LLM-Agent-Based Legal Compliance Evaluation System*

Nara Shin†, So-Hyun Park, Il-Gu Lee

Sungshin Women's University, Seoul, Korea
{narax0410, sohyun9625} @gmail.com, iglee@sungshin.ac.kr

Abstract

Recently, various artificial intelligence (AI) services have been applied across all industries, and the legal field is no exception. AI has searched for laws and precedents faster, easier, and more accurate than before. Nevertheless, many existing systems are limited to providing legal texts or precedents. This study proposes a real-time legal compliance diagnosis system based on a large language model (LLM)–driven multi-agent framework. The proposed system aims to (1) verify whether pre-submitted materials for regulatory sandbox applications meet relevant legal requirements, and (2) evaluate and monitor real-time compliance within the testing environment. Through this, the system goes beyond mere legal information retrieval to realize automated compliance support that integrates legal interpretation, application, and judgment.

1 Introduction

Testing of innovative services through regulatory sandboxes is increasing[1]. However, some projects are suspended due to unmet application requirements or legal and technical constraints during testing[2][3]. Although sandboxes relax certain regulations, legal boundaries that must be observed still exist. Extensive legal analysis is required to verify compliance, and recently, LLM-based analytical approaches have been introduced. Nevertheless, current AI systems are limited to searching laws and precedents, and actual compliance must be reviewed manually. Therefore, there is a need for an LLM-agent-based expert system that can automatically reflect the latest legal amendments and diagnose compliance with regulations in real time.

2 LLM-Agent-Based Legal Compliance Evaluation System

The proposed system operates in three main stages: data input, analysis, verification, and evaluation of legal compliance. First, users upload service-related materials—prepared according to applicable laws in areas such as consumer protection or vulnerability assessment—in file formats such as CSV, PDF, or MD. In the data analysis stage, the LLM Agent analyzes and verifies the uploaded data using pre-trained legal texts and service-related information. It autonomously updates relevant legal knowledge, such as revised statutes or case law through the Retrieval-Augmented Generation (RAG) technique. When external integration (e.g., data crawling or analytics) is required, the system employs the Model Context Protocol (MCP) to improve data processing efficiency. Finally, the system evaluates whether the analyzed data complies with applicable regulations. If non-compliance is detected, it filters out the portions requiring modification, refines them to meet legal standards, and reanalyzes the updated

* Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. P-75, December 16-18, 2025, Sapporo, Japan.
† Corresponding author

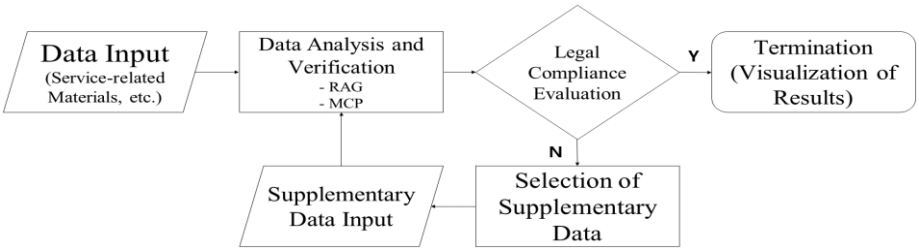

Fig. 1. Flowchart of the LLM-Based Real-Time Legal Compliance Evaluation System

dataset. The results and corrective actions are visualized and provided to the user to support efficient final decision-making.

Table 1 compares the traditional human-driven legal consulting approach, AI-based consulting systems [4][5], and the proposed system. Existing studies include legal search and LLM-based automation functions but have limited capabilities in real-time updates and compliance evaluation. In contrast, the proposed system distinguishes itself by employing an LLM-based multi-agent architecture that can update legal information in real time and automatically assess whether the input data complies with applicable regulations.

Category		Legal Search Function	LLM Utilization	Automated Pipeline	Real-Time Update	Legal Compliance Evaluation
Conventional	Legal Consulting	O	X	X	X	X
	AI-based Consulting Systems[4][5]	O	O	O	X	X
Proposed	LLM-Agent-Based System	O	O	O	O	O

Table 1. Comparison between Previous Studies and the Proposed System

3 Conclusion

In this study, how much each sensor contributed to the judgment of the drift detection model was visualized by calculating the Deep LIFT contribution. By analyzing this, it was confirmed that the drift sensor affected the abnormality judgment. This proved that the proposed method efficiently detects abnormal conditions from a model perspective, can identify which sensors were drifting at which point, and that dataset collection is an important aspect of artificial intelligence and an efficient method. Future studies plan to continue researching how to identify and correct drift intervals.

References

[1] Markellos, Raphael N., et al., Worldwide adoption of regulatory sandboxes: drivers, constraints and policies. *Constraints and Policies (March 6, 2024)*, 2024.

[2] Organisation for Economic Cooperation and Development (OECD), *Regulatory sandboxes in artificial intelligence, OECD Digital Economy Papers*, No. 356, Paris, France, 2023 [Online]. Available: https://doi.org/10.1787/8f80a0e6-en

[3] European Securities and Markets Authority (ESMA), European Banking Authority (EBA), and European Insurance and Occupational Pensions Authority (EIOPA), *Joint ESAs Report on Innovation Facilitators*, Paris, France, 2023. [Online]. Available: https://www.esma.europa.eu/press-news/esma-news/joint-esas-report-innovation-facilitators

[4] Kalaycioglu, Sean, et al., AI-Powered Legal Intelligence System Architecture: A Comprehensive Framework for Automated Legal Consultation and Analysis. *arXiv preprint arXiv:2508.17499*, 2025.

[5] Cui, Jiaxi, et al., Chatlaw: Open-source legal large language model with integrated external knowledge bases. *CoRR*, 2023.