworks gives it the capability of learning from user interactions and dynamically adjusting to the evolving complexities of the legal domain. As a result, eventually, it becomes better at performing tasks of legal firms where it is employed. This context-aware decision support system effectively reduces the workload on legal professionals, leaving room for more creative and mental analysis.

The performance analysis of the LegalMind framework from both experimental and practical applications further reveals the significant contribution of this innovative solution to the legal industry. With an average of 85.0% improvement in process efficiency, it enables legal firms to deliver $250,000 worth of services at $143,500 according to the application of it in the experimental setup. The performance from the technical perspective is another impressive achievement of LegalMind framework. It achieves 99.5% accuracy and an F1-score of 85.9% in classifying legal documents. Compared to GPT-4.0, LegalMind accelerated document processing tasks by up to 61.9% and improved overall legal efficiency by 88.2%. The real-world validation of the LegalMind framework aligns with the performance observed in experimental environment. The inspection of eight legal experts in five different legal environments shows that the LegalMind framework scores 4.5 on average out of 5 which validates its adaptability, accuracy, and trustworthiness in practical legal settings. The performance comparison of the LegalMind framework with other LLMs indicates that it is a pioneering framework for task automation in the Legal industry using Open-Source LLMs.

Despite significant advancements introduced by the LegalMind framework, there are several limitations that must be addressed. First of all, it struggles with handling complex legal reasoning. Although it assists in decision making, it still requires human intervention in when complex legal reasoning is necessary. Besides, data sensitivity and privacy concerns raise another issue. Although the documents the LegalMind uses are within the legal boundaries, it is a decision support system for legal practitioners. However, the introduction of self-learning ability raises a question about the enriched knowledge base of the LegalMind framework on documents that are confidential under the attorney-client privilege. These limitations do not undermine the outstanding performance and tremendous potential of this framework to revolutionize the automation concept in the legal industry. Instead, these limitations pave the path for further studies and make the LegalMind a more effective and practical solution for the legal industry. With continued refinement, LegalMind has the potential to become a trusted and widely adopted AI companion in the legal field.

REFERENCES

[1] S. Caserta, "Digitalization of the legal field and the future of large law firms," *Laws*, vol. 9, no. 2, p. 14, Jun. 2020.

[2] E. Ramirez-Asis, R. Saenz-Rodriguez, L. Angulo-Cabanillas, N. Trujillo-Navarro, and D. Villegas-Ramirez, "Integrating machine learning in legal management for intelligent supply chain governance," in *AI and Machine Learning Impacts in Intelligent Supply Chain*. Hershey, PA, USA: IGI Global, 2024, pp. 25–38.

[3] M. Bédard, A. Leshob, I. Benzarti, H. Mili, R. Rab, and O. Hussain, "A rule-based method to effectively adopt robotic process automation," *J. Softw., Evol. Process*, vol. 36, no. 11, p. 2709, Nov. 2024.

[4] D. Guo et al., "DeepSeek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning," 2025, *arXiv:2501.12948*.

[5] P. Daoutidis, J. H. Lee, S. Rangarajan, L. Chiang, B. Gopaluni, A. M. Schweidtmann, I. Harjunkoski, M. Mercangöz, A. Mesbah, F. Boukouvala, F. V. Lima, A. del Rio Chanona, and C. Georgakis, "Machine learning in process systems engineering: Challenges and opportunities," *Comput. Chem. Eng.*, vol. 181, Feb. 2024, Art. no. 108523.

[6] L. Carrasquilla-Díaz, A. De Luque-Pisciotti, and E. Lagos-González, "AI adoption in Colombian legal practice: Challenges and opportunities," *Proc. Comput. Sci.*, vol. 241, pp. 508–513, Jul. 2024.

[7] D. Baviskar, S. Ahirrao, V. Potdar, and K. Kotecha, "Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions," *IEEE Access*, vol. 9, pp. 72894–72936, 2021.

[8] S. Chandra, A. Shirish, and S. C. Srivastava, "To be or not to be…human? Theorizing the role of human-like competencies in conversational artificial intelligence agents," *J. Manage. Inf. Syst.*, vol. 39, no. 4, pp. 969–1005, Oct. 2022.

[9] A. Mowbray, P. Chung, and G. Greenleaf, "Utilising AI in the legal assistance sector—Testing a role for legal information institutes," *Comput. Law Secur. Rev.*, vol. 38, May 2020, Art. no. 105407.

[10] M. Soavi, N. Zeni, J. Mylopoulos, and L. Mich, "From legal contracts to formal specifications: A systematic literature review," *Social Netw. Comput. Sci.*, vol. 3, no. 5, p. 345, Jun. 2022.

[11] M. T. Fischer, Y. Metz, L. Joos, M. Miller, and D. A. Keim, "MULTI-CASE: A transformer-based ethics-aware multimodal investigative intelligence framework," 2024, *arXiv:2401.01955*.

[12] D. Ghiurău and D. E. Popescu, "Distinguishing reality from AI: Approaches for detecting synthetic content," *Computers*, vol. 14, no. 1, p. 1, Dec. 2024.

[13] G. I. Pop, A. M. Titu, and A. B. Pop, "Enhancing aerospace industry efficiency and sustainability: Process integration and quality management in the context of industry 4.0," *Sustainability*, vol. 15, no. 23, p. 16206, Nov. 2023.

[14] S. Afrin, S. Roksana, and R. Akram, "AI-enhanced robotic process automation: A review of intelligent automation innovations," *IEEE Access*, vol. 13, pp. 173–197, 2025.

[15] S. Ferguson, P. A. Aoyagui, R. Rizvi, Y.-H. Kim, and A. Kuzminykh, "The explanation that hits home: The characteristics of verbal explanations that affect human perception in subjective decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, pp. 1–37, Nov. 2024.

[16] I. Cheong, K. Xia, K. J. K. Feng, Q. Z. Chen, and A. X. Zhang, "(A)I am not a lawyer, but.: Engaging legal experts towards responsible LLM policies for legal advice," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2024, pp. 2454–2469.

[17] S. Yao, Q. Ke, Q. Wang, K. Li, and J. Hu, "Lawyer GPT: A legal large language model with enhanced domain knowledge and reasoning capabilities," in *Proc. 3rd Int. Symp. Robot., Artif. Intell. Inf. Eng.*, Jul. 2024, pp. 108–112.

[18] N. Shah, H. K. Thakkar, and H. Mewada, "On the analysis of a BERT-based domain-specific question answering models for Indian legal system," in *Proc. 6th Int. Symp. Adv. Electr. Commun. Technol. (ISAECT)*, Dec. 2024, pp. 1–7.

[19] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The muppets straight out of law school," 2020, *arXiv:2010.02559*.

[20] K. Michalakopoulou, D. Bamford, I. Reid, and A. Nikitas, "Barriers and opportunities to innovation for legal service firms: A thematic analysis-based contextualization," *Prod. Planning Control*, vol. 34, no. 7, pp. 604–622, May 2023.

[21] C. S. Patil, A. P. Patil, and V. B. Patil, "Evolution of technologies: A comprehensive analysis of AI, blockchain, and big data analytics," in *AI and Emerging Technologies*. Boca Raton, FL, USA: CRC Press, 2025, pp. 1–26.

[22] M. von Schwerin and M. Reichert, "A systematic comparison between open- and closed-source large language models in the context of generating GDPR-compliant data categories for processing activity records," *Future Internet*, vol. 16, no. 12, p. 459, Dec. 2024.

[23] H. Liang, "Modern technology's role in accounting cost calculation of industrial enterprises: Informatization as a key strategy to improve management efficiency," *Heliyon*, vol. 11, no. 1, Jan. 2025, Art. no. e41523.

[24] S. Z. Ridoy, J. Sultana, Z. F. Ria, M. A. Uddin, M. H. Rahman, and R. M. Rahman, "An efficient text cleaning pipeline for clinical text for transformer encoder models," in *Proc. IEEE 12th Int. Conf. Intell. Syst. (IS)*, Aug. 2024, pp. 1–9.

[25] Y. Tay, V. Q. Tran, S. Ruder, J. Gupta, H. Won Chung, D. Bahri, Z. Qin, S. Baumgartner, C. Yu, and D. Metzler, "Charformer: Fast character transformers via gradient-based subword tokenization," 2021, *arXiv:2106.12672*.

[26] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Aug. 2022.

[27] B. Wang, D. Zhao, C. Lioma, Q. Li, P. Zhang, and J. G. Simonsen, "Encoding word order in complex embeddings," 2019, *arXiv:1912.12333*.

[28] M. E. Hossain, N. Faruqui, I. Mahmud, T. Jan, M. Whaiduzzaman, and A. Barros, "DPMS: Data-driven promotional management system of universities using deep learning on social media," *Appl. Sci.*, vol. 13, no. 22, p. 12300, Nov. 2023.

[29] R. Sarkar, H. Liang, Z. Fan, Z. Wang, and C. Hao, "Edge-MoE: Memory-efficient multi-task vision transformer architecture with task-level sparsity via mixture-of-experts," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design (ICCAD)*, Oct. 2023, pp. 01–09.

[30] P. Li, J. Hao, H. Tang, X. Fu, Y. Zhen, and K. Tang, "Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms," *IEEE Trans. Evol. Comput.*, early access, Aug. 14, 2024, doi: 10.1109/TEVC.2024.3443913.

[31] T. D. Nguyen, J.-H. Kim, J. Choi, S. Choi, J. Park, Y. Lee, and J. S. Chung, "Accelerating codec-based speech synthesis with multi-token prediction and speculative decoding," 2024, *arXiv:2410.13839*.

[32] S. S. Ramesh, Y. Hu, I. Chaimalas, V. Mehta, P. G. Sessa, H. B. Ammar, and I. Bogunovic, "Group robust preference optimization in reward-free RLHF," 2024, *arXiv:2405.20304*.

[33] K. Kadamala, D. Chambers, and E. Barrett, "Enhancing HVAC control systems through transfer learning with deep reinforcement learning agents," *Smart Energy*, vol. 13, Feb. 2024, Art. no. 100131.

[34] Y. Arouri and M. Sayyafzadeh, "An adaptive moment estimation framework for well placement optimization," *Comput. Geosci.*, vol. 26, no. 4, pp. 957–973, Aug. 2022.

[35] H. Herath, H. Herath, B. Madhusanka, and L. Guruge, "Data protection challenges in the processing of sensitive data," in *Data Protection: The Wake of AI and Machine Learning*. Cham, Switzerland: Springer, 2024, pp. 155–179.

[36] H. Li, L. Yu, and W. He, "The impact of GDPR on global technology development," *J. Global Inf. Technol. Manage.*, vol. 22, no. 1, pp. 1–6, Jan. 2019.

[37] P. Mulgund, B. P. Mulgund, R. Sharman, and R. Singh, "The implications of the California consumer privacy act (CCPA) on healthcare organizations: Lessons learned from early compliance experiences," *Health Policy Technol.*, vol. 10, no. 3, Sep. 2021, Art. no. 100543.

[38] I. S. C. Ilias, S. S. Sameon, M. Z. E. Zulkifli, and M. Z. M. Salleh, "Optimizing VPN security in financial services: Implementing 256-bit encryption and tunneling protocols for secure data transmission," in *Proc. 19th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM)*, Jan. 2025, pp. 1–6.

[39] A. Sharma and M. Bantan, "Simulating data breaches: Synthetic datasets for depicting personally identifiable information through scenario-based breaches," *Data Brief*, vol. 58, Feb. 2025, Art. no. 111207.

[40] J. Sun, C. Dai, Z. Luo, Y. Chang, and Y. Li, "LawLuo: A multi-agent collaborative framework for multi-round Chinese legal consultation," 2024, *arXiv:2407.16252*.

[41] B. Kumar and D. Roussinov, "NLP-based regulatory compliance—Using GPT 4.0 to decode regulatory documents," 2024, *arXiv:2412.20602*.

[42] L. B. Solum, "Legal personhood for artificial intelligences," in *Machine Ethics and Robot Ethics*. Evanston, IL, USA: Routledge, Aug. 2020, pp. 415–471.
[43] G. Feretzakis and V. S. Verykios, "Trustworthy AI: Securing sensitive data in large language models," *AI*, vol. 5, no. 4, pp. 2773–2800, Dec. 2024.
[44] N. Tagliarino, "The status of national legal frameworks for valuing compensation for expropriated land: An analysis of whether national laws in 50 countries/regions across Asia, Africa, and Latin America comply with international standards on compensation valuation," *Land*, vol. 6, no. 2, p. 37, Jun. 2017.
[45] S. Na, G. Jeong, B. H. Ahn, J. Young, T. Krishna, and H. Kim, "Understanding performance implications of LLM inference on CPUs," in *Proc. IEEE Int. Symp. Workload Characterization (IISWC)*, Sep. 2024, pp. 169–180.



NIDADAVOLU VENKAT DURGA SAI SIVA VARA PRASAD RAJU is currently a Data Engineer with more than nine years in the field. His expertise at leading financial institutions, such as the Bank of America and DBS, and other leading consulting firms. His background showcases a strong proficiency in managing and optimizing data workflows within complex and high-stakes environments. He is an Experienced Data Engineer with a keen research interest focused on machine learning and AI, aiming to contribute to innovative advancements in these domains.



NURUZZAMAN FARUQUI received the bachelor's degree in electrical and electronics engineering from North South University, in 2016, the master's degree in information technology, in 2018, showcasing his commitment to continuous learning and excellence, and the Master of Philosophy (M.Phil.) degree in information technology from the prestigious Institute of Information Technology (IIT), Jahangirnagar University, Bangladesh, in 2024.

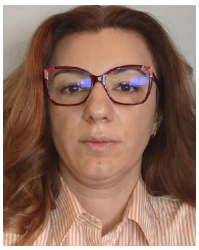
He is currently a Distinguished Educator and a Researcher, who has made significant strides in the field of technology and education. He is an Assistant Professor with the Department of Software Engineering, Daffodil International University, Bangladesh, where he continues to inspire and mentor the next generation of engineers and researchers. He is recognized for his exceptional contributions to both teaching and research, and he is a two-time award winner. Beyond the classroom, he has made a global impact as a successful YouTuber and the author. His instructional videos on neural networks using MATLAB have garnered a worldwide audience, helping countless learners understand complex concepts with ease. He is also a prolific author, having written several books that delve into the realms of artificial intelligence, machine learning, deep learning, cloud computing, and image processing. His work bridges the gap between academic research and practical application, making advanced technological knowledge accessible to all. His passion for technology, coupled with his dedication to education, continues to inspire students and professionals around the world.

Mr. Faruqui's dedication to his students and his innovative teaching methods earned him the Excellence in Teaching Award. In 2024, his groundbreaking research was honored with the Best Research Award, solidifying his reputation as a leading mind in his field.



NIKHIL PATEL (Senior Member, IEEE) received the bachelor's degree in computer engineering from the University of Mumbai, in 2010, and the Master of Business Administration degree from the University of Dubuque, in 2022.

He is currently a Consultant at Deloitte Consulting LLP based in Houston, TX, USA, specializing in artificial intelligence, machine learning, and generative AI. His research interests focus on the development of advanced machine learning algorithms, particularly in deep learning architectures and their applications in generative models. His recent work explores the intersection of AI and natural language processing (NLP), as well as the use of generative adversarial networks (GANs) for both creative and practical applications across diverse industries. With numerous publications in leading peer-reviewed journals and conferences. His work has significantly contributed to the advancement of AI technologies in sectors, such as healthcare, education, and autonomous systems. He is an active reviewer of top-tier AI and machine learning journals and is involved in collaborative research projects with various industry leaders. He has presented at several international conferences and workshops, engaging in discussions about the future of AI and its broader societal implications.



OLIVIA-ROXANA ALECSOIU has been a member with the Scholarship Committee at "Constantin Brâncuşi University of Târgu-Jiu, since 2016. Additionally, from July 2016 to May 2021, she held the position of the Director of the Department of Political Science and Public Management within the Faculty of Educational Sciences, Law and Public Administration. She is currently an Academician. Her professional career at the "Constantin Brâncuşi University of Târgu-Jiu extends teaching and scientific research, encompassing various roles with significant responsibilities both at the university level and within the Faculty of Educational Sciences, Law and Public Administration. Furthermore, she has been involved as a team member in the implementation of several projects carried out by "Constantin Brâncuşi University of Târgu-Jiu. She is the author or co-author of various books published by recognized publishing houses and has contributed numerous scientific papers to prestigious national and international journals and international conference volumes. Her research focuses on organizational change, human resources management, AI, and decision-making processes. Her research activity has been conducted alongside her teaching and individual academic development. She has also been part of the organizing committees for major national and international scientific communication sessions.



PRIYABRATA THATOI is currently a seasoned Data Scientist, with over eight years of extensive experience in the dynamic fields of machine learning, natural language processing, and predictive analytics. Throughout his career, he has honed his expertise in developing and deploying production-level code that empowers organizations to make informed, data-driven decisions. His work is characterized by his ability to transform complex data into actionable insights, driving innovation, and optimizing outcomes. With a proven track record of delivering impactful solutions, he consistently demonstrates a deep understanding

of both the technical and strategic aspects of data science, making him a valuable asset to any team or project he undertakes. His passion for leveraging data to solve real-world problems continues to inspire those around him, setting a standard of excellence in the ever-evolving landscape of data science.



SALEM A. ALYAMI (Member, IEEE) received the Ph.D. degree in bio-statistics from Monash University, Australia, in 2017. He is currently an Associate Professor with the School of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia. He has been contributing and leading several grants in statistics, and the Dean of Scientific Research with IMSIU, since 2020. His research interests include statistical inference, Bayesian statistics, sampling algorithms, statistical simulation, and the applications of statistics in biology and medicine.



AKM Azad is currently a Cancer Data Scientist and an AI Scientist in research. He is appointed as a Faculty Member with the Faculty of Science, Imam Mohammad Ibn Saud Islamic University. He has an excellent research career with a strong and highly focused track record on the development of advanced methodologies and computational resources for multi-omics (e.g., transcriptomics, proteomics, and genomics) data analysis and their relevant drug positioning. Earlier, he was a Cancer Data Scientist and Bioinformatician in multiple research institutes in Australia and South Korea, i.e., the Children's Medical Research Institute, Australia, the Kolling Institute of Medical Research, Australia, the University of New South Wales, Australia, and the Gwangju Institute of Science and Technology, South Korea. During his research career, he has gained experience in multi-omics data for identifying network-level biomarkers and analyzing their drug-targets in various cancers, e.g., Glioblastoma, Breast, and Pancreas. His research interests include AI-based biomedicine, data science, Bayesian methods, Bayesian networks, and drug repositioning.