LLM Driven Legal Text Analytics: A Case Study For Food Safety Violation Cases

Suyog Joshi[1], Soumyajit Basu[1], Lipika Dey[1], Partha Pratim Das[1]

[1]Ashoka University, Sonepat, India

Correspondence: suyog.joshi@alumni.ashoka.edu.in, soumyajit.basu@alumni.ashoka.edu.in
lipika.dey@ashoka.edu.in, ppd@ashoka.edu.in

Abstract

Despite comprehensive food safety regulations worldwide, violations continue to pose significant public health challenges. This paper presents an LLM-driven pipeline for analyzing legal texts to identify structural and procedural gaps in food safety enforcement. We develop an end-to-end system that leverages Large Language Models to extract structured entities from legal judgments, construct statute-and-provision-level knowledge graphs, and perform semantic clustering of cases. Applying our approach to 782 Indian food safety violation cases filed between 2022-2024, we uncover critical insights: 96% of cases were filed by individuals and organizations against state authorities, with 60% resulting in decisions favoring appellants. Through automated clustering and analysis, we identify major procedural lapses including unclear jurisdictional boundaries between enforcement agencies, insufficient evidence collection, and ambiguous penalty guidelines. Our findings reveal concrete weaknesses in current enforcement practices and demonstrate the practical value of LLMs for legal analysis at scale.

1 Introduction

In 2024, the World Health Organization (WHO) reported that unsafe food causes around 600 million cases of foodborne diseases, and 420,000 deaths annually. Despite the presence of comprehensive food regulatory acts across countries, food safety violations continue to pose a significant challenge to global public health. The complexity of food supply chains coupled with evolving legal frameworks, often results in inconsistent enforcement and delayed policy response. Though the exact statistics are not available for India, research articles report that among several other developing nations in South East Asia, food adulteration is widespread in India[1]. The Food Safety and Standards Act of India was passed in 2006 and thereafter, various provisions of the Act have to force through several notifications. Despite this, food safety violations remain a national concern. According to a recent article published by National Law Institute University, the Food Safety and Standards Authority of India (FSSAI), responsible for regulating and overseeing food safety in India, faces several challenges, including limitations of their regulatory purview, infrastructure deficiencies, limited scope of regulations, etc.[2]. In a study conducted by Shukla et al. (2014) in 2014, it was emphasized that when it comes to food safety assessment and implementation of quality, there exists a gap in infrastructure and risk-based approach in both implementation and enforcement. While this could possibly explain why food safety violation remains critically high, ours is an evidence-based approach to objectively unearth these gap areas through analysis of court cases. Our investigation revealed some interesting insights. While the number of court cases are not overwhelming, our study reveals that most of these were filed by those accused of violation against the authorities, and a large number of them are also won by the appellants. This prompted us to dive deeper into the case files to gain insights about what is happening and identify the possible loopholes, wherever they are.

Legal text analytics plays a critical role in identifying gaps, performing causal analysis, and thereby contributing towards strengthening regulatory oversight. By automatically analyzing statutes, compliance guidelines, inspection reports, and judicial rulings, legal text analytics systems can identify

[1]https://ncdc.mohfw.gov.in/wp-content/uploads/2024/04/Food-borne-Diseases-and-Food-Safety-in-India.pdf

[2]https://nliulawreview.nliu.ac.in/blog/fssais-regulatory-apathy-and-indias-marginal-consumers-a-case-for-decentralized-food-safety/

Proceedings of the 1st Workshop on NLP for Empowering Justice (JUST-NLP 2025), pages 41–52

emerging patterns of non-compliance, inconsistencies across regional laws, and gaps between regulations and enforcement practices. This work was initiated to obtain possible insights about the food safety violation landscape, the type of cases reported, the food items or malpractices frequently associated, and most importantly the possible reasons for the practice not reducing, despite regular actions by FSSAI. The intent was to analyze the case proceedings and judgments in the context of the Food Safety and Standards Act of India, 2006, to identify possible causal factors. In this paper we propose an evidence-based approach that uses Natural Language Processing techniques to identify the structural and procedural factors that could be influencing the case volumes in the judicial system.

One of the key bottlenecks faced while designing legal text analytics systems earlier was the lack of annotated corpora to train the systems. This has been substantially eased by the Large Language Models (LLMs), which have already been exposed to fair amounts of legal corpora from across the world. LLMs can play a crucial role in legal text analytics by summarizing regulatory provisions and case documents, extracting obligations, penalties and other relevant key entities, and also in answering questions about how the judicial reasoning progressed. Together, these can provide deeper insight into the legal landscape associated with food safety in India. The insights can enable regulators, policymakers, and researchers to gain a data-driven understanding of legal landscapes, facilitating proactive interventions and harmonization of global food safety standards. Ultimately, the integration of legal text analytics into food governance frameworks can enhance transparency, improve compliance monitoring, and contribute to safer and more accountable food systems around the world.

This paper presents how Large Language Models (LLMs) and knowledge graphs can be used for legal text analytics. The research develops an end-to-end pipeline that leverages LLMs to extract structured entities and factual context from legal judgments, constructs a statute-and-provision-level knowledge graph, and aligns India's Food Safety and Standards Act, 2006 with equivalent laws in Finland and the United Kingdom using a hybrid semantic similarity and LLM-based matching process.

2 Review of Related Work

Natural language processing of legal texts poses unique challenges due to their lengths, denseness and use of specialized vocabulary. However their analysis is important to understand the effectiveness of the legal system. While manual processing is time-consuming and prone to error, automated analysis is difficult due to the complexity of syntax, archaic jargon, and strict semantics. The text processing tasks traditionally included summarization, named entity recognition and structured information extraction. Ariai et al. (2024) present a comprehensive view of language processing tasks for legal text. The integration of Large Language Models (LLMs) in legal technology is rapidly transforming the landscape of legal research and document analysis, and several new applications like predicting missing citations, legal analytics are also gaining popularity with these models. The rapidly changing landscape of legal technologies centered around the use of LLMs are presented in many articles like Mayer (2023), Siino et al. (2025), Padiu et al. (2024), Ződi (2024). Named Entity Recognition (NER) in legal texts are designed to identify entities like statute names, case parties, courts, etc. Early approaches treated NER as a token classification task (Skylaki et al., 2020) but use of abbreviations, synonyms, acronyms make this task heavily error-prone. A recent survey evaluating approaches ranging from traditional rule-based systems to LLMs reports that `GPT-4` achieves superior performance on legal NER benchmarks, outperforming smaller models (Deußer et al., 2024). This suggests that LLMs can better use context to resolve ambiguities e.g. linking "Section 420 of FSS" to the correct law. NER based approaches have been also used to extract structured facts like case facts, charges, outcomes, etc. from judgments. Recent work like LegalLens reports the use of `GPT-4` to label violation entities in unstructured legal text (Hagag et al., 2024). In their work Deußer et al. (2024), also report that `GPT-4` outperformed smaller language models on legal NER and text classification tasks.

Use of LLMs have enhanced the quality of structured information extraction from legal texts, in turn increasing their use in legal text analytics. In a recent publication, Pereira et al. (2025) have shown the utility of using GPT-4o to extract factors for Brazilian consumer law judgments, without any further fine-tuning. LLMs have also been reported for immigration policy analysis by Brown (2025).

Li et al. (2025) propose LegalAgentBench, a comprehensive benchmark specifically designed to evaluate LLM Agents in the Chinese legal domain.

The quality of summarizing long legal documents like contract agreements also improved phenomenally with the use of LLMs. The work reported in Davenport (2025) proposed the use of LLMs for hierarchical segmentation of large documents in combination with chain-of-thought prompting and multi-stage summarization techniques. This article reports that OpenAI's API can not only summarize and analyze long contracts quite well, but also capture critical obligations and clauses with high accuracy and efficiency. Similar results were reported in Litaina et al. (2024), who presents results from an experiment in which they use GPT-3.5 and Gemini to analyze type and subject of contracts to obtain insights about involved named entities and their relationship(s). Handwritten contracts were first digitized into plain text format using an AI-powered tool called Transkribus. Their experiments demonstrated that the LLM-generated responses outperformed humans in precision, but not in recall. The performance of LLMs can be further improved by the use of legal Knowledge Graphs (KG). A legal knowledge graph typically represents cases, laws, and concepts as nodes, with edges for relationships like citations or involvement. In a different approach, the work done by Dhani et al. (2024) presents an Indian legal KG built with regulatory documents and case documents. The nodes include cases, judicial orders, statutes, parties involves, while edges capture citations, applicable statutes etc. A Graph Neural Network, is trained on the knowledge graph to predict similar cases and thereafter suggest missing citations. The KG-based approach is reported as yielding higher precision while retrieving similar cases. Using LLMs for legal reasoning is increasingly gaining traction. A framework for unified legal reasoning that combines rule-based, abductive, and case-based approaches, and then investigate possible methods for their integration with LLMs is proposed in Nguyen et al. (2025).

3 Knowledge Graph Design for Regulation Act and Case Documents

A knowledge graph was designed to capture some of the key information that is relevant for insightful analysis of cases, not just for the present work, but also to be used for other kinds of comparative analytics of legal documents. The design was based on concepts and relationships presented in Leone et al. (2019) enhanced by the main features of the provisions and statutes as followed by Indian legal framework. The statutes represent the broader issues being discussed. For example, a tax-related case may cite the Income Tax Act, while the provisions represent the specific clauses and subsections that provide the precise context for the case's arguments and judicial reasoning. Another portion of the knowledge graph contains details about individual cases. These include case-specific details like the judge overseeing the case, names of petitioners and respondents, the acts, statutes, section numbers cited in the case, the names of locations and organizations mentioned in the document, the final judgment in the case.

Figure 1 presents the design of the knowledge graph that stores country-specific regulatory information created from regulatory documents. This graph contains as nodes the names of different acts, statutes, provisions and their descriptions. These are connected to each other using the part-of relationship. Initially, a graphRAG based approach was also tried. However, with very little control over the knowledge graph that was created, it was difficult to perform analytical queries of the kinds that we wanted to.

On the right hand side of Figure 1, the knowledge graph design for individual cases is presented. Each case gets connected to those nodes of the law knowledge graph, that are cited in it. The prompt used for the extraction of different entities, their resolution and creating entries for the knowledge graph is given in A.2.

Figure 2 presents the full knowledge graph created from the Food Safety Regulatory document published in 2006. Connecting the cases to the statutes and provisions in the knowledge graph allows for meaningful insights to be extracted about the nature of the cases present in a corpus. The graph can be queried for quantitative insights. This includes information like the most commonly cited statutes across cases, clustering of cases based on citation similarity, and finding the cases most similar to a chosen case. These queries can provide information about patterns across the corpus, such as understanding the frequency of specific legal interpretations, or identifying the most common reasoning provided in specific types of cases.

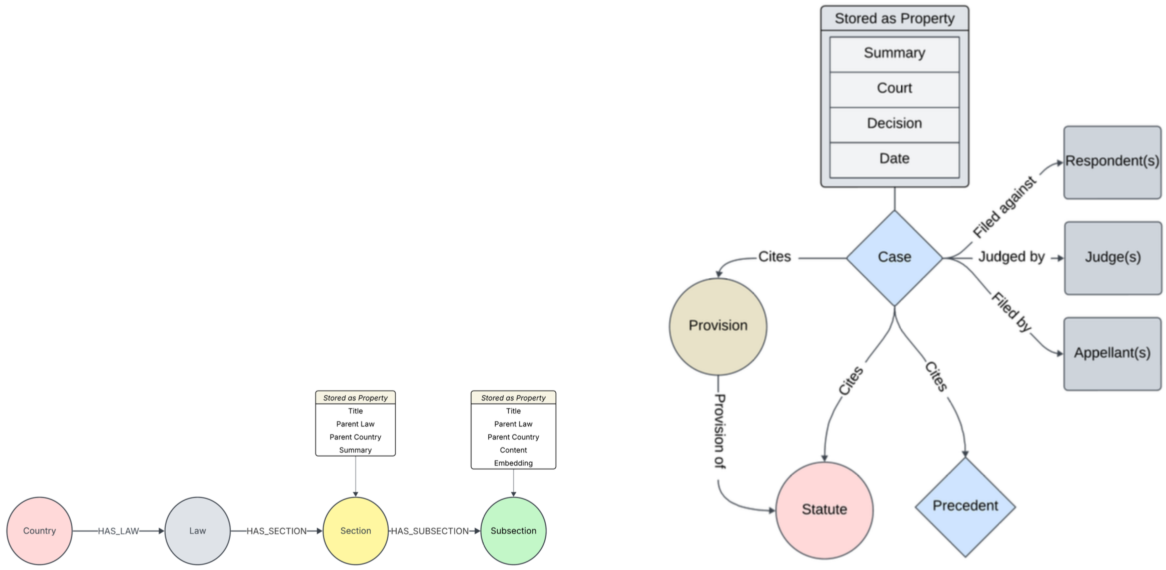

Figure 1: (left) Knowledge Graph Model for storing Regulatory information about statutes and sections (right) Knowledge graph for each case document

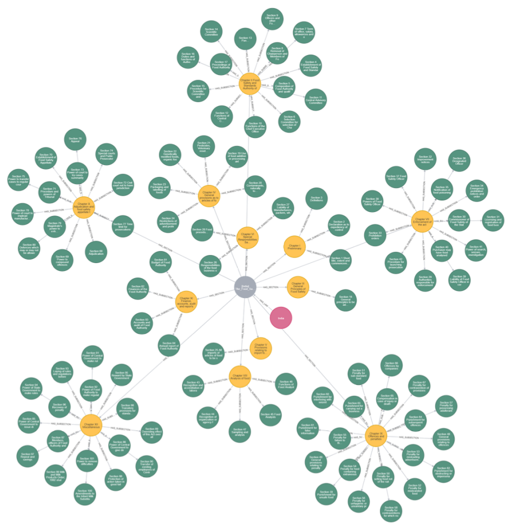

Figure 2: Knowledge Graph curated from Food Safety and Standards Act 2006, India

4 LLM Driven Pipeline to Obtain Insights from Case files

Figure 4 presents the pipeline that we have deployed for LLM-driven case document analysis. While the core activity is to extract different kinds of information that can help in analytics-driven insight generation, the content analytics pipeline consists of two different threads. Case documents were downloaded from a website called Indiakanoon.org using a custom-designed crawler. The pipeline consists of two independent threads, each of which processes each case document. The analytical activities utilizes the outputs of both. Thread 1 analyzes each document and creates a structured summary of the court proceedings. The prompt for creating the summary is given in A.3

Thread 2 parses each case to extract different types of entities that are required to populate the knowledge graph. The end result of the pipeline is a structured representation for the documents, their summaries and all the case-specific details that are extracted from the document. This consolidated list is now ready for further analysis to obtain insights.

5 Clustering Case Summaries, Decisions and Reasons Given for the Decisions

To discover the patterns across the food safety violation cases, we explored a semantic similarity based clustering methodology using the structured case summaries obtained earlier. Each case summary contains the case overview, key facts from the case, the decision and the judge's reasoning. Contextual embeddings of these summaries are created using different mechanisms. With the summaries now represented as vectors, we use the k-means algorithm to cluster the cases into meaningful groups. The optimal number of clusters is determined using the silhouette score and Davies-Bouldin (DB) index together. Specifically, we optimize $k^* = \arg\max_{k\in[2,20]} \frac{S(k)}{DB(k)}$, where $S(k)$ and $DB(k)$ are the silhouette score and the DB index,

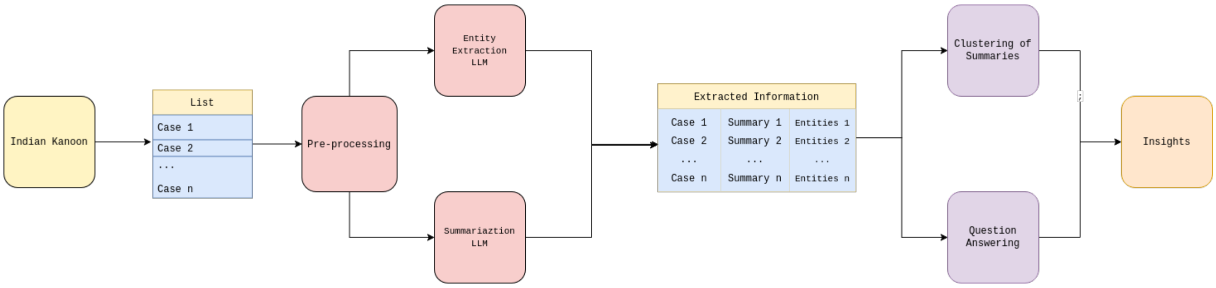

Figure 3: LLM-driven case document processing Pipeline - Information extraction from individual documents followed by analytics over the repository

respectively, for k clusters. Finally, each cluster is passed through an LLM to generate a descriptive labels for the cluster. In this paper, we report results for three different embedding model and LLM combinations. The first set of embeddings were created using OpenAI's `text-embedding-3-large` model, and the corresponding cluster labels were generated using `GPT-4o`. The second set used Google's `gemini-embedding-001` for embedding generation and `Gemini-2.5-Flash` for generating the cluster labels. A third set of results were obtained using Alibaba's `qwen3-embedding-8b` for embedding generation and `Qwen3-235B-A22B-Instruct-2507` for cluster labeling. The intent was to study the susceptibility of the analytical results to different embedding generation process. For a robust analysis, the case summaries should cluster more or less identically across all three embedding spaces, and also be assigned semantically similar labels. Results from Section 6.4 show that this indeed is the case.

6 Analysis of Food Safety Violation cases filed between 2022 - 2024

6.1 Data Collection

Court proceedings related to food safety were collected from indiankanoon.org using a webscraper built using Selenium and BeautifulSoup. Prior consent to use judgments available on the website was received from the IndianKanoon team. To ensure all cases mentioning "Food Safety" were included, a search term with a week-long window was created, which ensured that cases were downloaded one week at a time. This is because IndianKanoon limits search results to 400 cases, so limiting to one week ensures all cases from the period are downloaded. Judgements passed between January 1, 2022 and December 31, 2024 were downloaded. A total of 7233 cases were collected, out of which duplicates were dropped along with files containing incorrect formatting and files exceeding the `GPT-4o-mini` context length of 128,000 tokens. The remaining 7040 files were selected for processing after minimal pre-processing. Links common to all documents, such as links pointing to IndianKanoon or judis.nic.in were removed. The remaining texts were passed downstream as-is for extracting the information about petitioners, defendants, court and the date of case filed. First level analysis revealed that a majority of them were related to tobacco products. Though tobacco itself is not considered as a food product under the Food Safety and Standards Act (FSS Act), its use is banned in edible items. This might have resulted in a high number of cases fetched during crawling and hence were dropped. A total of 782 unique cases finally remained which were used for downstream analysis.

6.2 Case Document Processing and Summarization

Each case document was passed through the summarization and information extraction pipelines implemented using `GPT-3.5`. The pipeline extracts structured information including petitioners, respondents, court details, case overview, key facts, legal issues and arguments, court's reasoning, and the final decision. A sample output demonstrating the extracted structure is provided in Appendix A.1.

To assess the faithfulness of the LLM-generated summaries to the original content, we computed the BERT and ROUGE-L scores between the generated summaries and the original case documents. An average BERT score of 0.624 represents a fair amount of semantic similarity between the summaries and their parent documents. Average ROUGE-L score of 0.275 represents structural similarity, that is represented by the longest common subsequences present in both the document and its summary. This

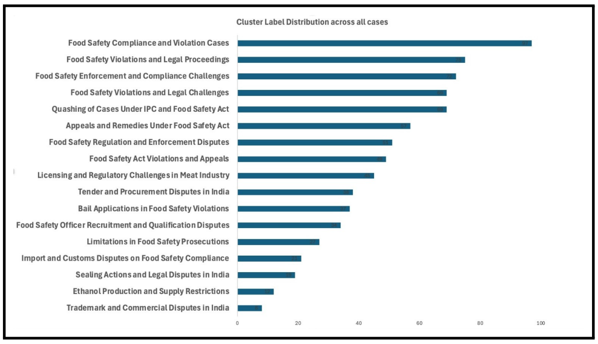

Figure 4: Distribution of cluster labels across all cases

part of the experiment could not be replicated due to resource constraints.

6.3 Results of Clustering of Food-safety related Cases Decided Between 2022 - 2024

The experiments were repeated many times. On an average, GPT obtained 17 clusters, Gemini based analysis yielded 18 and Qwen found the optimal number to be 16. Figure 5 presents the distribution of the cases across cluster labels for all the three alternatives, generated using t-SNE, which essentially produces two-dimensional plots for high-dimensional embeddings. The colours are randomly assigned by the visualizer for each plot, and has no other significance. In other words, blue dots across the plots are not necessarily associated to the same underlying documents across the different embedding spaces. Indeed, the plots do show that the case summaries cluster quite similarly across different embedding spaces, and the optimal numbers are also pretty close.

6.4 Comparing Cluster Labels for Robustness Analysis

The thematic alignment between the cluster labels generated by the three models was obtained using a semantic consensus metric based on Sentence-BERT (SBERT) embeddings for the cluster labels generated by the three LLMs. These embeddings were generated using `all-MiniLM-L6-v2` model, and their pairwise cosine similarities were computed to determine semantic equivalence. A similarity threshold of $\tau = 0.55$ was established to account for terminological variations, such as “Ethanol Production” versus “Ethephon and Ethanol Regulatory Disputes”, while maintaining semantic precision. The analysis yielded a robust average consensus score of 90.3% across all model pairs. The highest alignment was observed between `Gemini-2.5-Flash` and `Qwen-3` (94.4%), followed by `GPT-4o` and `Gemini-2.5-Flash` (88.2%), and `GPT-4o` and `Qwen-3` (88.2%). Detailed pairwise similarity scores for all cluster label comparisons are provided in Appendix A.5.

A granular examination of the divergent labels (similarity < 0.55) revealed that the remaining discrepancies were largely structural rather than substantive. For instance, `GPT-4o` tended to isolate specific administrative actions like “Sealing Actions”, while other models grouped these with broader regulatory enforcement themes. Figure 6 presents the agreement between the labels generated by different methods. The strong cross-model agreement on labels validates the robustness of our clustering approach and confirms that the identified legal themes are model-independent.

On manual inspection, GPT-generated labels were found to be most comprehendible as well as at the right granular levels. For example, the most frequent label *Quashing Food Safety and Allied Criminal Proceedings* generated by `GPT-4o`, links to three different labels generated by Gemini, namely *Food Safety Act: Limitation and Quashing Proceedings*, *Food Safety Procedural Violations and Quashing* and *Food Safety Act Violations and Quashment*. Case-wise label assignments were also reviewed manually for robustness check. Case document wise, it is found that 90% of the cases that are assigned to the cluster labeled *Quashing Food Safety and Allied Criminal Proceedings* by `GPT-4o` are assigned to the cluster labeled *Food Safety Act: Limitation and Quashing Proceedings* by Gemini. The next highest category of cases belong to the cluster labeled *Food Safety Compliance and Violation Cases*. 65% of these cases belong to the cluster labeled *Food Safety Act Procedural Violations* and the remaining 35% cases belonged to the cluster labeled *Food Safety Procedural Violations and Quashing*. After reviewing all the labels, for the final insight extraction, we decided to use the `GPT-4o` labels for clarity and distinctiveness. Figure 4 presents the distribution of the cluster labels generated by `GPT-4o`.

6.5 Insights Extracted

A detailed analysis of the 782 cases is now presented. Analysis of petitioners and respondents shows that 96% of these cases were filed by individuals and organizations against a representative

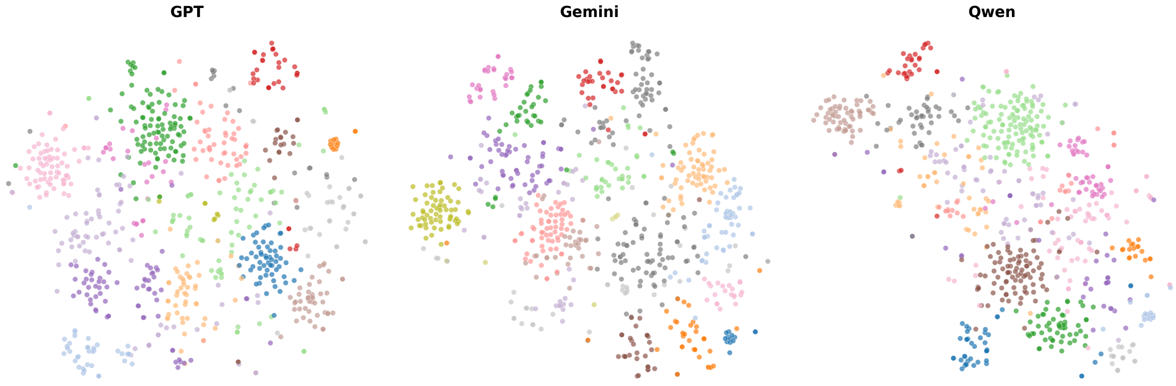

Figure 5: t-SNE Visualization of clusters generated by different embedding methods; distinct clusters are obtained for all methods showing the separability of the cases irrespective of the embedding space

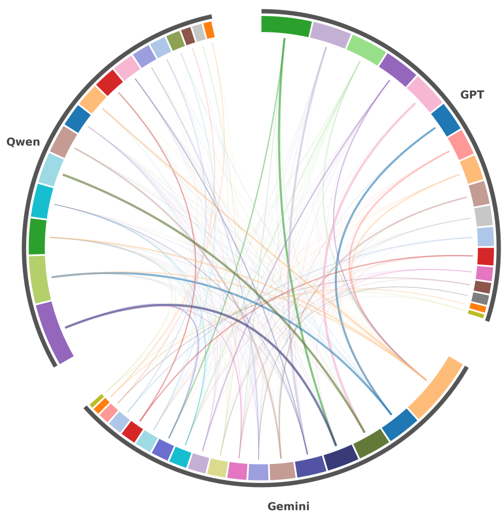

Figure 6: Agreement between the labels generated by different embedding methods. Gemini-generated labels are good proximities to the other two sets. Granularity of GPT generated labels appear to be best as the most frequent labels are more uniformly distributed. Each corresponds to one big and one small cluster of Gemini.

of the state authority. These cases were mostly filed challenging an action or a decision taken by a representative of a State agency. In a few cases, these were petition against government announcements. Based on the extracted information about in whose favour the judge's decision go, Figure 7 shows that 60% of the cases filed against state authorities were won by the appellants or petitioners. Assuming that the state authorities would have acted against an individual or organization only if a food safety violation act was detected, this statistics by itself is quite intriguing. It reveals that most of the appeals against the authority's actions were given a judgment which went in favour of the appellant or petitioner, and not the state authority. Figure 8 shows the distribution of the cluster labels obtained using `GPT-4o` for those cases which were won by individuals or organizations against state agency representatives. We opted for `GPT-4o` labels for the same reason as cited earlier.

To understand the situation better, we present the cluster label distributions for the above cases separately in . Further analysis of the top 5 labels, other than those for bail applications are given below:

- Quashing of Cases Under IPC and Food Safety Act: These were cases where the court quashed the FIR registered against the petitioners. For many cases, the court observed that the FSS Act allows only Food Safety Inspectors to initiate prosecution, and not Police or other authorities. Some were quashed since the action was taken after stipulated time-period from FSS violations were observed. Similarly, if a case was filed under provisions of both IPC and FSS, the court observed that the dual role of the complainant and investigating officer compromises on fair trial rights of the appellant.

- Food Safety Compliance and Violation Cases: This category majorly includes all cases in which the judge's decision went in favour of the appellants citing procedural lapses on part of the state authorities while conducting the necessary tests for establishing food safety violations

- Food Safety Enforcement and Compliance Challenges: In this category, the court most often found the respondents' arguments unacceptable. the details vary from case to case. For example, in a case titled "R Shanmughasundaram vs The Food Safety Officer on 21 September 2023", that involved a fine imposed for misbranding oil products, the petitioner challenged the fine amount of Rs. 3,00,000, which they found to be excessive. The state authorities argued that a sample of groundnut oil was actually found to be Palm Oil and there is a significant price difference between the two. Though there was merit in the case, the court noted that the District Revenue Officer did not fully consider the factors outlined in Section 49 of the Food Safety and Standards Act, 2006, when determining the fine. The petitioner was let off with an undertaking to prevent future misbranding, a smaller amount of fine and an order for the state authorities to vacate the entire premises without any qualifications.

- Food Safety Violations and Legal Proceedings: These were mostly cases which dragged on for a long time. In some cases, the petitioners pleaded not guilty to the crime as they were only resellers or employees. In some cases, though the crimes were acknowledged, since the petitioners had spent substantial time in imprisonment, the court decided to replace the sentence of imprisonment with penalties.

- Food Safety Regulation and Enforcement Disputes: Under this category, the court noted that the allegations of violations reported were not substantiated by sufficient evidence or test results that were acceptable.

Detailed analysis of all the clusters revealed that there were major procedural issues due to which many food safety violation cases were not being penalized sufficiently. It includes lack of knowledge or actual lack of concrete boundaries between the responsibilities on the part of the Food Safety officers and other law enforcement agencies like the police. Unclear rules about the penalties to be applied also allows petitioners to possibly get away with their crimes. In many cases, resellers or employees were penalized, which doesn't affect the root cause of safety violation, and could be also one of the reasons for the violations to continue.

The analysis clearly establishes the need for safeguarding the interests of public health through re-

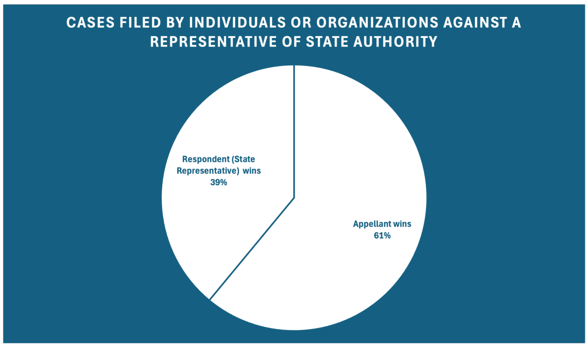

Figure 7: Distribution of judge's decision for cases filed by individuals or organizations against state authorities - most cases are lost by the State authorities

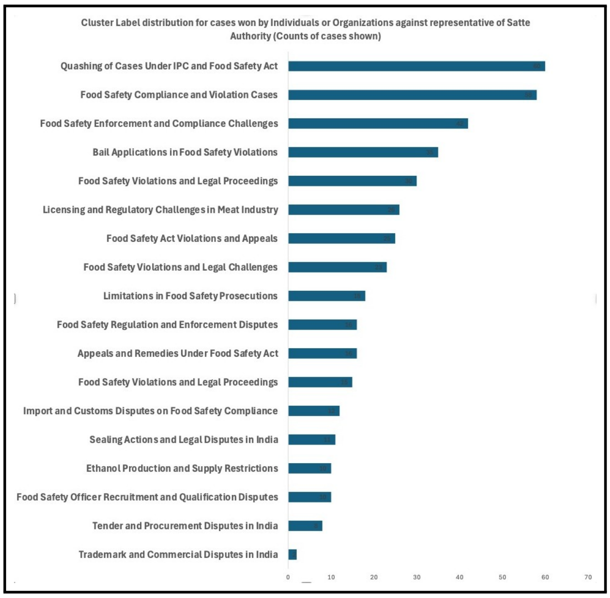

Figure 8: Distribution of cluster labels for cases won by individuals or organizations against state authorities

moval of these loopholes in the regulation and its implementation along with education and awareness among all citizens.

7 Conclusions and Future Work

In this work, we have demonstrated the potential of Large Language Models (LLMs) in establishing an evidential approach to understand the landscape of food safety violations by analyzing the food safety–related legal cases. This work has shown that by leveraging the advanced capabilities of LLMs in analyzing legal text, both information extraction and contextual reasoning can play significant roles in identifying the possible causes of recurrent violations. The insights that are revealed include the large number of cases that end up being quashed due to procedural errors and also an equally large number of them showing up the lacunae in the regulatory framework itself as well as enforcement challenged. The case study finds evidences in support of known issues that exist between the intentions of the regulation and implementation challenges. Lack of awareness among the authorities of law enforcement also surface as a key issue.

The research also initiated knowledge modeling for the regulations themselves. In future, we intend to continue our work towards comparative analysis of regulations from different countries, their implementations and the outcomes observed. The work will continue to focus on developing easily deployable explainable and domain-tuned LLMs that can integrate legal ontologies and cross-jurisdictional data to ensure trust and accountability. This area can contribute significantly towards strengthening of food safety systems across the globe.

Limitations

The proposed pipeline has been applied on food safety-related cases filed in Indian courts over three years from 2022 to 2024. The entire pipeline from initial document processing to clustering could not be repeated with multiple LLMS due to resource constraints. Only the clustering part was repeated for three different systems. Our future work would focus on employing the entire pipeline for larger set of case documents to obtain insights at multiple levels, including those related to food items and their propensity for safety violations. This would make the work more complete.

Acknowledgments

We are thankful to the Mphasis AI & Applied Tech Lab at Ashoka - a collaboration between Ashoka University and Mphasis Limited - for their support.

References

Farid Ariai, Joel Mackenzie, and Gianluca Demartini. 2024. Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges. *ACM Computing Surveys*.

Crystal Brown. 2025. *Leveraging generative AI and system dynamics for enhanced immigration policy analysis*. Ph.D. thesis, Worcester Polytechnic Institute.

Mark J Davenport. 2025. Enhancing legal document analysis with large language models: A structured approach to accuracy, context preservation, and risk mitigation. *Open Journal of Modern Linguistics*, 15(2):232–280.

Tobias Deußer, Cong Zhao, Lorenz Sparrenberg, Daniel Uedelhoven, Armin Berger, Maren Pielka, Lars Hillebrand, Christian Bauckhage, and Rafet Sifa. 2024. A comparative study of large language models for named entity recognition in the legal domain. In *2024 IEEE International Conference on Big Data (BigData)*, pages 4737–4742.

Jaspreet Singh Dhani, Ruchika Bhatt, Balaji Ganesan, Parikshet Sirohi, and Vasudha Bhatnagar. 2024. Similar cases recommendation using legal knowledge graphs. *Preprint*, arXiv:2107.04771.

Ben Hagag, Liav Harpaz, Gil Semo, Dor Bernsohn, Rohit Saha, Pashootan Vaezipoor, Kyryl Truskovskyi, and Gerasimos Spanakis. 2024. LegalLens shared task 2024: Legal violation identification in unstructured text. *Preprint*, arXiv:2410.12064.

Valentina Leone, Luigi Di Caro, and Serena Villata. 2019. Taking stock of legal ontologies: a feature-based comparative analysis. *Artificial Intelligence and Law*, 28(2):207–235.

Haitao Li, Junjie Chen, Jingli Yang, Qingyao Ai, Wei Jia, Youfeng Liu, Kai Lin, Yueyue Wu, Guozhi Yuan, Yiran Hu, Wuyue Wang, Yiqun Liu, and Minlie Huang. 2025. LegalAgentBench: Evaluating LLM agents in legal domain. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2322–2344, Vienna, Austria. Association for Computational Linguistics.

Tania Litaina, Andreas Soularidis, Georgios Bouchouras, Konstantinos Kotis, and Evangelia Kavakli. 2024. Towards LLM-based semantic analysis of historical legal documents. In *SemDH2024: First International Workshop of Semantic Digital Humanities, co-located with ESWC2024.*

T Mayer. 2023. AI and LLMs in legal technology: Revolutionizing research and document analysis. *Advances in Computer Sciences*, 6(1).

Ha Thanh Nguyen, Wachara Fungwacharakorn, May Myo Zin, Randy Goebel, Francesca Toni, Kostas Stathis, and Ken Satoh. 2025. LLMs for legal reasoning: A unified framework and future perspectives. *Computer Law & Security Review*, 58:106165.

Bogdan Padiu, Radu Iacob, Traian Rebedea, and Mihai Dascalu. 2024. To what extent have LLMs reshaped the legal domain so far? a scoping literature review. *Information*, 15(11):662.

Lucas de Castro Rodrigues Pereira, Maykon Marcos Junior, Guilherme de Brito Santos, Isabela Cristina Sabo, Thiago Raulino Dal Pont, Andressa Silveira Viana Maurmann, Luísa Bollmann, Maite Fortes Vieira, João Gabriel Mohr, Cristian Alexandre Alchini, and 1 others. 2025. Using

GPT-4o as a factor extractor for Brazilian consumer law judgments. *Artificial Intelligence and Law*, pages 1–29.

S. Shukla, R. Shankar, and S. P. Singh. 2014. Food safety regulatory model in India. *Food Control*, 37:401–413.

Marco Siino, Mariana Falco, Daniele Croce, and Paolo Rosso. 2025. Exploring LLMs applications in law: A literature review on current legal NLP approaches. *IEEE Access*.

Stavroula Skylaki, Ali Oskooei, Omar Bari, Nadja Herger, and Zac Kriegman. 2020. Named entity recognition in the legal domain using a pointer generator network. *Preprint*, arXiv:2012.09936.

Zsolt Ződi. 2024. Legal technology in the service of access to justice. *Hungarian Journal of Legal Studies*, 64(3):323–335.

A Appendix

A.1 Sample Case Output

Below is a sample output for a case titled *M/S Itc Limited vs Food Safety Inspector on 12 October, 2023*, demonstrating the structured extraction format:

<Begin Output Summary>
Petitioner: M/S Itc Limited

Respondent: FOOD SAFETY INSPECTOR, DEPARTMENT OF THE FOOD SAFETY AND STANDARDS KALABURAGI

Court: Karnataka High Court

Case overview: The case involves a writ petition filed by M/S ITC Limited against the Food Safety Inspector and the Designated Officer of the Department of Food Safety and Standards, challenging an order dated December 29, 2015, which imposed a fine for misbranding of their product, Sunfeast Yippee Noodles. The petition was heard in the High Court of Karnataka, Kalaburagi Bench.

Key facts:

- The petitioner, M/S ITC Limited, manufactures instant noodles and was accused of misbranding due to the label stating 'No Added MSG'.
- A sample was collected on June 8, 2015, and tested, with the Chief Food Analyst reporting it as misbranded.
- The petitioner claimed they were not given proper notice or the right to appeal against the Food Analyst's report.
- The Designated Officer imposed a fine of Rs. 10,000 based on the Food Analyst's report.

Legal issues and arguments:

- The petitioner argued that the product complied with all standards and that the misbranding claim was unfounded, as there was no evidence that MSG was added during manufacturing.
- They contended that the prosecution was initiated without proper inspection of the manufacturing premises and without affording them the right to appeal as per Section 46(4) of the Food Safety and Standards Act, 2006.
- The respondents maintained that the product was misbranded due to the label claim and that the prosecution was justified.

Court's reasoning:

- The court noted that the Chief Food Analyst's report indicated compliance with standards but also stated that the product was misbranded due to the label claim.
- It highlighted that there was no analytical method to determine if MSG was added or naturally present, and that prosecution should not occur without ascertaining the addition of MSG during manufacturing.
- The court found that the Designated Officer did not inspect the manufacturing unit, which was necessary to determine the validity of the misbranding claim.
- The court emphasized that the petitioner was not given the opportunity to appeal, violating principles of natural justice.

Decision or judgment: The court allowed the writ petition, quashing the order passed by the Additional District Magistrate cum Adjudicating Authority in C.C.No.316/203/54/2015-16 dated December 29, 2015, against M/S ITC Limited.

Type: Case won by Appellant

<End Output Summary>

A.2 Entity Extraction Prompt

You will be provided with a legal judgment of a food safety related case. Your goal is to extract key information following the schema provided. Please ensure that the extracted information is accurate and complete. Any references to the Food Safety and Standards Act should be extracted as "Section x of the Food Safety and Standards Act, 2006" or "Food Safety and Standards Act, 2006". Always use full forms of abbreviations, e.g. "Supreme Court" instead of "SC", or "Indian Penal Code" instead of "IPC". Names of courts should be standardized and follow correct capitalization, e.g. "Supreme Court of India", "High Court of Delhi", NOT "Supreme Court", "Delhi High Court" or "DELHI HIGH COURT".

Here is a description of the parameters to be extracted:

- court: name of the court that issued the judgment.
- petitioners: array of strings containing the names of ALL appellant(s) in the case.

- respondents: array of strings containing the names of ALL respondent(s) in the case.
- judges: array of strings containing the names of the judge(s) in the case.
- date: date of the judgment, as a string in the format "DD-MM-YYYY".
- org: array of strings containing the names of all organizations, companies, or government entities mentioned in the case, if any.
- gpe: array of strings containing the names of all geographical locations mentioned in the case, if any.
- provisions: array of strings containing the provisions of ALL statutes cited in the judgment. Provide these in the format "Section x of y". In case of references to multiple sections of the same statute, list them all separately.
- statutes: array of strings containing the names of ALL acts or laws cited in the judgment.
- precedents: array of strings containing the names of ALL precedents cited in the judgement.
- key facts: key facts about the case. This should be very concise, and include the background of the case, and the main arguments made by both parties. If no information is provided, leave this field empty.
- type of case: the type of case, e.g. bail application, civil appeal, criminal appeal, public interest litigation, etc.
- decision: the decision of the court in the case, if provided. If there is a verdict, respond with 'in favour of appellant' or 'in favour of respondent'. If not, leave this field empty.

A.3 Summary Generation Prompt

You will be provided with a legal judgment of a food safety-related case. Your goal is to provide a detailed summary of the judgment. Only include information explicitly stated in the document. Do not infer, interpret, or add new information. Use concise language while preserving all critical legal details, such as statute names, provisions, case outcomes, and involved parties. Organize the summary logically, grouping related points. For example:

- Case overview
- Key facts
- Legal issues and arguments
- Court's reasoning
- Decision or judgment

Return plain text, do not include any markdown or HTML formatting.

A.4 Cluster Labeling Prompt

<Task>
You are an expert document analyst tasked with analyzing clusters of similar documents and generating concise, descriptive labels that capture the common theme or topic of the cluster.
</Task>

<Documents>
{documents}
</Documents>

<Instructions>

1. Read through all the provided documents carefully.
2. Identify the common themes, topics, or patterns across the documents.
3. Focus on:
   - Legal subject matter (if applicable)
   - Key parties or entities involved
   - Types of cases or proceedings
   - Common legal issues or questions
   - Geographic or jurisdictional patterns
   - Temporal patterns or time periods
4. Generate a SHORT, descriptive label (3-10 words maximum) that captures the essence of the cluster.
5. The label should be:
   - Specific enough to distinguish this cluster from others
   - General enough to encompass all documents in the cluster
   - Clear and understandable to someone unfamiliar with the documents
   - Professional and concise

<Instructions>

<OutputFormat>
Return ONLY the label text without any additional explanation, prefixes, or formatting. Do NOT include phrases like "Label:", "Cluster:", or "Theme:". Just provide the descriptive label itself. Examples of good labels are "Food Safety Violations and Regulatory Actions", "Municipal Tax and Assessment Challenges", "Employment Termination and Labor Rights", "Environmental Compliance and Pollution Cases"
<OutputFormat>

A.5 Cluster Label Semantic Similarity Analysis

Tables 1, 2, and 3 provide detailed pairwise semantic similarity scores between cluster labels generated by GPT-4o, Gemini-2.5-Flash, and Qwen-3, as described in Section 6.4. Labels with SBERT similarity scores above our established threshold of 0.55 indicate semantic consensus. The high proportion of aligned labels across all model pairs validates the robustness of our clustering approach.

GPT	Gemini	SBERT Score
Appeals and Remedies Under Food Safety Act	Food Safety Act Appeals and Remedies	0.99
Bail Applications in Food Safety Violations	Food Safety Violations and Bail Applications	0.99
Food Safety Officer Recruitment Disputes	Food Safety Officer Recruitment Disputes	0.98
Food Safety Act Violations and Appeals	Food Safety and Standards Act Violations	0.93
Food Safety Compliance and Violation Cases	Food Safety and Standards Act Violations	0.91
Import and Customs Disputes	Food Safety and Import Regulation Disputes	0.90
Food Safety Violations and Legal Challenges	Food Safety and Standards Act Violations	0.90
Food Safety Violations and Legal Proceedings	Food Safety Act Cases and Quashing Proceedings	0.89
Food Safety Enforcement Challenges	Food Safety and Standards Act Violations	0.86
Food Safety Regulation Disputes	Food Safety and Import Regulation Disputes	0.86
Limitations in Food Safety Prosecutions	Food Safety Act: Limitation and Quashing	0.85
Quashing of Cases Under IPC and FSS Act	Food Safety Act Cases and Quashing Proceedings	0.83
Ethanol Production and Supply Restrictions	Ethanol Production Regulation and Sugar Supply	0.81
Trademark and Commercial Disputes	Trademark Infringement and Product Disputes	0.72
Licensing and Regulatory Challenges in Meat	Food Business Licensing Under FSS Act	0.60
Divergent / Low Consensus Matches (Score < 0.55)		
Tender and Procurement Disputes	Food Procurement and Safety Disputes	0.42
Sealing Actions and Legal Disputes	Food Safety Act Cases and Quashing Proceedings	0.37

Table 1: Semantic alignment between GPT and Gemini cluster labels.

Gemini	Qwen	SBERT Score
Food Safety and Standards Act Violations	Food Safety and Standards Act Violations	0.97
Food Safety Violations and Bail Applications	Food Safety Violations and Anticipatory Bail Cases	0.94
Food Safety Act Appeals and Remedies	Food Safety Act Enforcement and Appeals	0.93
Food Safety Act Violations and Quashment	FSS Act Violations Quashed by Precedent	0.88
Food Safety Act: Limitation and Quashing	Food Safety Act Enforcement and Appeals	0.87
Food Safety Procedural Violations and Quashing	Food Safety Prosecution Challenges	0.87
Food Safety Act Cases and Quashing	Food Safety and Standards Act Cases	0.86
Food Safety and Import Regulation Disputes	Food Import Clearance and Regulatory Compliance	0.84
Quashing Food Safety Proceedings	Food Safety Prosecution Challenges	0.81
Food Safety, Seizure, and Business Operations	Food Safety and Seizure Disputes	0.80
Food Safety Act Convictions and Sentence	FSS Act Violations Quashed by Precedent	0.78
Food Procurement and Safety Disputes	Public Procurement and Tender Disputes	0.76
Food Safety Act Cases in Madhya Pradesh	Food Safety and Standards Act Cases	0.71
Food Business Licensing Under FSS Act	Food Safety and Adulteration Cases	0.70
Food Safety Officer Recruitment Disputes	Anganwadi Workers and Officer Recruitment	0.68
Ethanol Production Regulation	Ethephon and Ethanol Regulatory Disputes	0.57
Divergent / Low Consensus Matches (Score < 0.55)		
Trademark Infringement and Product Disputes	Food Safety Licensing and Meat Business	0.38

Table 2: Semantic alignment between Gemini and Qwen cluster labels.

GPT	Qwen	SBERT Score
Food Safety Act Violations and Appeals	Food Safety and Standards Act Violations	0.97
Appeals and Remedies Under Food Safety Act	Food Safety Act Enforcement and Appeals	0.92
Bail Applications in Food Safety Violations	Food Safety Violations and Anticipatory Bail Cases	0.92
Food Safety Compliance and Violation Cases	FSS Act Violations and Regulatory Compliance	0.91
Food Safety Enforcement Challenges	FSS Act Violations and Regulatory Compliance	0.90
Food Safety Violations and Legal Challenges	Food Safety Prosecution Challenges	0.89
Licensing and Regulatory Challenges in Meat	Food Safety Licensing and Meat Business	0.88
Quashing of Cases Under IPC and FSS Act	Food Safety and Standards Act vs. IPC Cases	0.88
Food Safety Violations and Legal Proceedings	Food Safety and Standards Act Violations	0.87
Limitations in Food Safety Prosecutions	Food Safety Prosecution Challenges	0.87
Food Safety Regulation Disputes	Food Safety Act Enforcement and Appeals	0.85
Import and Customs Disputes	Food Import Clearance and Regulatory Compliance	0.84
Food Safety Officer Recruitment Disputes	Anganwadi Workers and Officer Recruitment	0.71
Ethanol Production and Supply Restrictions	Ethephon and Ethanol Regulatory Disputes	0.62
Tender and Procurement Disputes	Public Procurement and Tender Disputes	0.61
Divergent / Low Consensus Matches (Score < 0.55)		
Trademark and Commercial Disputes	Food Safety Licensing and Meat Business	0.35
Sealing Actions and Legal Disputes	Food Safety and Standards Act Litigation	0.32

Table 3: Semantic alignment between GPT and Qwen cluster labels.