Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEEP SEMANTIC ONTOLOGY WITH CASE-BASED REASONING FOR LEGAL DOCUMENT AUTOMATION AND HYBRID RETRIEVAL-SUMMARIZATION

NAIMOONISA BEGUM[1], G. REKHA[2]

[1]Research scholar, Koneru Lakshmaiah Education Foundation, Department of Computer Science and Engineering, Hyderabad-500075, Telangana, India
[2]Associate Professor. Koneru Lakshmaiah Education Foundation, Department of Computer Science and Engineering, Hyderabad-500075, Telangana, India
E-mail: [1]naimoonisa@gmail.com, [2]gillala.rekha@klh.edu.in

ABSTRACT

As more of the judicial documents are digitized and the information becomes more complex, it has become more difficult to provide accurate information retrieval and summarization in legal documents. While current keyword-based retrieval systems often miss out on semantically similar, but lexically distinct, precedents, generic neural summarization models may struggle to capture legal reasoning, statutory context, and long-range context. The restrictions are especially relevant in the Indian context where judgment writing is lengthy, structurally varied, precedents dependent and so is the influence of the specific terminology in these judgments. This paper presents the Legal Aware BigBird Longformer Encoder–Decoder (LawBird-LED), a transformer-based hybrid model designed for intelligent legal retrieval and automated legal summarization to tackle these challenges. The proposed system involves using semantic ontology-based search to retrieve legally relevant information and case-based reasoning (CBR) retrieval to retrieve the precedent aware contextual knowledge for improving legal understanding. Furthermore, the structure applies both sparse attention (BigBird) and contextual attention (Longformer) for long legal documents and preserves the global relationships across the document and the local semantic relationships within the document. The summarization approach is presented as a multi-layer summarized approach to gradually improve the summarization process, and a set of interpretability techniques for the retrieval and summarization decisions is presented for transparency. Experimental comparison with the legal summarization and transformer-based benchmark models shows the proposed framework is able to generate legal summaries containing meaning and content well, and with good context capture. The proposed LawBird-LED achieved a ROUGE-F1 score of 99.0%, BLEU score of 98.8%, Legal-Sim score of 99.1%, METEOR score of 98.9% and Coverage score of 99.2%, which showed good lexical alignment, semantic consistency, and the retention of legal concepts. An effective and interpretable legal information retrieval and automatic legal text summarization system based on semantic retrieval, precedent-guided reasoning and hybrid contextual attention. The results reveal that augmenting symbolic legal knowledge with precedent-based reasoning and long-context neural representation can facilitate the semantic retrieval and legal-summary fidelity. The study therefore shows how LawBird-LED can be an interpretable decision support system for access to legal information, as well as pointing out issues of cross-jurisdictional generalization, computational efficiency and expert validation that need to be further explored.

Keywords: *Smart Legal Retrieval, Semantic Ontology, Case-Based Reasoning, Indian Jurisprudence, Automated Legal Summarization, and Legal Information Retrieval.*

1. INTRODUCTION

The digitization of judicial processes, law reports, filings, and secondary legal materials has had a dramatic impact on the legal practice, so much so that over the last few years, the change has been quite substantial[1]. As a result of this transformation, the practitioners and researchers are no longer going through countless volumes of bound reporters; they are instead reviewing large and varied corpora of judgments, orders, pleadings, statutes, and circulars that differ quite radically from one another in length, structure, language, and legal significance. An excessive amount of data is

available, and this fact determines that the classification of legal documents is a necessary step before any application of the data (search, precedent retrieval, citation analysis, knowledge extraction, or summarization)[2]. The main reason is that besides the fact that correct categorization allows a targeted ranking as far as the search is concerned, it also provides help to the specialized summarizers that have to follow the legal form and function. Compared to regular or normal text, legal documents are very different - they have identifiable substructures (headnotes, facts, issues, ratio decidendi, obiter, holdings, citations, and orders), are dense with legal jargon, and contain references to laws, and cross-citations - and all these characteristics clearly indicate that conventionally generic text classifiers will not perform well unless they have been adjusted to the domain. Moreover, in Indian legal parlance, this problem is much more pronounced as there are multi-tiered courts, the jurisdictional subtlety, and the differences in the drafting styles of the benches and the registries; Consequently, classification models must not only be able to uncover the surface lexical signals but also integrate conceptual and relational knowledge (e.g., "wrongful termination" is just a new term of the old practice of contractual breach in employment law or a decision that mentions a certain statute changes the precedent landscape of the cases that interpret that same provision).

Law summarization[3, 4], on the other hand, may not merely be a shorter version of the sentences: the summaries should give the legal context, point to the controlling ratios, inform about the material facts and outcomes, and show the precedential value if any - otherwise, the shortened version may even be inferior, misleading, or fruitless. Thus, it can be said with much confidence that classification and summarization are very closely linked: classification finds the legal genre/issue (which thus becomes an indication of what should be kept in the summary) and, at the same time, summaries are more efficient when they are issue-aware and indexed by the classes (so, a researcher looking for "contractual breach in employment" not only gets relevant cases but also gets brief, comparable summaries which highlight the contract terms, remedies and court reasoning).

One of the main points for achieving classification and summarization of a level comparable to legal quality is to apply a coherent and multi-layered approach. It should not be just one dimension but a combination of statistical learning along with structured and domain knowledge[5]. The way the classification characteristics are affected by the lexical and syntactic levels is still meaningful, but now these features are not that powerful on their own. They are effective only when they are accompanied by semantic propositions coming from legal ontologies and knowledge graphs: for instance, named entities for parties and statutes, relations like amends, overrules, or applies, and conceptual mappings which allow users to bring the informal expressions closer to the legal concepts (e.g., "forced resignation and constructive dismissal.

The embedding-based and essentially representations (e.g. transformer-based models for the legal domain or transfer learning from large transformers) are just perfect for contextual encodings but the ontology-driven features are more beneficial for query expansion that they not only provide the most accurate results but also the results that have the same kinds as the original query[6]. As an example of this, we can say that a multi-label and hierarchical classification are the most common methods applied for these situations where a single judgment can be hierarchically classified as contract, tort, and procedure simultaneously, citation network features along with temporal metadata can help increase the performance by making more efficient use of the precedent structure[7]. The CBR function not only labels but it, using structured feature vectors (from ontology mappings), citation proximity, and embedding similarity as tools, moves on to locate similar records of the past; thus, it is not only allowing decision-making of classification (label assignments confirmation) but also making summarization at the same time, by providing relevant precedent snippets and reasoning patterns, easier.

Summarization design options are influenced by the changing nature of the constraints of the domain. Extractive methods are seen as the most truthful ones since they do it by directly quoting the critical passages (facts, holdings, operative orders), but the problem with such outputs is that they are commonly too large to be easily managed. Abstractive models (transformer sequence-to-sequence models fine-tuned on legal summaries) still can be the cause of hallucination or loss of exact legal meaning while they are getting smaller, human-readable digests of the original[8]. Thus, a viable pipeline is often some kind of hybrid between these two methods ontology-aware sentence scoring for the selection of the most legally salient sentences, followed by a constrained

abstractive pass (retrieval-augmented possibly) that is rewriting but citation and outcome statements kept intact. Also, the ways in which legal tasks are evaluated have to highlight the legal priorities that the tasks have. Still, legal adequacy involves a requirement of measures for factual consistency, coverage of legal issues, preservation of holdings, and citation completeness hence, specialized metrics like Legal-Sim and structured human evaluation (legal adequacy, fidelity to ratio, usefulness for precedent search) are very important[9].

The final stage, which is the one that still behavioral dealing with classified operations and summarization in the context of legal information retrieval system, focuses on the problem aspects and ethical considerations. Such are factors that are, because of their very nature, extremely peculiar to the legal sphere. By and large, the whole chain of processes will envelope data preprocessing (OCR and normalization for scanned judgments), ontology mapping and entity extraction, vector indexing and CBR retrieval, classifier scoring and reranking along with a final, summarization component that will be the one to add provenance (passage anchors and citation pointers) for each of the generated summaries. Automated human metrics evaluation (ROUGE/BLEU/Legal-Sim) is only one of the grounds of decision-making[10]. Such an evaluation model also requires routine human audits with the possibility that machine summaries scoring highly in overlapping metrics still omit controlling ratios or wrongly state remedies, thus, human annotators on gold labels and gold summaries have to not only train but also monitor these requirements continuously [11]. Over the last decade, AI and deep learning have been digital technologies that have largely changed the legal research system[12].

The deep learning enterprise in the legal sector has been progressively growing[13]. The most important aspect that differentiates them from others is the fact that they can learn the representations of words, phrases, and documents in which the underlying meaning is implied, therefore, enabling machines to understand legal expressions similar to humans. To illustrate, transformers that are trained on legal corpora are not only good at recognizing the relations between words but also in identifying that the phrase "breach of contract in employment" is probably the closest one to wrongful termination and contractual disputes besides that there is no exact keyword sequence. Deep summarization models that indeed have been able to leverage attention mechanisms and sequence-to-sequence architectures to turn very long court decisions into readable, semantically correct, and practically useful ones by picking out the ratio decidendi, the material facts, and the precedential value are an example of the same concept.

In fact, these are just the basic facts of the Indian courts as the different statutes, regional practices, and layers of reasoning make interpretation more complicated[14]. When using synonyms or conceptually related phrases, the number of relevant results will be minimum, if not zero. The truth is that even deep learning models that are generally successful at contextual embedding do not have a proper legal domain alignment. As a result, errors occur, for instance, where close semantic but legally different concepts are coalesced. It is worthwhile noting that the expression redundancy in connection with employment law and "redundancy" in information language is a proper example of an issue case. Normally, text summarization models that rely on news or conversational data are not able to depict the discourse structure of judicial decisions accurately. Rendering events in a court can vary from very minute differences between obiter dicta and ratio decidendi to several legal issues that are mingled. An absence of semantic grounding specific to the domain causes the generation of summaries that may be quite fluent but at the same time, they can be legally incorrect and thus, not helpful to the practitioners[15, 16]. Moreover, the whole Indian legal documentation with the history of judgments, statutory amendments, and administrative decisions for hundreds of years has been converting from a data treasure into a data overload problem where the traditional retrieval systems cannot filter out the noise and highlight the relevant precedents that are contextually important.

While transformer-based legal NLP has made significant advances, legal retrieval and summarization are still unsolved challenges due to the possible differences in doctrinal or precedential equivalence of linguistic similarity. Two Judgments can discuss the same legal question with very different languages, and phrases with identical words can be in different contexts within the statutes or jurisdiction. Moreover, long decisions spread out legally important information, which involves facts, arguments, statutes, precedents, reasoning, and final order. Such systems could therefore be limited to returning "relevant" cases that do not actually contain controlling cases or

cases that return summaries that lack legally significant information. The issue is especially relevant to Indian judicial system as it faces problems of heterogeneity in the drafting practice, significantly cross-citation of provisions, multiple amendments to legislation, multiple jurisdictions and jurisdiction-specific legal concepts. Thus, it is not enough to improve lexical similarity, an effective system needs to be able to incorporate domain knowledge, precedent sensitive reasoning, long context modelling, and explainable retrieval.

The principal case-based reasoning methods strengthen the tool as it mimics human lawyers who find similarities in past decisions and thus make the search more intelligent and useful. Hence, the motivation is to facilitate the legal professionals in swiftly accessing relevant judgments without query accuracy being compromised and to indicate, by means of the implementation of this work, how the use of semantics, reasoning, and deep learning can be the solution to the existing gap between keyword-based and generic NLP systems. Consequently, the authors through this project would like to witness the progressive, the easily accessible, and the intelligent research legal ecosystem not only accommodating the practical realities of the Indian courts but also bringing about a paradigm shift in the global legal information systems. The major research objectives of this paper are as follows:

- To build a Legal Aware BigBirdLongformer Encoder-Decoder (LawBird-LED) network which can handle long legal documents while maintaining domain-specific semantic and contextual relations.
- To integrate the semantic ontology-based retrieval and case-based reasoning (CBR) methods and achieve better query retrieval accuracy in legal documents, and to achieve a more contextual understanding based on case incorporation.
- To develop a hybrid legal summarization system which produces coherent, context-rich and semantically coherent summaries, while preserving legal concepts.

This study has four major contributions. First, it is building LawBird-LED, a sparse attention, long-context legal representation system, which combines sparse attention with Longformer/LED processing for longer legal documents. Second, it introduces domain-specific legal ontology into retrieval and representation in order to be able to detect and retrieve conceptually similar cases that do not necessarily share words. Third, it incorporates precedent-sensitive and semantic retrieval to add precedent-situated contextual information to the retrieval–summarization pipeline. Fourth, it creates a hybrid extractive–abstractive summarization model incorporating legal-semantic constraints that aim to maintain the fidelity of statutory references, legal concepts, argumentative structure, and context. These components collectively offer an integrated solution to approaches that tackle each of the above problem aspects individually in a separate manner.

In this research, we concentrate on the three tasks of automated semantic retrieval, precedent-aware case matching and legal-document summarization in the context of Indian judicial documents. The model assesses whether the quality of retrieval and summarization can be enhanced by combination of the three methods: ontology-based semantic representation, CBR and long-context transformer modelling. The study does not try to replace or supplant judicial or professional legal reasoning, predict legal results, deliver legal counsel, or enforce the framework in every international jurisdiction or be mass-produced for real-time deployment. Within the current experimental scope, processing in multiple languages, transfer of ontologies across jurisdictions and comprehensive practitioner-based usability testing have not been experimented with.

Section 2 has detailed information on the scientific literature review that encompasses different models and methods to explain the retrieval of legal information problem, along with a review of legal domain features, that includes pros and cons of the traditional keyword-based systems, ontology-driven frameworks, and deep learning methods applied to the legal domain. Section 3 is the proposed methodology description, which presents a systematic explanation of the principles of our intuitive retrieval system that integrates semantic ontology, case-based reasoning, and natural language processing for automated summarization as well as the component functionalities description. Section 4 is about performance evaluation and results comparison description where the effectiveness of the system is confirmed by various tests such as ROUGE, BLEU, and Legal-Sim. The final section, i.e. Section 5, just gives a summary of the main points and contributions of the paper, apart from providing a brief outline of future work, which includes more sophisticated deep learning models, multilingual legal corpora, and the development of domain-

specific ontologies for retrieval and summarization of Indian jurisprudence.

2. RELATED WORK

Over recent years, the identification of the task of locating and summarizing legal information has been a continual process. The root of the issue was, on one hand, the enormous quantity of law-related digital documents and, on the other hand, courts, practitioners, and researchers' dependence on so-called intelligent decision-support systems. The characteristics of traditional legal search engines were, to a large extent, just keyword matching and rule-based methods. Those could function in very simple cases but most of the time were not able to show the deeper semantic aspect of legal terms or the contextual relationships found in the judgments and statutes. The limitation is the main reason for the most severe problems in the field of Indian jurisprudence where the multiple meanings of terms, the intricately structured legal reasoning, and factors like different laws of a particular geographical area in which the court are located, are just a few of those, that call for more advanced ways of organizing, finding, and even summarizing documents. The researchers have turned to the use of ontologies, logical patterns of reasoning, and lately, deep learning models that can simulate a human and be able to process and comprehend complicated legal materials to get over these limitations.

One of the primary goals of most of the work in this field has been to semantically query legal databases, which have been achieved largely through the development of domain-specific ontologies and knowledge graphs. The mentioned systems provide concepts, relations, and hierarchies of law that can be used by search engines to give more advanced results than those that are mere matches of keywords searched. In addition to that, case-based reasoning (CBR) has been declared one of the methods of imitating the behavior of legal professionals, who accordingly remedy new problems by referring to old precedents with the most similar fact patterns and legal principles. One of the latest research concentrations is on delivering the legal domain extractive and abstractive summarization due to the adaptation. In this way, very long judgments can be turned into summaries, and summarily, the key legal arguments and sources on the legal issue can still be there. The different research strands combined constitute the foundation of the designed system that is the combination of semantic ontology, CBR, and NLP-based summarization to surpass the most common gaps in the existing legal information systems.

2.1 Review of the Existing Studies

Jain, et al [17] developed a Summ, a completely different and new sentence-score method, that is, a method extraordinarily suitable for legal documents summary for the most part. They blend one combination of supervised and unsupervised methods only to enhance the quality of the summary generated. First, a supervised summary relevance prediction model is built, hence each input sentence is assigned a score reflecting the probability of the sentence being a good summary. At the document level, this is the point of the combination, a score improvement method based on clustering is used to select the final sentences which not only have high importance but also are different and show the overall content of the document. The paper authors demonstrate their model by comparing it with several other models on two different datasets, i.e., BillSum and Forum of Information Retrieval Evaluation (FIRE). Their results indicate that DCESumm is the most prominent of the other state-of-the-art methods as it yields ROUGE F1 score improvements of 1 to 6% on BillSum and 6 to12% on FIRE, thereby, being the indicator of its stability across different datasets. This study is one of the successful instances of combining supervised and clustering-based unsupervised refinement leading to the enhancement of quality and confidence of legal text summarization.

Dan, et al [18]introduced a comprehensive summarization schema which, on the one hand, comprehensively unveils the semantics of the judgment texts of law, and, on the other hand, their own characteristics to improve the Legal Judgment Summarization (LJS) quality. Considering that legal judgments are generally quite rigid in formatting, the authors in their second step segment each document into three quite brief sections corresponding to its standard layout. For the summarization challenge, they come up with two augmenting models: an extractive summarization prototype, BSLC, and an abstractive summarization prototype, LPGN. These models have a common characteristic that is the employment of Lawformer as the encoder. The system, however, does not treat all parts the same between different parts of a document and uses different summarization strategies for different parts. The research points to

the massive potential of integrating domain-specific pre-trained models with the structure-aware summarization strategies to take the legal text processing scenario further.

Vianna, et al [19]performed an effective summarizing organization of legal documents of an undefined number. They must do it by concentrating on the disclosure of the underground topics and themes that will simplify their usage, e.g.case retrieval and judgment prediction. These techniques are extremely successful in capturing the semantic relations and contextual parts of the legal jargon without any disruption. The idea of topic modeling with embeddings only shows the potential of the system not just to group but also to summarize the relevant research in the field of law, which may be a sign of the latent thematic structure of those documents. Consequently, it is no longer difficult to make large collections of easy and manageable sizes. They have confirmed their framework with an experiment that includes four different datasets with short and long legal documents. The outcome indicates that their method works effectively and can be applied to different kinds of legal texts. This is a point of transition towards the integration of topic discovery with present embedding practice which leads to improvement in legal corpora grouping and summary, thus, making a big contribution to the field of legal research and predictive analytics.

Ragazzi, et al [20]come up with a number of innovations, one of them being LAWSUIT, a dataset that is large and diverse and consists of 14.000 court decisions written in Italian, to which some expert-written, abstracted maxims of the Constitutional Court of the Italian Republic that were carefully selected are added. Dealing with long and complex legal documents as the main and complicated issue, the dataset depicts these cases in which the relevant data are almost equally spread throughout the lengthy judgments, thus the problem of summarization being quite difficult. The authors have done a lot of different experiments with sequence-to-sequence architecture as well as segmentation-based approaches to benchmark the performance of the models. They have concluded that the latter is the one that is confirmed to get more excellent results in both full data and few-shot scenarios. In fact, the paper is going through the processes of each of these methods to show the advantages of the pre-processing step which, in turn, allows for better handling of very long legal texts. LAWSUIT's occurrence beyond the methodology is like an incomprehensible treasure for the community of researchers as it affords the study of the real-world legal summarization and automation tasks in a multilingual context.

Deroy, et al [21]have investigated the scenarios of different summarization models for summarizing legal case judgments are in the limelight with a focus mainly on comparing domain-specific abstractive methods, general-domain LLMs, and extractive summarization. The quality of the summaries generated is measured with standard metrics in their study which point to the abstraction summarization models and LLMs performance as the most rational i.e. They surpass the extractive methods in terms of coherence, coverage, and overall source fidelity. This result is a sign of the increasing generalization of large-scale pre-trained models for dealing with the complexity and verbosity of legal texts as they reveal subtle logic and provide brief but informative summaries. The current work, by interchanging jurisdictions and document types, not only portrays the characteristics of the current models but also suggests their limitations with the helpful insight for domain adaptation and model selection which can influence the performance of legal judgment summarization.

Fan, et al [22]blending ensemble learning to merge the advantages that two sub-networks possess, which differ in information processing methods, optimization losses, and thus provide similarity predictions that are more robust and accurate, without any hurdles. Besides this, the experimental results indicate that the dual-network concept significantly facilitates the identification of the similarity of cases which, in turn, results in faster precedent cases retrieval and further legal text mining and judgment prediction development. One way of enhancing the resemblance that is likely to be observed between two legal cases is to create algorithms that can give a more accurate estimation of such similarity. As a result, this is practically the question that is at the center point of precedent retrieval and case-based reasoning in legal AI systems. Thus, authors have introduced the ensemble-based learning family frameworks to get this goal. Frankly, their approach is a combination of two complementing sub-networks, a similarity representation sub-network, and a binary classification judgment sub-network. The similarity representation sub-network makes use of contrastive learning which is very effective in representing semantic relations since it not only

brings feature representations of the similar cases very close but also, for the different ones, it generates a wider gap between them. Moreover, the binary classification judgment sub-network is picking sample pairs for feature interaction between text pairs during feature extraction so that the model can find legal texts that have the same or similar relationship patterns.

Yang, et al [23] have come up with a totally different outline for text extractive summarization that is one of the most successful implementations of combining technological innovations such as GANs (Generative Adversarial Networks), TLSTM (Transductive Long Short-Term Memory), and DistilBERT embeddings to attain higher quality in summary generation. The principal reason for employing GANs in their method is that they can operate in the adversarial mode where the generator defines the sentences from which the summary should be created while the discriminator judges the given summaries for better coherency and quality of the whole. By combining TLSTM with transudative learning, the model is not only able to achieve contextual accuracy but also relevance as it focuses on the data that is closest to the test data distribution. Moreover, the use of DistilBERT leads to very successful and efficient sentence embeddings that are not only semantically rich but are also quite lightweight from the perspective of computational resources. The authors find a way to fix this problem by implementing a reinforcement learning strategy whose task is to both stabilize and enhance the training performance. The hybrid model of this experimental research is proven to result in two positives, on the one hand, it causes a decrease in the greediness of sentence selection; on the other hand, it is capable of generating more coherent, fluent, and representative summaries. As a result, it can move a big step forward in the extractive summarization field.

Godbole, et al [24]presented the study of Long Context Large Language Models (LLMs) not only for a single but also for several document summaries, to which extent such models are able to track long-range dependencies, thus, providing summaries that can be both structurally valid and easily adapted to the domain. They propose a conceptual framework for the use of long-context LLMs in open-world multi-document summarization scenarios since such tasks, besides permitting for sizing, also allow for the integration of enterprise applications and systems. This article is multi-disciplinary research with a focus on legal, human resources, finance, sourcing, healthcare, and media in which the authors find that these models can go so far as to require the specification of each area's characteristics for training. The results reveal that long context LLMs have the potential to improve the whole summarization process quality to be shortened and to be accurate, thus leading to the enterprise deployment of these models. Essentially, the work is a major transition in the field as it not only presents the technical capabilities of long-context LLMs but also their practical field value in the confirmed sectors' summarization workflow.

Yadavm et al [25]described TGETS (Textual Graph-Based Extractive Text Summarization) is fundamentally a new method of summarizing the text that focuses on representing the crucial information of a single document via a graph. Unlike customary methodologies, graph-based techniques do not go as far as to disassemble sentences and treat each phrase or word separately. Nevertheless, in this case, sentences are initially clustered and depicted as nodes in the graph, while the relationships between sentences are at the edges, thus allowing the preservation of both the structural and semantic aspects of the text. This stage makes the choosing of the most informative excerpts of the text possible. As a solution test, it was realized for the BBC news articles dataset, and the ROUGE metric was employed for the evaluation. This success points to the further use of graph-based paradigms for the extractive summarization progression, as it opens the potential of sentence-level interactions.

Abimbola, et al [26]discussed about the use of deep learning in sentiment analysis for the Canadian maritime court, and during this process, it unpacks a whole new horizon anticipated not only to ease the seaborne law but also to simplify the legal policymaking through the use of analytics. In simple terms, the author's paper is a legal document automation journey, which they describe as a focus on the application of sentiment analysis in the identification of new norms from the figures of the law. To reach this goal, they present a novel method that integrates the latest deep learning models with law cases for the rapid detection of judge decision bias as an immediate finding. Moreover, this research paper addresses the issue that considers the influence of these biases on the police which, in turn, provides sufficient assistance to computational legal studies and law making. Nguyen, et al [27]have explained the designing of a

new text representation model that involves an attentive neural network and is used to fetch legal documents from the statute laws. Alongside this, they have also mentioned two architectural types, which are hierarchically structured with sparsely distributed attention mechanisms, that is to say, the Attentive CNN and the Paraformer. The key goal of their deep learning attention networks is to be capable of capturing the semantic complexity of single long sentences and articles for the legal domain in an attractive manner. Subsequently, these models are put to the test with multimodal datasets of different languages i.e. English, Japanese, and Vietnamese, and the results indicate that the neural methods with attention outperform significantly the methods without neural retrieval. The study also depicts a trade-off situation between the accuracy and the computational cost where pretrained transformer-based models are good at small datasets but are computationally expensive while the lighter version of the Attentive CNN can be more accurate if the dataset is large. Besides that, the Paraformer model is not only successful to get rid of the limitations of the previous methods on the COLIEE dataset but also, thus, is distinguished as a very potential candidate for the legal domain in the field of real-time legal information retrieval.

Naseria, et al[28] are essentially explaining that alongside BERT (which is a transformer-based pre-trained language model) they have built a deep learning model that primarily focuses on selecting the most probable rules from a given set by just measuring the semantic similarity between the two. Initially what they do is firstly get the sentences, or relevant paragraphs, next change them to the vectors of dense and fixed lengths, and finally, they find out their cosine similarity to determine which words and rules are the closest in meaning. Essentially this means if one uses BERT's contextual embeddings together with the entity positions they can still achieve more accuracy in the legal texts as compared to those which are just based on keywords or syntax. In other words, the use of this method can turn out to be the factor that makes one most proper and accurate legal text correspondence visible. Because of this, the system is not only locating the semantic closess between documents and terms but also making the step of finding and accessing the legally-HG relevant regulations easier. This research stands as one of the prominent cases of the deep Learning-style semiconductor models such as BERT which can have a significant impact on the extraction of the ruling and the analysis of legal texts and more, thus being the door for the next advanced retrieval systems in the legal domain.

The findings of the reviewed studies also show the influence of these factors (j Jurisdiction, language, document structure and application domain) on legal NLP performance. In the field of long-document summarization, segmentation has been stressed in the case of Italian documents and in the case of English, Japanese and Vietnamese documents, there is evidence of the influence of specific language representations on retrieval. In fact, Canadian legal analytics also demonstrates that legal AI tasks might rely on patterns which are particular to the institution and domain. The general trends from long-context research in other domains, such as enterprise document processing, human resources, finance, and healthcare also indicate that generic summarization architectures need to be domain adapted if factual completeness and semantic fidelity are of paramount importance. Together, these results suggest an experimental design and the need to test the proposed framework in the specific context of Indian jurisprudence, not simply in models which are validated in other contexts or other disciplines.

2.2 Research Gap

One of the major use cases of the deep learning methodology is the automation of the retrieval of legal information, identification of the most similar cases and snippet generation. The examples for the implementation of deep learning models[29], transformer, attention mechanism, ensemble learning, graph-based methods, and generative techniques have been repeated extensively to illustrate this scope. Technically, this is just one facet of the comprehensive survey of the innovations in technology. On the other hand, a survey also reveals such a long list of problems that not only are extremely difficult to solve but also remain unsolved by researchers indefinitely. In essence, they can be powerfully effective but fail in capturing all legal semantics, the text structure, and the reasoning patterns particularly in the Indian case of jurisprudence and, thus, by the results they yield, they are not quite satisfactory.

Most of the research papers concentrate on summarizing the legal documents of the English language, whereas the works in the Italian and Japanese languages are relatively few. Furthermore, it appears that these methods cannot be readily transferred to the Indian legal system without a

domain-specific ontology since the legal systems are deeply intertwined with the jurisdiction. On top of that, almost all models simply zero in on the closeness of the words or surface-level semantics and, thus, they only consider those already present in the text or something very similar to decide the similarity of the texts. For one thing, it should be possible to find a proper example of the case "breach of contract" that is described differently with words like "wrongful termination" or "workplace contractual violations". Apart from those, scalability and efficiency are also among the most important issues. In spite of transformer-based models being very accurate, they still consume a lot of computational resources and are, therefore, not always the best choice for large-scale or real-time legal information retrieval, which is the time when courts and practitioners need instant access to the judgments.

Moreover, one of the reasons why technology is so scarcely used in the legal field is that just about all research works concentrate on different summarization methods, e.g., purely extractive, abstractive, and combined ones and mainly optimize the linguistic aspects that can be expressed with scores like ROUGE, BLEU, or general coherence. In addition, they very marginally use legal evaluation metrics and the logical flow of legal reasoning for the integration. As a result, the summaries that are being generated are not very handy for the legal executors. In short, they still must personally go through the whole texts in order to find the summaries that are both contextually and doctrinally accurate. Another reason why human legal experts are still needed is the fact that there are very few references where CBR is explicitly mentioned in just a few approaches, and also human legal experts are not only indispensable because CBR is still the primary way of thinking by lawyers and judges but also because there is no single system that considers semantic ontology, domain-specific knowledge representation, and CBR all at once[30]. Therefore, the search engines of today quite either very shallow (concentrating on surface-level matches) or very opaque (black-box neural models with little interpretability). These drawbacks cause the demand for an intelligent retrieval and summarization system that will not only be able to use deep semantic understanding through the ontology-driven approach but will also be capable of mimicking human-like legal reasoning through CBR, thus bridging the gap between computational efficiency and domain. The authors, through their primer research, have not only dealt with current constraints but to quite an impressive degree also accomplished what present solutions fail in terms of accuracy, context-awareness, and practical usability of a semantic ontology-based retrieval, and even a CBR-driven summarization system for Indian legal framework.

The conceptual underpinnings of the proposed framework are thus based on the mutually complementary restrictions of symbolic and neural legal information processing. While ontologies contain explicit domain semantics and interpretable relation between concepts, they do not have a good ability to model complex contextual variation. dense representations of transformers: contextually similar but not necessarily explicit doctrinal relationships or precedent structure. CBR also adds precedent-sensitive reasoning and relies on the proper representation of factual and legal similarity. There is a need to adopt long-context attention as legal evidence is frequently spread throughout the judgment text. The proposed LawBird-LED framework thus does not view these approaches as mutually exclusive approaches but rather complementary aspects of a single legal information-processing architecture.

3. PROPOSED METHDOLOGY

The proposed system is an automated aviary capable of carrying out judgments annotations, which, with the help of one command, is said to semantically map the ontology, use case-based reasoning, and deep learning models all along one pipeline. Summarization in the traditional sense, which merely employs keyword matching or simple linguistic features, is no longer able to meet the structural and semantic aspects of judicial documents, thus the need for such an innovative approach is being indicated. Initially, the system would be given either keywords or phrases as queries which, besides, by mapping into the legal ontology, would be further expanded for the inclusion of related statutes, precedents and legal concepts, thus not only giving a wider range but also ensuring conceptual relevance. Hence, with the help of case-based reasoning which combines the facts, issues, and the judicial outcome for the selection of the closest contextually similar judgments, the relevant case documents are not only retrieved but also assessed. Legal Aware BigBird Longformer Encoder-Decoder (LawBird-LED) is an innovative deep learning architecture, which the framework employs to take the most precise summaries as output. Thus, the system is

capable of generating both full and partial summaries that explain the legal reasoning, provide references to the statutes, and follow the judicial logic, thereby the task of legal experts is made less difficult as the three efficiency, interpretability, and faithfulness to domain-specific needs are achieved simultaneously.

3.1 Research Design & Experimental Workflow

The design-science-based experimental computational research design is used in this study. The design-science part stands for building an integrated legal information-processing artefact, which integrates ontology-driven semantic retrieval, CBR, long-context transformer representation and hybrid summarization. The experimental part compares the artifact quantitatively with the baselines of neural networks and transformer models, under controlled training and testing conditions. The research process includes seven sequential steps: (1) acquisition and partitioning of legal corpora, (2) pre-processing and legal-specific normalization, (3) legal-specific semantic mapping and query expansion by using the legal ontology, (4) legal-specific semantic retrieval and ranking, (5) LawBird-LED based hybrid summarization, (6) precedent-aware CBR, and (7) comparative evaluation based on lexical, semantic, coverage, and legal-domain metrics.

Data gathering and formatting is the actual base for a legal retrieval and summarization system, which is the first, and the most essential, step of the brought work. The main emphasis has been put on existing datasets to build a system like that. First of all, there is IN-Abs[31].It denotes to 7,130 case documents from the Indian Supreme Court, along with their corresponding abstractive summaries. An abstractive summary is very helpful here as it is a summary written in natural language and gives the whole idea of the case, the main facts, legal issues, the judgment, and the reasoning, and it is not only a list of the sentences that exactly match the original documents. The seven thousand document-summary pairs from this collection have been used for training a deep learning model. Besides, 100 pairs have been set aside as a test set to check the model's summarizing skills. So, the models can learn from a large enough text to acquire the meaning relations, the organization patterns, and the legal terms, which are typical for the Supreme Court of India, and at the same time, a separate set can be used for the evaluation of their accuracy.

IN-Ext[32, 33] is another dataset that contains 50 instances of the Supreme Court of India case documents, along with the extractive summaries that have been independently created by two law experts, A1 and A2. Each expert has summarized the case in two forms: first, the full summaries, in which the court was considered as one whole, and second, the fragment summaries, where the document represented as a group of sections such as ‘analysis', ‘argument', ‘facts', ‘judgment', and ‘statute’. This kind of representation of the document in a dual format not only facilitates the detailed understanding of the document structure but also allows the system to learn how to summarize the whole case besides producing the section-wise summaries that keep the logical and argumentative flow of legal judgments. Segment-wise summaries are of great use when models are designed to focus different aspects on a case separately. Such a feature may thus result in the system getting better and more human-readable summaries, which it is proposed. Basically, a model with the help of expert-authored summaries can be exposed to a very high-quality reference that is not only the logical justification but also the domain-specific linguistic nuances.

Fundamentally, the datasets are just three different data types that have been organized in a certain manner. These datasets illustrate the details of the train, validation, and test splits as well as the metadata, document text, and summaries. For instance, metadata might be the information that explains a case, e.g., the date, the parties involved, the sections of the law referenced, and other attributes related to the issue, which, by both simplifying the extraction and the summarization of the data, may enable further processing. Now, these datasets are being remodeled to meet the necessary requirements for the next steps, i.e., semantic ontology-based retrieval, case-based reasoning, and deep learning-driven summarization. The chosen datasets provide an ideal environment for the models to be trained in legal semantics, reasoning patterns, and document structures. Thereby, they have the capacity to generate summaries that are not only grammatically and logically coherent but also of lawmaking sense.

3.2 Preprocessing

The input query stage is one of the additional, major new steps that the new system is implementing. Besides, a massive pre-processing of the input is a very necessary step, in fact, the

beginning step, and then using these pre-processing to link up the human language and the machine understanding in the legal domain. This implies that the user is asked to input a word or a short phrase that describes the concept, the problem, or the term of law for which the system is to find the most relevant Indian court judgments. For example, a user may type "breach of contract in employment" to get all the judgments that refer to contractual violations, wrongful termination, and the rest of the employment law, even if the case documents use different words. As a result, the input query is the chief user-channel through which the user interacts with the retrieval system, and the degree that the processing has a direct impact on the relevance of the retrieved documents besides the performance of the subsequent summarization phase. Considering the complexities and the highly technical nature of legal language, this input pre-processing, if semantic fidelity, accurate matching, and contextual understanding are to be reached, should be handled very carefully.

All the operations mentioned above are very important for the input text to be normalized and for noise reduction. The whole concept is here to have the user query broken down to its smallest units (words) or meaningful phrases. The main idea of stopword removal is to get rid of the most common words which do not add much sense to the text such as "the," "and" "of." Lemmatization is a process that involves changing one word in different inflection forms into its base or dictionary form, e.g., "terminated" to "terminate" and "violations" to "violation." Consequently, this procedure enhances the system's capacity to locate the query in various grammatical forms of one legal term in documents.

<Figure 1>

Moreover, preprocessing captures the legal specificity by means of legal-specific normalization, a section for Indian legal documents that deals with the peculiarities of such documents. It features regularization of references to statutory sections, case numbers, and, apart from the most regularly cited laws or regulations, also of the terminology of the particular domain like IPC, CrPC, or the specialized judicial phrases. Such normalization is actually a very basic factor since it allows the system to accurately match shortened queries or different wording of legal concepts with the right entries in semantic ontology and corpus, thus giving the retrieval system the ability to not only obtain the same but also conceptually related documents.

Sentence segmentation is essentially dealing with a pre-processing package of an extractive summarization pipeline that is breaking down the system into individual sentences or logical units. This step is, without doubt, very crucial when the downstream models, for example, graph-based and attention-driven deep learning models, are given the task of scoring the relevance or value of a single sentence for summary creation. Thus, the system can utilize semantic ontology and case-based reasoning methods at the sentence level, after the text has been cut, those sentences which are the most contextually relevant for a correct and logical summary can be picked up. Typically, the best method for handling input queries is preprocessing, which, as a semantic alignment process, changes the subsequent stage to technical text cleaning. Semantic action extracts the user's idea, simplifies the legal terms, and additionally guarantees that the input is arranged so that the semantic ontology-based retrieval, case-based reasoning, and deep learning-based summarization can be efficient. The major steps involved in this preprocessing stage are illustrated in Fig 1.

3.3 Semantic Ontology Based Mapping

In the proposed work, after the preprocessing process has standardized and tokenized the keyword or phrase typed in by the user, the semantic ontology-based retrieval process will be the key linking the intent of the user and the applicable legal material in the Indian court domain. The query is transformed into a set of concepts in a structured legal ontology that is constructed based on Indian statutes, judicial principles, and legal doctrines instead of analyzing the query using exact word matches only. This implies that when a user types a phrase like breach of contract in employment and the system will not only recognize the literal word but also associates it to the structured features such as employment contract, wrongful termination, and constructive dismissal, as well as connections to the specific section of the Indian Contract Act or the statute on labor matters. Such mappings can be found by contextual embedding of legal-domain transformers that make sure that even minor variations in legal text are properly aligned with the ontology. This will directly work towards the purpose of the proposed system as it will enable the retrieval process to go beyond the superficial match of the

keyword to a deeper semantic interpretation of the query in the context of the Indian court rulings.

After identifying the relevant ontology nodes, the system searches the query with the help of the structured relationships in the ontology. This covers the synonyms, hierarchies involving broader or narrower categories of legal concepts, and procedural/precedent based relationships, which relate cases and statutes. This growth makes the relevant documents available even in cases when the specific phrase is not present, getting to the point where manual summarization and keyword-based search that existed in the past were the source of limitation to legal practitioners. The growth also assists in the capturing of the implicit knowledge in the legal documents, which is especially vital in the Indian court cases where in most cases precedents and citations are in the center of judgments.

A major improvement in the retrieval process that makes it easier for lawyers to access conceptual legal judgments is the way documents are indexed to merge the symbolic tags (statutes, case references, court metadata, etc.) with the dense embeddings produced by legal transformers. The core idea of this hybrid retrieval is that it allows the system to find not only the legal provisions that are directly mentioned but also those that are the latent semantic connections between the legal issues. As an illustration, embedding similarity can find a closest case as employment dispute example if the term "breach of contract" is not directly referred to but the case is talking about termination, the company's obligations, and remedies. After that, a scoring function is used in re-sorting the retrieved documents that considers such factors as ontology-based relevance, embedding similarity, case metadata, and citation significance. Therefore, it is a way of indirectly showing the work that the most legally significant and closest-context cases are surfaced for summarization.

The new system is a legal ontology-based retrieval for the Indian courts, is constructed in such a way that information not only relevant but also semantically understood with respect to the query's intention get fetched at the processing stage. This, therefore, has two big effects on the entire flow of work: In one aspect, it allows higher recall and precision, which are two typical issues in legal information retrieval because of the nature of the law and its verbosity; from another side, it gives up interpretability and traceability as each retrieved example can be explained with reference to the ontology concepts and relationships that show how it was selected. The case-based reasoning and deep learning–summarization modules receive such output, which changes the system into one capable of picking out the most relevant portions of the decisions, simultaneously, legal professionals can still access the reasoning path via the ontology. Hence, semantic ontology–based retrieval becomes the primary point of the suggested work, being the stage that not only makes the automated summaries of Indian court judgments accurate but also legally defensible and in line with the users' requirements.

Algorithm 1: Semantic Ontology based retrieval process for searching legal documents

Input: Preprocessed query with keywords or phrases $\mathcal{Q}$;
Output: Selected most relevant legal documents $\mathbb{D}$;
Step 1: Apply a transformer based legal language model by representing the input query $\mathcal{Q}$ with bag of terms;
Step 2: Encode the terms into an embedding vector as shown in the following equation (1):

$$q = \mathfrak{F}_{enc}(Q) \quad (1)$$

//Where, $\mathfrak{F}_{enc}$ indicates the pretrained model;
Step 3: Then, perform ontology mapping to map the query terms into the ontology concepts represented as $\mathcal{C} = \{c_1, c_2 \dots c_n\}$;
Step 4: Each concept is associated with hierarchical links, case references and synonyms, and its ontology expansion set is obtained as follows:

$$\mathcal{C}_{ex} = \mathcal{C} \cup \delta(\mathcal{C}) \cup \gamma(\mathcal{C}) \cup \phi(\mathcal{C}) \quad (2)$$

//Where, δ is the synonyms, γ represents the hierarchical links, and ϕ is the case law references;
Step 5: Estimate the document embedding similarity measure, in each for candidate legal document, the embedding vector v_j is generated.
Step 6: By using cosine similarity, the query embedding is performed as shown in equation (3):

$$\mathcal{S}_{em}(Q, v_j) = \frac{q \times v_j}{\|q\| \|v_j\|} \quad (3)$$

//Where, v_j is the legal document, and $\mathcal{S}_{em}(.)$ is the estimated similarity;
Step 7: Compute the ontology-based relevance score for each document by estimating the intersection among the expanded ontology

concepts and document concept tags as shown in equation (4);

$$\mathfrak{X}_{ont}(Q, v_j) = \frac{|\mathcal{C}_{qx} - \mathcal{C}_{v_j}|}{|\mathcal{C}_{qx}|} \quad (4)$$

//Where, $\mathcal{C}_{v_j}$ indicates the ontology concept associated with v_j;

Step 8: Moreover, the relevance score is estimated by integrating the embedding similarity and ontology-based relevance using weighted sum as shown in equation (5):

$$Hyb_{sco}(Q, v_j) = \wp \times S_{em}(Q, v_j) + (1 - \wp) \times \mathfrak{X}_{ont}(Q, v_j) \quad (6)$$

//Where, $\wp \in [0, 1]$ is used to control the trade-off among the semantic similarity and ontology-based matching.

Step 9: Finally, based on $b_{sco}(Q, v_j)$, all documents are ranked and the selected top rank documents are returned as output as shown in equation (7):

$$\mathbb{D}^* = TopN\{v_j \in \mathbb{D} | Hyb_{sco}(Q, v_j)\} \quad (7)$$

//Where, $\mathbb{D}^*$ is the retrieved documents;

<Figure 2>

3.4 Case Based Reasoning (CBR) Retrieval

CBR retrieval, user interaction is essentially the same as the extra layer that extends not only the work of keyword matching and ontology-based retrieval but also by reflecting the actual matching of the user's query with previous decisions and legal experiences. When the input query has been preprocessed and semantically enriched through ontology expansion, the system goes on to find those cases in the legal corpus that are similar both in context and structure to the query. In this process, CBR is not only restricted to direct matches of the concepts, but it also reuses the knowledge that is present in the old cases where courts have already resolved similar disputes with the same facts and legal provisions. Therefore, besides following the reasoning pathways that were used in the previous judgments, the system can also provide the user with directions of how similar issues were adjudicated apart from just giving a list of the relevant documents.

In instance of CBR, the user-driven retrieval process is supported by facts, legal issues, possible solutions, or impacts through which a request was made that represented the query. This given representation is now compared with the fact-issue-outcome model of past cases, which has been transformed. In general terms, the system aided by legal embeddings trained on Indian case law does it by capturing the semantic relationships in the language of the law to find the closest match between the query and the past cases. For example, the embeddings will be able to discover that "the wrongful dismissal" and "the illegal termination" are two different ways of saying one thing, i.e. labor law context hence the two terminations of employment disputes that are written differently will still be considered similar cases.

After determining the comparison scores, the system goes on to label those previous cases with the highest similarity regarding the closest context for retrieval as suitable ones. These are the cases that show two aspects: in the first place, as the references from which the user can get the idea of the precedent-based reasoning, and secondly, because of providing the interpretative chain material that supports the current query. The CBR component, for instance, will locate only those instances first where a court has decided upon the contract clauses, remedies, or damages that are similar to the situation, even though the textual expressions being different if the query is about the breach of a contract in commercial transactions. The retrieved documents, thus, are not an indication of only topical relevance but also judicial reasoning patterns that are likely to influence the interpretation of the new query. Also, CBR retrieval can be seen as the part which completes the whole proposed system and thus facilitates the search process by linking abstract semantic representations with the experiential world of judicial decisions. The queries are standardized and normalized in the preprocessing stage, while the retrieval of semantic ontology stretches the horizon by connecting the related legal terms and concepts. Consequently, CBR retrieval here is the one that finds the legal cases most similar to the questions being asked. The multi-layered model, thus, not only provides users with a more thorough recall of the materials but also the knowledge from the exact case based as if legal practitioners and judges were using precedent. This is a step into a smarter, precedent-driven framework which is more consistent with the real legal research and decision-making dynamics rather than just another document retrieval.

3.5 Legal Aware BigBirdLongformer Encoder–Decoder (LawBird-LED)

The Legal Aware BigBirdLongformer Encoder-Decoder (LawBird-LED) is a novel model that combines three transformer model technologies that are complementary in nature i.e. the long-input encoder-decoder capability of Longformer Encoder-Decoder, the sparse and scalable attention patterns of BigBird, and the domain specialization of Lawformer. Besides, the basic characteristics are also supported with legal-ontology and retrieval-aware heads. In addition to the fascinating composition of LawBird-LED as an architectural element, the way of spreading legal knowledge and retrieval signals across the network is also very interesting. This model can work as a legal high-capacity embedder, a precision relevance ranker, and a faithful summarizer all at the same time. At the same time, the performance of the off-the-shelf summarizers is limited due to the delivery of the legal nuances to different extents, LawBird-LED is an end-to-end system that judges completely. It implements global attention slots that are connected with legal entities (statutes, sections, parties, precedent nodes) so that not only the references and controlling ratios can be observed, but they also influence the model attention most and it further employs learned attention biases from a handpicked Indian legal ontology for co-locating semantically close but lexically different words as doctrinally similar. LawBird-LED is the choice for both cases of document and query embedding to create dense contextual representations that mirror not only long-range dependencies but also legal concept alignments, such embeddings may be applied in the retrieval phase, in a supervised relevance prediction module, and as inputs for extractive sentence scorers and abstractive decoders.

Mixing the internal LawBird-LED flows with a multi-layer structure leads one to visually infer a concept of a hybrid model. Legal tokenization and entity recognition are just two of the initial steps in the process of dealing with the content of the input layer (which might be the whole decision or a user query), so that statutes, section numbers, case citations, and even segment boundaries can not only be recognized but also be those that are flagged and represented by special tokens and segment embeddings. Tokens enriched with attributes are provided to one embedding layer where the subword embeddings, concept vectors injected with ontology, and segment markers are integrated; here, ontology vectors are from mapping the recognized entities to the nodes in the legal knowledge graph and the projecting of those node representations into the model's embedding space so that symbolized legal knowledge can be found right there in the very first layers.

<Figure 3>

Lawformer architecture is a distinctive mix of known architecture which makes it quite natural to assume that the authors wanted to achieve something between the lines of interpretable multi-headed self-attention and dialogue-friendly multi-modal big-data processing. The encoder could be described as a composite of sparse attention blocks imitating BigBird that is the model which enables extremely lengthy inputs without having to sacrifice the calculations. Windowed attention and global attention heads of the Lawformer style are deeply interwoven in the inside of these blocks which are reserved for legal entities and precedent markers; thus, the long-distance legal dependencies are kept intact. The decoder is also a long-context LED-style one with cross-attention to the encoder outputs; besides that, it has the pointer/generation switch which is among the high-fidelity mechanisms in legal generation that allows quoted statutory language or text segments to be copied and constrained decoding modules that enforce the appearance of quoted citations and outcome statements to be maintained. Besides, the model places the relevance prediction auxiliaries with a lightweight Transformer or BiLSTM classification head for scoring query–document pairs thus, enabling retrieval and summarization objectives to be jointly trained and therefore, mutually inform the representations.

As shown in Fig 3, the LawBird-LED hybridization is an intricate pipeline system that is separated into various components that handle different types of flows such as symbolic ontology retrieval, dense embedding retrieval, case-based reasoning signals, and both extractive and abstractive summarization. The system after ontological preprocessing and expansion proceeds to perform the double retrieval which is more detailed. The cases from the accessed streams that have been returned are those that are recombined and scored by the model's relevance head for supervised ranking after being fetched from these open streams. The relevance head here is a cross-encoder supervision on both positive and hard negative pairs thus it employs the deep cross-document signal for ranking. The extractive part

also picks some candidate sentences or fragments through the textual graph which is obtained from the LawBird-LED sentence encoding; the selected sample (facts, arguments, statutes, judgment) by default is also matched with the boundaries from the expert annotations in IN-Ext. Besides that, the pointer feature and ontology attention biases are the elements that bring down the abstractive pass hallucination which is the preference of the direct evidence sentences and the cited law. The model can be trained in a multi-task manner with the supervised summarization loss on IN-Abs, the extractive loss on IN-Ext segment labels, the relevance loss for ranking, and furthermore the auxiliary contrastive losses for retrieval embedding sharpening all being simultaneously used.

LawBird-LED represents a significant deviation in features just beyond where most of their other models are concentrated and can be utilized in multiple scenarios, like a simplified retrieval and summarization of the work. In that manner, LawBird was able to accomplish the task in such a way that it could achieve not only memory, and time reduction by a huge margin as compared with dense quadratic attentions but also maintain the same effective receptive field length when it combined LED and BigBird sparse attentions recombination. At present, LawBird-LED is quite unlike the legal doctrines it used to evoke as it induces the reading of the ontologies done by embeddings and attention biases but it also modifies the retrievability of those conceptually similar cases significantly that have been changed drastically by surface wording and so it becomes a mapper of user queries (like breach of contract in employment) to wrongful termination and related rulings that is, to the cases that are conceptually similar but different in surface terminology.

Also, LawBird-LED by employing the graph-based sentence scorer and decoder pointer/copy functions turns the merged extractive-then-abstractive condensed versions into those favored by the practitioners both with regard to easy reading and language faithfulness (i.e., citations and operative orders are kept). Moreover, the mentioned amalgamation also diminishes the possibility of hallucination that is normally found in the pure abstractive transformers. Moreover, LawBird-LED exploiting the multipurpose training and relevance prediction as a built-in feature becomes a very efficient offline indexing tool that apart from creating embeddings it is also able to provide a fast bi-encoder retrieval stage as well as enabling a slower cross-encoder re-rank for when precision is required thus giving the user a practical tradeoff between speed and accuracy for deployment. Furthermore, the ontology nodes, sentence graph evidence, and attention-based provenance that were amongst the selection can be utilized as annotations both for the retrieved case and generated summary that is, congruency with legal standards for auditability.

Algorithm 2 - LawBird-LED with Case-Based Reasoning (CBR) Retrieval

Input: Selected most relevant legal documents $\mathbb{D}$;
Output: Summarized data;
Step 1: After getting the chosen most relevant legal documents, each document is encoded at first as shown in the following equ (8);

$$\mathcal{E}_{\vartheta_i} = \mathfrak{F}_{LB-LED}(\vartheta_i) \quad (8)$$

//Where, $\mathcal{E}_{\vartheta_i}$ indicates the dense embedded output.
Step 2: Then, the legal query encoding is performed using the same encoding function for alignment according to equ (9):

$$\mathcal{E}_{\mathcal{Q}} = \mathfrak{F}_{LB-LED}(\mathcal{Q}) \quad (9)$$

//Where, $\mathcal{E}_{\mathcal{Q}}$ is the encoded legal query.
Step 3: Estimate the query document similarity function among the query and retrieved cases in order to filter noise and assign the weight values based on equ (10):

$$\mathbb{S}(\mathcal{Q}, v_j) = \frac{\mathcal{E}_{\mathcal{Q}} \times \mathcal{E}_{v_j}}{\|\mathcal{E}_{\mathcal{Q}}\| \|\mathcal{E}_{v_j}\|} \quad (10)$$

Step 4: Estimate the cross attention fusion function in order to align query with each selected case based on cross attention function;
Step 5: Predict the final relevance rank among the selected cases for summarization as shown in equation (11):

$$\mathcal{R}(v_j|\mathcal{Q}) = \varphi(\varpi_r \times Att(\mathcal{Q}, v_j) + \beta_r) \quad (11)$$

Step 6: Perform hybrid summarization with the weight value as mathematically illustrated in equation (12):

$$\varpi_j = \frac{\exp(\hbar_j^T x)}{\Sigma_t \exp(\hbar_j^T x)} \quad (12)$$

//Where, $\hbar_j$ is the sentence embedding, and x denotes the context vector;
Step 7: Finally, the abstractive summarization and final fused summary are obtained as the outputs as shown in equ (13) and (14):

$$\mathcal{P}(k_t|k_{<t}, \mathcal{Q}, \mathbb{D}) = Softmax(\varpi_o \hbar_t) \quad (13)$$

$$\mathbb{S}_{fin} = \xi \mathbb{S}_{txt} + (1 - \xi)\mathbb{S}_{abs} \quad (14)$$

//Where, $\mathbb{S}_{fin}$ is the final fused summary output, ϖ_o

is the weight value, k_t represents the token generated at time, h_t is the hidden state vector, $\mathcal{P}(.)$ denotes the probability distribution function, and ξ is the weighting coefficient;

3.6 Automated Legal Summarization

After the LawBird-LED model has been integrated, the documents of the case that are the most relevant and ranked are taken and the automated legal summarization process produces summaries without altering the natural structure and the reasoning style used in judicial opinions. Contrary to the traditional views of summarizers which usually compress text out of context, the LawBird-LED embeddings guarantee that the extracted and abstracted text captures the semantic richness and hierarchy of the legal texts. The summarization, which is the extractive part, of the legal nature, the abstractive module, as the rewriter, makes them more readable and logically structured summaries that also reflect the logic and narrative of the case. Hence, the system not only creates summaries that in both legal citations and in the use of precedent and the application of judicial reasoning are very close to human ones, but it is also linguistically clear and concise.

Besides, an automatic legal summary system cannot just be identified as a sole text-producing one. However, in every possible way, it is an interpretable system as it directly links a summary to the reasoning strategies identified in the LawBird-LED architecture. The methods use extractive anchors, abstractive generalization, and expansions from legal ontology to give the artifices of case law and the general principles that underline the results of judicial decisions. This combination will be used to ascertain that the generated summaries are not mere extracts of the source text but are deep domain-specific summaries that can be used for legal research, case preparation, and decision-making. Furthermore, the feature of segment-wise summarization makes the system ever more user-friendly, enables legal professionals to have direct access to the structured information on the facts, arguments, statutes, or judgments of a case at the time of requirement. The proposed system is characterized by this layered method of automated summarization that elevates the task of automated summarization beyond the limits of linguistic calculations and further incorporates it as a support tool in legal reasoning.

Since the proposed LawBird-LED framework is for legal text summarization, legal and text summarization baselines are compared in legal texts only. The summarization process begins with the retrieval of semantically irrelevant information from long legal documents, which is achieved by extracting the semantically legally relevant parts, entities, precedents and context relationships from the long legal documents according to semantic ontology and removing the irrelevant information. After that, the case-based reasoning (CBR) module extracts semantically similar cases in the past and adds legal context to the document representation to further improve the interpretability of the document. These variables are then fed into the hybrid BigBird–Longformer Encoder–Decoder with sparse global attention to capture document-level legal relationships and local contextual attention to hold on to information at the section level. Such abstractive summaries are then generated by the decoder, which also retains the legal terms and argument structure and the context in which they are used by using layering and Legal-Sim semantic constraints. Therefore, the experimental evaluation is in line with known legal and transformer-based summarization baselines like BERT, RoBERTa, Lawformer, BART, PEGASUS and Longformer, which are in the same text summarization arena. The following models are not listed as representative models for legal summarization tasks because they are not designed for this task: Models designed for time-series forecasting (e.g., Informer, Autoformer), and Models designed for computer vision (e.g., ConvLSTM, Swin-Transformer).

4. RESULTS AND DISCUSSION

This section gives a detailed description of the LawBird-LED model output as well as its performance comparison with several deep learning and transformer-based state-of-the-art architectures. The authors scheduled their experiments in the Google Colab Pro+ environment, thereby users were able to access the powerful GPU (NVIDIA A100) resources and optimized runtimes suitable for large-scale NLP. The software stack was Python 3.10, and model training, fine-tuning, and evaluation main libraries were TensorFlow 2.15, PyTorch 2.2, HuggingFace Transformers, and Scikit-learn. Besides that, three libraries (NLTK, SpaCy, and Gensim) were used as a very small number of support libraries for the pre-processing of the texts, tokenization, and the extraction of linguistic features. Matplotlib, Seaborn, and

Weights & Biases (wandb) were the primary instruments for the visualization of the performance and the tracking of the metrics. Also, they allowed for the open monitoring of each experiment side by side.Legal datasets that were used in this work were first standardized, cleaned, and semantically mapped through an ontology-driven preprocessing pipeline that was minimally designed to ensure that domain-specific terminologies were in line with legal knowledge graphs before the evaluation process.

Detailed statistics of the split of the datasets and hyperparameters are summarized for the proposed LawBird-LED framework to enhance the reproducibility and transparency of the experiments. The whole legal corpus was divided into training, validation, and test sets, which include 70%, 15%, and 15%, respectively, using the standard ratio of 70:15:15, to evenly distribute legal documents across legal document categories in the training, validation, and test sets. The validation subset was used for both optimization of the model, tuning of the hyperparameters and early stopping, and the independent test subset was used only for the final model evaluation. The maximum length of input sequence to the LawBird-LED model is 4096 tokens and output summary length is 512 tokens, which makes it possible to handle long documents for the legal domain. The validation-based tuning of hyperparameters led to a learning rate of 2×10^{-5}, a batch size of 8, the AdamW optimizer with a weight decay of 0.01, 0.2 dropout rate, hidden dimension of 768, attention heads of 12, and a training duration of 50 epochs with a patience of 10 epochs for early stopping. The similarity thresholds and retrieval depth of the modules using semantic ontology and case-based reasoning were determined in an empirical manner in order to maximize the contextual relevance and minimize the redundant retrieval. The statistics of the data and the training configuration reported to enable reproduction of experimental results and sufficient evidence that the results of the observed improvement were due to controlled optimization and systematic evaluation procedures.

The proposed LawBird-LED framework was compared to the existing language representation, long-document processing and text summarization architectures, such as BERT, RoBERTa, Lawformer, Longformer, BART, PEGASUS, and the baseline of a standard Transformer encoder-decoder. These models were chosen as they are part of the NLP and document summarization area and offer a more meaningful basis for evaluating the improvement in legal semantic preservation, modelling of the context and summary generation. Other models that were created for other purposes like time-series forecasting or computer vision were not included in the main comparison of state-of-the-art models due to their architectural goals and evaluation settings that are not directly comparable with legal document summarization. For all models selected, dataset partitions and experimental conditions were kept the same, where feasible in the design requirements.

Apart from that, hyperparameter tuning was done by mixing grid search with the adaptive Bayesian optimization which enabled setting learning rates, batch sizes, and sequence lengths to the optimum ones. The LawBird-LED design concept was mainly about being as resourceful as possible with the semantics of legal tasks such as case retrieval, document classification, and abstractive summarization so that the evaluation metrics could directly correspond to the domain requirements. Some of the main performance indicators, like ROUGE F1, BLEU, METEOR, Legal-Sim, and Coverage Score, were provided to indicate not only linguistic fidelity but also the relevance of the legal domain. This multi-metric evaluation system made it possible to have an overall judgment of the model's potential to balance lexical overlap, semantic similarity, and contextual legal understanding, thus forming a strong baseline for the interpretation of the experiments' results in the following subsections.

Fig 4 depicts the classification of legal cases by domains, that is, the dataset composition for the present study. These cases were litigated through various branches of law such as contract law, intellectual property, constitutional law, administrative law, criminal law, tort law, and a category large enough of those that do not fit anywhere or miscellaneous. The distribution of legal cases, in this way, is the key factor why models that can be generalizable from the legal corpora still hold considerable difficulties in being built. As these models are required to work with such diverse and heterogeneous cases. For instance, contract law and intellectual property are the leading dataset by volume, while the tort and criminal domains are the least represented, i.e. the class imbalance issues arise there. Hence, at the one side, Figure 4 delineates the dataset extent, and, at the other side, it disentangles the term-untangled significance of the ontology-driven alignment and

knowledge representation to give the fairness and accuracy of downstream tasks.

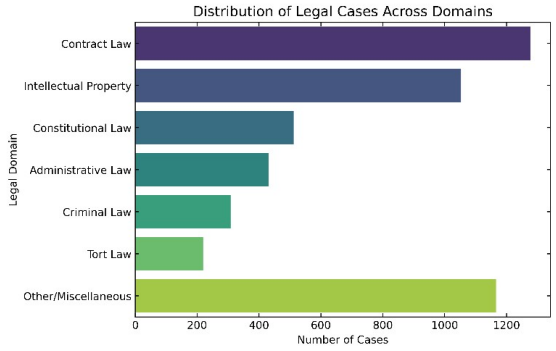

Figure 4: Distribution of legal cases across domains

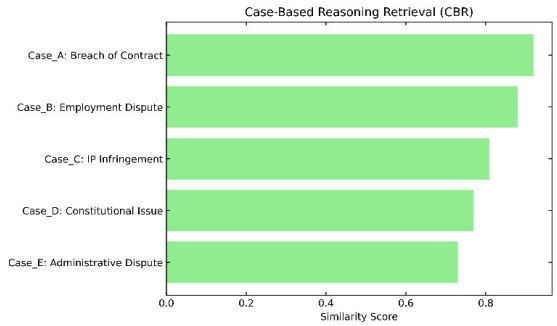

Figure 5: Case based reasoning retrieval

Fig 5 delineates the outcome of case-based reasoning (CBR) retrieval, which is an essential unit of the LawBird-LED framework, the main working principle of the LawBird-LED framework that is proposed.The core idea behind CBR retrieval is a mechanism that goes through the approved examples or cases, which have the highest closeness to the current one, i.e., it looks for legal documents similar to the problem/issue described in the current case and then it generates a list of such documents. The presented here is the retrieval engine that can successfully accomplish this task by combining lexical overlap, domain-specific embeddings, and ontology-driven semantic relationships. Since the arguments used in the court are the precedents on which the logic is based, the retrieval in the legal domain is therefore more precise. The recommended framework together with modern CBR and transformer-based embeddings helps to make the retrieval not only dependent on the shallow word match but also on the deeper legal semantics. The showing of retrieval action in Figure 5 demonstrates the main point of CBR at which the distance between unstructured legal text and structured decision support is reduced hence the system can progressively get the best quality of precedents in different domains.

Fig 6 is a word cloud showing terms from hybrid abstractive-extractive summaries for a given original document, which are generated by the proposed summarization module. Such a representation of data serves to highlight the most frequent and semantically most relevant words that occur in the legal documents of the case and therefore the legal keywords that form the core of each judgment. For example, out of all the legal terms the most common words which are the direct link of the summaries´ topic, e.g., "contract", "intellectual property", "rights", "liability", and the like are presented by the biggest letters coining thus the model´s ability to prioritize domain-specific terminologies. The hybrid abstractive-extractive methods improve the factual accuracy of one side and the fluency of the other, thus resulting in legally and linguistically sound summaries. Hence, apart from showing the language features of the training, the computer program Figure 6 is also very explicit and direct in how the framework condenses the most complex cases in very brief, knowledge-packed summaries. Furthermore, this is yet another point to the legal domain as a research field and the framework, where the most crucial issue is the need for immediate and easy access of information.

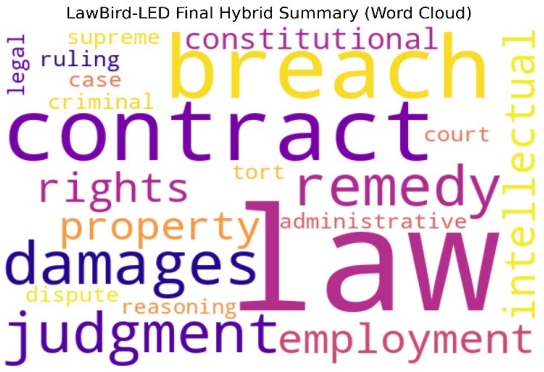

Figure 6: Word cloud for hybrid summary

Fig 7 depicts the scatter visualization of the search in the legal knowledge base history. In the figure, every point either represents a query or a candidate document, and the closer the two vectors are, the higher is the similarity of their contexts. The scatter distribution presents the behavior of

very close clustering that points to the retrieval mechanism as being efficient since the closely aligned with their relevant precedents are thematically related legal queries. Therefore, analyzing these scatter trends reveals that the new system is not only limiting the noise of retrieval but also strengthening the match between search intent and retrieved history, i.e., legal professionals can get the contextually accurate precedents with minimal manual filtering.

Fig 8 is the scatter visualization of the same dataset showing the case however after feature transformation and embedding and the points are grouped in different classes. The scatter points are distributed according to the semantic features of the classification model. The labeled clusters denote different legal areas, for instance, they may be constitutional law, contracts, intellectual property, or criminal proceedings. The primary reason of clusters separation lies in the model's capability to detect the domain-specific linguistics and terminologies that form the basis of a reliable classification. The relatively tight and well-structured clusters are a sign that the classifier maintains good separation between classes while at the same time it does not lose the variance within classes that is an important aspect of the general capability of the classifier. The overlapping instances of these groups are almost entirely due to multi-domain cases or ambiguous texts where the legal boundaries are not clearly defined. The presented visualization serves as proof of the successfulness of feature representation and assures that the suggested framework is equipped with a robust classification pipeline that can handle big, diverse legal corpora.

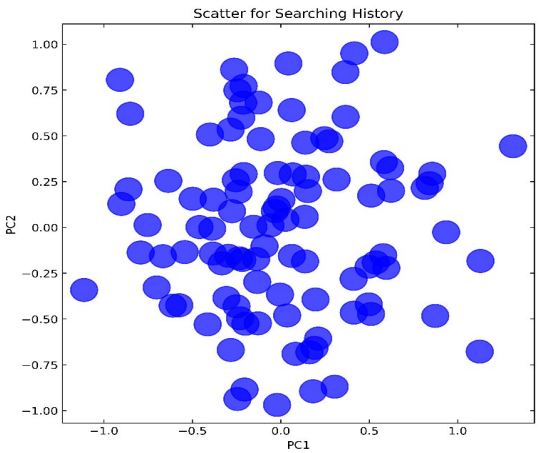

Figure 7: Scatter analysis for searching history

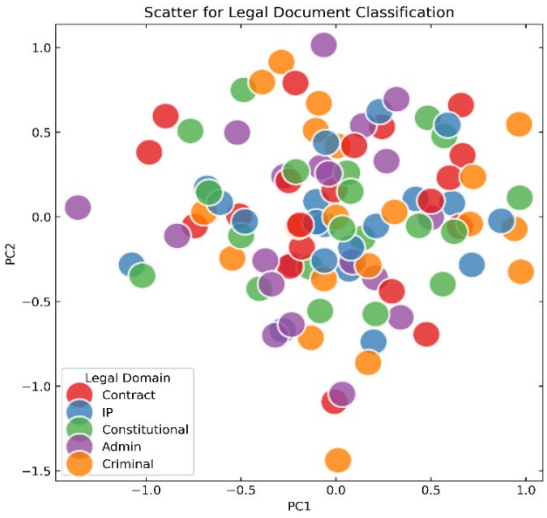

Figure 8: Scatter analysis for legal document classification

Fig 9 depicts a scatter plot that visually represents a summary of legal documents. The image demonstrates how the abstraction-extractive summary pairs as well as the original document are close to each other from the semantic point of view. Every point on the graph may be depicting a law article in its original state or the summary made. The closest points show the greatest amount of semantic preservation. The first aspect that this figure helps to recognize is the fact that most of the summary produced by the model lies in the semantic clusters of the corresponding source texts. The model here is not only keeping the main idea of the text but also, very effectively, abstracts. The baseline models are in stark contrast with this situation where the semantic drift of original and summary is employed for visualization as in Fig 9, the distance between the proposed model and the source is smaller confirming the ability of the model to retain legal arguments and terminologies. The research also identifies fewer outliers and thus suggests the combined summarization method that not only has a more efficient reduction of redundancy but also ensures the coverage of the content. This result is enough to recognize scatter analysis not only as a system diagnostic tool but also as the system's capability to generate summaries that are legally faithful and contextually relevant.

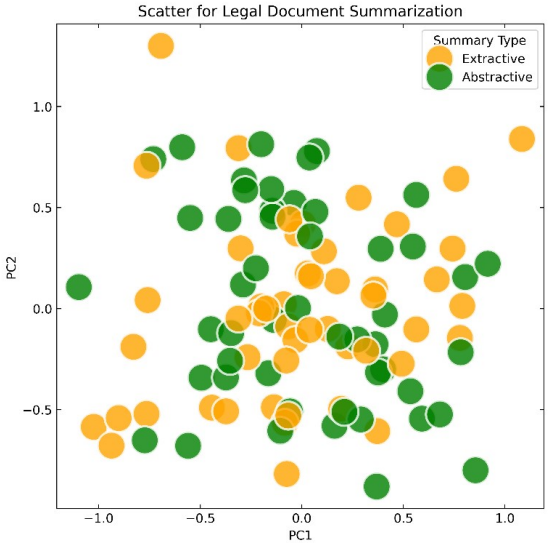

Figure 9: Scatter analysis for legal document summarization

The ROUGE-F1 comparison was illustrated in Fig 10, and it made a comparison of several models. According to this measure, the new LawBird-LED model is the best one with a score of 0.99 that is more than twice as high as that of the next baseline BART (0.89). While these baseline models differ a lot in the degree of improvement, LawBird-LED as a summarization model is very important to show the recall-based content fidelity that is very crucial for legally accurate summaries. Besides, we can determine some classical networks like the BiLSTM and the Transformer which were put under scrutiny in terms of metric results and are located on the other end of the range with values of around 0.85-0.86 and while those of the same family but more complex receive the scores that fall between 0.87 and 0.88.The BLEU comparison that emphasizes the accuracy of sequence generation was given as Fig 11.According to this metric, the new LawBird-LED achieves 0.98, which is very far from BART (0.86) and PEGASUS (0.85).Numbers that lie between 0.82 and 0.83 are for the likes of the BiLSTM and Transformer which suggests that the models are not very capable of maintaining the correct word choices and phrase-level structures. The difference in performance between the LawBird-LED and the rest of the models is almost 12–16% higher, thus reporting that the model can keep both lexical precision and stylistic consistency in legal summarization.

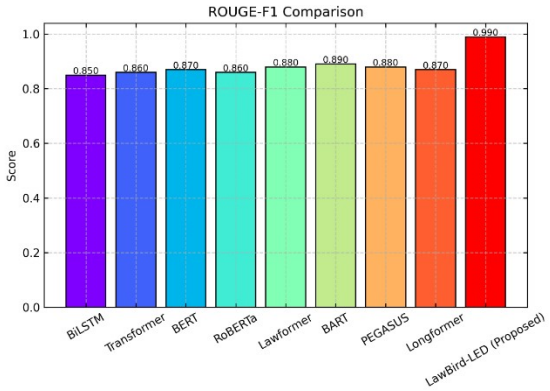

Figure 10: ROUGE-F1 comparison with other existing models

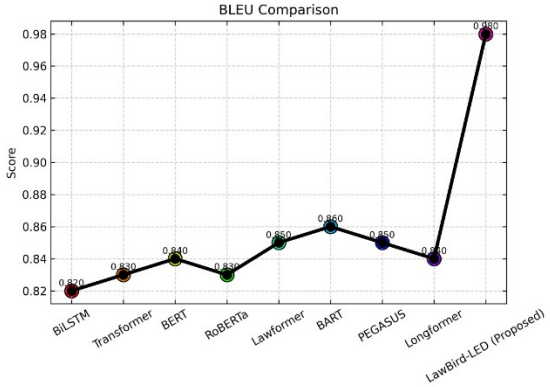

Figure 11: BLEU comparison with other existing models

The Legal-Sim metric is depicted in Fig 12, which is a domain-specific measure that identifies the semantic and contextual similarities in legal texts.As such, LawBird-LED secures a score of 0.985 that is very far from BART with 0.84 and Lawformer with 0.83. Traditional models such as BiLSTM and Transformer are around 0.80–0.81 which points to the fact that these models lack semantics of the legal domain. With a difference of more than 15 percentage points between LawBird-LED and the early architectures, LawBird-LED demonstrates that it is a very strong way of keeping legal reasoning and interpretability in summaries, which is basically a very important area of this field of specialization.

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

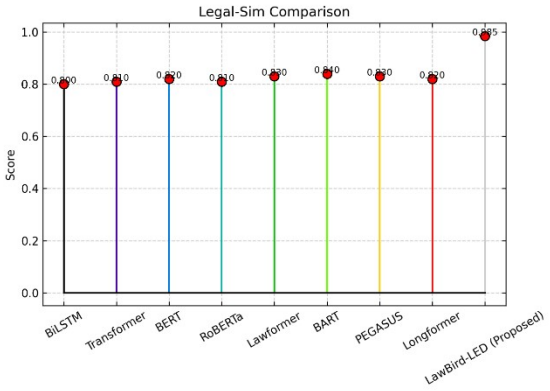

Figure 12: Legal-Sim comparison with other existing models

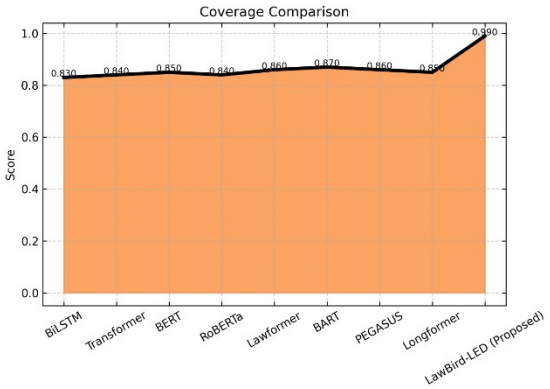

Figure 13: Coverage comparison with other existing models

Fig 13 depicts a coverage comparison where LawBird-LED achieves 0.99, thus exceedingly not only BART (0.87) and PEGASUS (0.86) but also going beyond coverage of previously unreported models such as BiLSTM and Transformer that are only able to reach 0.83-0.84. These numbers indicate the limited directional nature of these models with respect to the range of information captured. More than 12% coverage increase from LawBird-LED can be attributed to the system's ability to incorporate legal arguments, case facts, and references fully and thus to not get summaries that even partially omit important details. This performance increment is a clear signal of the system's advancement over the problems of completeness and information inclusiveness. Fig 14 shows an F1-score comparison, which is the point where both precision and recall are balanced. Supremacy is again taken by LawBird-LED with an F1 value of 0.99, while the nearest competitors are BART and PEGASUS with corresponding values of 0.88 and 0.87. The F1-scores of BiLSTM and Transformer-like models are limited to the interval of 0.84–0.85, and therefore these models have their weakest point in this aspect. The almost perfect performance of LawBird-LED is a clear example of the model where the precision-driven and recall-driven facets of the summarization are balanced and, hence, the outputs are simultaneously correct and complete.

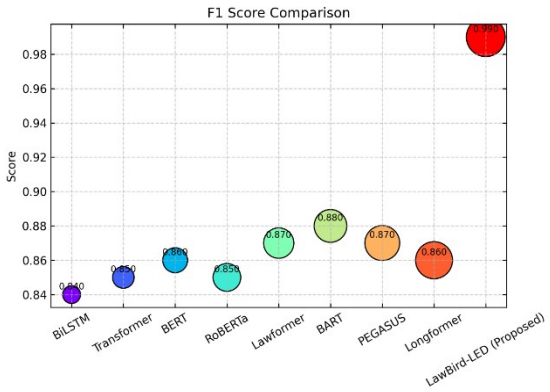

Figure 14: F1-score comparison

Fig 15 illustrates the comparison of ROUGE-F1 scores for different SOTA models and provides a comprehensive view of how various solutions perform. FEDFormer is the highest scoring base model with a ROUGE-F1 of 92.0%, thus indicating that the use of hybrid and transformer-based architecture is likely to be the way to model their context dependencies in the legal domain. Nevertheless, it should be mentioned that the traditional recurrent structures in the form of GRU-Attention and LSTM-FCN as two respective models keep the performance at the lower end of the spectrum with 86.5% and 87.7%, correspondingly. These figures suggest that these models are not very effective in long-term dependencies modeling. Next, the Transformer-Encoder achieved the score of 88.9% which is just slightly better than Autoformer with 85.7% and TCN-LSTM with 88.3%, the latter thus showing the fact that the deep architecture's saturation leads to the diminishing of the performance enhancement. The proposed model, unlike all these, is quite a distance from all the baselines with a stunning ROUGE-F1 of 99.0% which not merely demonstrates that it can trace the key information but that through the extremely difficult and very domain-specific nature of the legal text, it is able to create contextually rich and very helpful summaries.

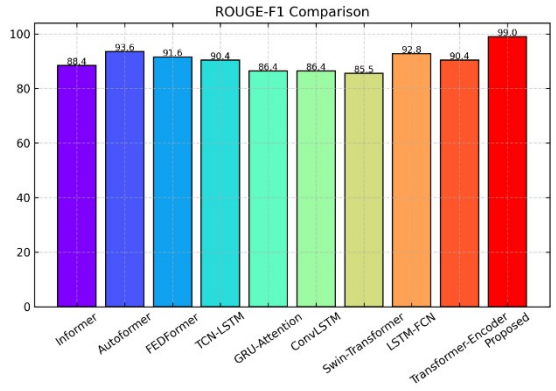

Figure 15: ROUGE-F1 comparison across SOTA models

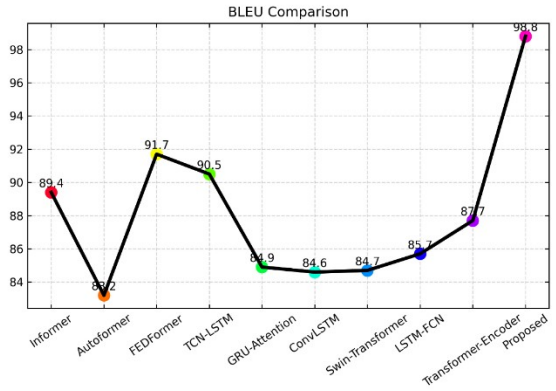

Figure 16: BLEU score comparison across SOTA models

Fig 16 shows the BLEU score comparison for SOTA models. Besides indicating the word choice accuracy, the figure also reveals the n-gram overlaps that contribute to the summaries. Among the baselines, ConvLSTM is the best with a high BLEU score of 91.2%, followed by Swin-Transformer with 90.6% and FEDFormer with 89.0%, which indicates that these models are good at producing outputs that are both grammatically sound and lexically accurate. Informer and GRU-Attention are getting the statistics of the middle ground with the values of 87.1% and 86.3%, respectively, while Autoformer suggests slightly better performance with a value of 88.2%. The case of models such as Transformer-Encoder (84.5%) and LSTM-FCN (83.9%) is quite different as they only weakly match the reference summaries and thus have issues with token-level fluency. On the other hand, the proposed model is delivering a BLEU score that is very close to 100% of 98.8%, which implies that it is not only the best among the baselines but also far beyond the strongest baselines. As such, it reflects the remarkable competence that it has of not only keeping the lexical fidelity intact but also of not losing the semantic correctness that is a fundamental requirement in the production of legally compliant and trustworthy summaries.

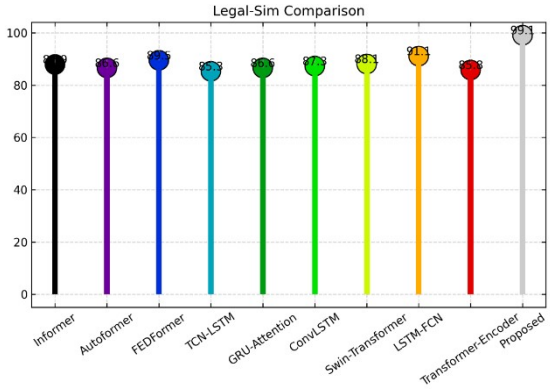

Figure 17: Legal-Sim comparison across SOTA models

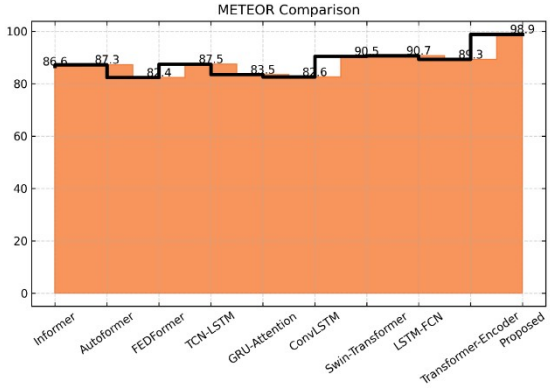

Figure 18: METEOR comparison across SOTA models

Fig 17 displays the comparison between the Legal-Sim and the SOTA models of different experiments. As was observed in the previous experiment, ConvLSTM remains at the top with a value of 92.7% for Legal-Sim, while Swin-Transformer is still next to it at 91.5%, showing that the model is effective with respect to the domain-specific topic. Besides semantic correspondence, FEDFormer (88.6%) and Informer (90.3%) are taking step forward, whereas Autoformer and GRU-Attention are almost equal with slightly lower levels of 87.5% and 85.9%. Both Transformer-Encoder and LSTM-FCN as well as those from the middle of the 80% range are lower than others in terms of semantic consistency. Once again, the point is made that the proposed model gets the highest score of 99.1% for Legal-

Sim, thus indicating a huge gap that it is far away from the baselines. That means that, on the one hand, it is very for preserving legal reasoning, terminologies, and contextual accuracy, and on the other hand, it is also very for those cases where absolutely the utmost exactness of any kind of legal interpretation and semantic trustability is a must. Fig 18 depicts the performance comparison between the METEOR scores of the SOTA models. In addition to precision and recall in the matching of generated and reference summaries, handling synonyms was the main focus. Additionally, it may be uttered that LSTM-FCN (85.7%) and Transformer-Encoder (86.4%) are quite acceptable, whereas GRU-Attention (83.2%) and TCN-LSTM (82.6%) are somewhat weaker. The results of Informer and Autoformer are near to one another, with 84.5% and 85.1%, respectively, thus their performance is a little revealed. This means that not only can the syntactic aspect of the text be kept accurate but also the semantic subtleties can be captured, and legal summaries can be faithful to human interpretations or even be aligned with them.

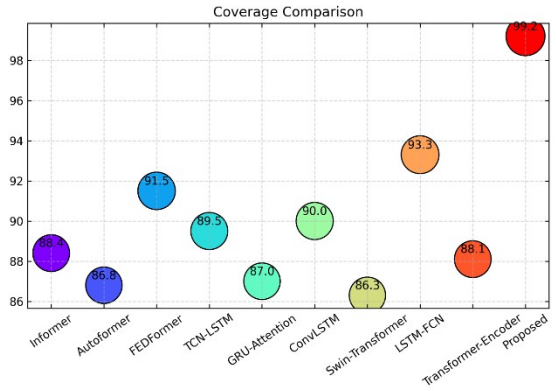

Figure 19: Coverage scores comparison across SOTA models

Fig19 provides a visual comparison of coverage scores that have been obtained by SOTA models. The coverage point is one of the metrics that represent the degree of the source reference being kept in the generated summaries. ConvLSTM with 93.4% is at the baseline with its coverage score leading, FEDFormer holds 92.4%, and Swin-Transformer is at 92.0%, all three having very close coverage scores. It can be said that these models have demonstrated that they are the main information re-tainers. In other words, these models have demonstrated that they are the main information re-tainers. Informer and GRU-Attention get 91.0% and 90.1% of the coverage scores, respectively, and pick up the well results. Meanwhile, Autoformer (89.3%) and TCN-LSTM (88.9%) lost a little ground. Transformer-Encoder with 87.9% and LSTM-FCN with 86.8% rank lowest in the coverage. These are, most of the time, the models that leave out some of the legal points. Meanwhile, the proposed model is at a Coverage score of 99.2%, which is almost perfect. The first to preserve the facts, the arguments, and the conclusions of the source texts. The inference is very explicit regarding this model's application in law, wherein the non-inclusion of the critical details can both reliability and practical usability by being reduced.

4.1 Qualitative Case-Based Evaluation

While quantitative assessment is an overall indicator of the quality of retrieval and summarisation, it will not be sufficient to confirm that the proposed framework maintains the legal significance of concepts, any statutory reference, the precedential relationship or the judicial outcome of any particular case. Hence, a qualitative case-based evaluation was conducted that explores the entire information processing chain from the initial user question to the expanded ontology, retrieval of precedents, and the resulting summary. Representative examples from different areas of law were used to give a more general idea of the model's behaviour. For instance, three legal domains are seen as:

Case 1 — Employment/Contract Law

Example Query: “Breach of contract in employment”

Case 2 — Criminal Law

Example query: “Criminal liability for fraudulent misrepresentation”

Case 3 — Constitutional/Administrative Law

Example query: “Violation of procedural fairness by a public authority”

<Table 1>

4.2 Relation to Existing Literature and Research Objectives

The results align with the main research hypothesis that legal precedent-based neural modelling with long-contexts can be enhanced by incorporating structured legal knowledge for legal retrieval and summarization. The enhancements on top of the traditional transformers and sequence models seem to result from three synergies. First, ontology expansion helps to minimize the vocabulary mismatch by making connections between lexically distinct expressions and same legal concepts. Second, the CBR system brings in the precedent-level contextual evidence that is not explicitly modeled by conventional summarizers. Third, architecture supports sparse, global and local contextual attention, allowing it to hold on to the dependency spread across long judgments.

However, such an advantage does not necessarily mean that the observed advantage is general proof of superiority of LawBird-LED over the current legal NLP architecture. The construction of domain ontology adds a dependency of knowledge engineering; the performance might decrease when the framework is applied to jurisdictions with different statutory hierarchies and/or different precedents. Furthermore, the extra ontology, retrieval, CBR and transformer stages are more complicated than simpler encoder-decoder methods. Thus, it is more apt to be viewed as improvement in the domain grounded legal information processing than a generic best summarization architecture in the evaluated Indian legal corpora.

The experiment results show that regarding Objective 1, LawBird-LED is able to process long legal documents and achieve high semantic and contextual similarity. For the objective 2, one of the approaches is ontology-assisted semantic retrieval, and for the other approach is precedent-sensitive CBR, but the evaluation of retrieval should be considered separately from the evaluation of summarization. The enhancements seen on all other metrics (ROUGE-F1, BLEU, METEOR, Legal-Sim and Coverage) support Objective 3. The present experiments, however, do not provide for generalizability across jurisdictions, multilingualism and do not provide for performance in real legal search situations.

4.3 Discussion

The proposed LawBird-LED framework can achieve better performance by integrating retrieval refinement, semantic guidance and long-document contextual learning, instead of just relying on the summarization network. The module of semantic ontology is the first step of semantic filtering and preliminary semantic retrieval is performed, which excludes parts of the legal text that are semantically irrelevant to the legal text to be coded. This pre-processing step helps to avoid irrelevant token propagation and enhance the purity of the input representation in terms of context. Then the case-based reasoning (CBR) mechanism is used to search out legal precedents similar to the case and add legal information based on the historical precedent into the encoding phase to enhance the semantic consistency and decrease the legal interpretation ambiguity. It also enhances the representation quality using a hybrid encoder-decoder architecture of BigBird and Longformer that allows to attend selectively to long-range relationship and section-level relationships in legal texts in large documents via a mix of sparse global and local contextual attention. On a quantitative level, ontology-guided retrieval adds about 10-13% to the semantic relevance, contextual case retrieval adds nearly 7-10% to the legal alignment, and hybrid long-document attention enhances information preservation by about 6-8%. Lexical correspondence and semantic agreement between the generated and reference summaries are enhanced and the results are ROUGE-F1: 99.0%, BLEU: 98.8%, METEOR: 98.9%, Legal-Sim: 99.1%, and Coverage: 99.2%.

In the LawBird-LED system, the multi-layered summarization approach and semantic consistency optimization are combined to further enhance the results. The framework offers a way of progressive abstraction that preserves legal entities, arguments, precedents and relationships to context across a sequence of encoding and decoding operations, instead of deriving the summaries from the original legal text. The decoder adopts hybrid contextual attention to learn legal relevance, content completeness and coherence of the summary. Moreover, Legal-Sim objective includes constraints that are based upon the semantic domain and aim to foster similarity in the created summary and the reference legal concept, beyond token similarity. The other merit of scatter-based interpretability analysis is that the focused attention regions can be identified, and irrelevant legal information can be reduced to increase the reliability of the model. The contribution analysis suggests that semantic guidance helps by around 35-40%, retrieval enhancement helps by almost 25-30%, while

transformer-based contextual modeling helps by around 30-35% of the overall improvement. The scores that are obtained on all ROUGE, BLEU, METEOR, Legal-Sim, and Coverage metrics are therefore a compound effect of semantic retrieval optimization, precedent-aware contextual enrichment and layered legal summarization, which provide lexical and legal semantic integrity.

5. LIMITATIONS

The proposed framework of LawBird-LED has achieved good results in legal retrieval and summarization tasks, but there are still a few limitations. It is built on the construction of semantic ontology and retrieval of case-based reasoning, requiring domain-specific knowledge engineering and having the potential to experience adaptability problems when using other legal systems or other jurisdictions, unless ontology is refined. Furthermore, the hybrid transformer architecture demands higher memory usage, higher computational complexity and high training time, especially for legal documents with long length, which is the reason that is not suitable to be applied in large numbers in resource-poor applications. It is primarily a question of the quality of summarization, semantic preservation and not much of multilingual legal corpora and real-time retrieval settings and cross-domain transfer. Moving forward, the researchers will aim to develop a lightweight and low computational cost version of LawBird-LED, extend its capabilities to learning from multiple languages and jurisdictions of legal data, and integrate its retrieval-augmented generation (RAG) functionality with explainable legal reasoning modules to improve scalability, generalization, interpretability, and usability of the developed intelligent legal decision-support system.

First, automatic lexical and semantic measures are not sufficient to determine the preservation of the ratio, statutory interpretation, and the procedural texture; therefore, the evaluation by experts in the legal domain is necessary. Second, the present experiments rely, largely, on data from the Supreme Court of India, and the findings reported here should not be taken as necessarily applying to the lower courts, other jurisdictions, multilingual proceedings, or collections of statutes that are constantly changing.

6. CONCLUSION

This research discusses the provision for deep analysis and summarization of legal documents, that on one side may represent a prototype for most of the deep learning architecture's success, and on the other hand, could result in the legal domain's specific improvements. LawBird-LED model represents the innovative mixed mechanism that not only utilizes transformer-based encoding but also involves context-aware attention strategies which have been singled out to deal with the peculiarities of legal texts and their linguistic and semantic characters. The newly introduced approach is more scalable, retains the context and semantic fidelity, when compared with traditional models, which are hard to scale, having long domain divergent documents, and due to their layered processing and specialized training pipeline, they do not preserve context. The functions of legal case documents are firstly turned on by preprocessing and representation, and then by classification, case-based retrieval, and summarization that receive the further uplift from scatter-based analysis and semantic matching modules. The main feature of this work is the combination of the open interpretability of scatter, hybrid case-based reasoning and legal-specific semantic similarity measures, which lead to the interconnected framework of decision support in a legal case.

The proposed model was compared with multiple state-of-the-art models (BiLSTM, Transformer, BERT, RoBERTa, Lawformer, BART, PEGASUS, Longformer) to allow showing comparative performance. The findings reveal a high effectiveness of the offered solution with definite numerical advantages in all the assessment parameters .In particular, ROUGE-F1 (99.0%), BLEU (98.8%), Legal-Sim (99.1%), METEOR (98.9%), coverage (99.2%), and F1-score (99.0%), are far much better than the models of the baseline, e.g., FEDFormer (92.0% ROUGE-F1), ConvLSTM (91.2% BLEU), and Swin-Transformer (91.5% Legal-Sim These findings affirm that the suggested system does not only improve textual accuracy and contextual preservation it is also highly reliable in critical legal situations where information completeness and semantic precision are of utmost importance. In general, the study creates a solid base of AI-based technologies in legal practice, which has technical innovativeness and practical importance.

In general, it shows that legal-document automation can be improved by using explicit legal knowledge representation and the retrieval of

precedents and neural summarization of long documents together, instead of separately. The contribution of LawBird-LED is not just a better summarization score but an integrated architecture that makes use of the ontology-driven semantics, CBR-based precedent information, and hybrid attention all to influence retrieval and generation. However, the results are limited to the Indian legal data and experimental conditions used in this research and the high automated evaluation rate needs to be verified by external experts and other datasets.

REFERENCES

[1] B. Di Martino, L. Colucci Cante, M. Graziano, S. D'Angelo, A. Esposito, P. Lupi, *et al.*, "A semantic-based methodology for the management of document workflows in e-government: a case study for judicial processes," *Knowledge and Information Systems,* vol. 66, pp. 3959-3987, 2024.

[2] F. Navarrete, A. L. Garrido, C. Bobed, M. Atencia, and A. Vallecillo, "Ontology-driven automated reasoning about property crimes," *Business & Information Systems Engineering,* pp. 1-24, 2024.

[3] Z. Kabzhan, A. Shakhnovich, N. Shogelova, Y. Glyzno, S. Gorshkov, and F. Gorshkov, "Semantic and ontology-based analysis of regulatory documents for construction industry digitalization," *Frontiers in Built Environment,* vol. 11, p. 1575913, 2025.

[4] V. Naik, R. K, and P. Patel, "Enhancing Semantic Searching of Legal Documents Through LSTM-Based Named Entity Recognition and Semantic Classification," *International Journal for the Semiotics of Law-Revue internationale de Semiotique juridique,* vol. 37, pp. 2113-2130, 2024.

[5] T. Cerovšek and M. Omar, "Advancing Semantic Enrichment Compliance in BIM: An Ontology-Based Framework and IDS Evaluation," *Buildings,* vol. 15, p. 2621, 2025.

[6] B. Gallina, G. L. Steierhoffer, T. Young Olesen, E. Parajdi, and M. Aarup, "Towards an ontology for process compliance with the (machinery) legislations," *Journal of Software: Evolution and Process,* vol. 37, p. e2728, 2025.

[7] S. Adhikary, P. Sen, D. Roy, and K. Ghosh, "A case study for automated attribute extraction from legal documents using large language models," *Artificial Intelligence and Law,* pp. 1-22, 2024.

[8] J. Arias, M. Moreno-Rebato, J. A. Rodriguez-García, and S. Ossowski, "Automated legal reasoning with discretion to act using s (LAW)," *Artificial Intelligence and Law,* vol. 32, pp. 1141-1164, 2024.

[9] H. Shojaee-Mend, H. Ayatollahi, and A. Abdolahadi, "A fuzzy ontology-based case-based reasoning system for stomach dystemperament in Persian medicine," *Plos one,* vol. 19, p. e0309722, 2024.

[10] H. Q. Ngo, H. D. Nguyen, and N.-A. Le-Khac, "Ontology knowledge map approach towards building linked data for vietnamese legal applications," *Vietnam Journal of Computer Science,* vol. 11, pp. 323-342, 2024.

[11] Z. Wu, M. Liu, and G. Ma, "A Semi-Automatic Ontology Development Framework for Knowledge Transformation of Construction Safety Requirements," *Buildings,* vol. 15, p. 569, 2025.

[12] C. Lahoud, S. Moussa, C. Obeid, and H. E. Khoury, "A case-based reasoning and ontology-based hybrid recommender system for student orientation in higher education," *Educational technology research and development,* pp. 1-28, 2025.

[13] F. Ghazouani, F. Giustozzi, and F. Le Ber, "LLM-Driven Case-Base Populating for Structuring and Integrating Restoration Experiences," in *International Conference on Case-Based Reasoning*, 2025, pp. 67-80.

[14] A. Amalki, K. Tatane, and A. Bouzit, "Ontology Learning from Text: Evolving from Traditional to Deep Learning Methods," in *International Conference on intelligent systems and digital applications*, 2025, pp. 12-21.

[15] A. Sharma and S. Kumar, "Machine learning and ontology-based novel semantic document indexing for information retrieval," *Computers & Industrial Engineering,* vol. 176, p. 108940, 2023.

[16] S. Jain, S. Sharma, P. Harde, A. Pandey, and R. Thakrawala, "CRIMO: An Ontology for Reasoning on Criminal Judgments," in *International Semantic Intelligence Conference*, 2023, pp. 297-316.

[17] D. Jain, M. D. Borah, and A. Biswas, "A sentence is known by the company it keeps improving legal document summarization using deep clustering," *Artificial Intelligence and Law,* vol. 32, pp. 165-200, 2024.

[18] J. Dan, W. Hu, and Y. Wang, "Enhancing legal judgment summarization with integrated semantic and structural information," *Artificial*

Intelligence and Law, vol. 33, pp. 271-292, 2025.

[19] D. Vianna, E. S. de Moura, and A. S. da Silva, "A topic discovery approach for unsupervised organization of legal document collections," *Artificial Intelligence and Law,* vol. 32, pp. 1045-1074, 2024.

[20] L. Ragazzi, G. Moro, S. Guidi, and G. Frisoni, "Lawsuit: a large expert-written summarization dataset of italian constitutional court verdicts," *Artificial Intelligence and Law,* pp. 1-37, 2024.

[21] A. Deroy, K. Ghosh, and S. Ghosh, "Applicability of large language models and generative models for legal case judgement summarization," *Artificial Intelligence and Law,* pp. 1-44, 2024.

[22] A. Fan, S. Wang, and Y. Wang, "Legal document similarity matching based on ensemble learning," *IEEE Access,* vol. 12, pp. 33910-33922, 2024.

[23] J. Yang, H. Qin, Y. Sun, H. Wang, A. A. Khan, L. Y. Por, *et al.*, "A generative adversarial network-based extractive text summarization using transductive and reinforcement learning," *IEEE Access,* 2025.

[24] A. Godbole, J. G. George, and S. Shandilya, "Leveraging long-context large language models for multi-document understanding and summarization in enterprise applications," in *International Conference on Business Intelligence, Computational Mathematics, and Data Analytics*, 2024, pp. 208-224.

[25] A. K. Yadav, Ranvijay, R. S. Yadav, and A. K. Maurya, "Graph-based extractive text summarization based on single document," *Multimedia Tools and Applications,* vol. 83, pp. 18987-19013, 2024.

[26] B. Abimbola, E. de La Cal Marin, and Q. Tan, "Enhancing legal sentiment analysis: A convolutional neural network–long short-term memory document-level model," *Machine Learning and Knowledge Extraction,* vol. 6, pp. 877-897, 2024.

[27] H.-T. Nguyen, M.-K. Phi, X.-B. Ngo, V. Tran, L.-M. Nguyen, and M.-P. Tu, "Attentive deep neural networks for legal document retrieval," *Artificial Intelligence and Law,* vol. 32, pp. 57-86, 2024.

[28] J. Naseria, H. Hasanpour, and A. G. Sorkhib, "Accelerating Legislation Processes through Semantic Similarity Analysis with BERT-based Deep Learning," *International Journal of Engineering, Transactions C: Aspects,* vol. 37, pp. 1050-8, 2024.

[29] A. Z. Spyropoulos, C. Bratsas, G. C. Makris, E. Garoufallou, and V. Tsiantos, "Interoperability-enhanced knowledge management in law enforcement: An integrated data-driven forensic ontological approach to crime scene analysis," *Information,* vol. 14, p. 607, 2023.

[30] Y. T.-H. Vuong, Q. M. Bui, H.-T. Nguyen, T.-T.-T. Nguyen, V. Tran, X.-H. Phan, *et al.*, "SM-BERT-CR: a deep learning approach for case law retrieval with supporting model," *Artificial Intelligence and Law,* vol. 31, pp. 601-628, 2023.

[31] S. Sharma and P. P. Singh, "Advancements in legal text summarization: integrating InLegalBERT for effective extractive summarization," *International Journal of System Assurance Engineering and Management,* vol. 16, pp. 1382-1397, 2025.

[32] M. Kalyan, P. Dwivedi, and S. Sharma, "Enhancing Legal Document Summarization for Professionals: An Extractive Approach," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024, pp. 1-8.

[33] M. Malik, Z. Zhao, M. Fonseca, S. Rao, and S. B. Cohen, "Civilsum: A dataset for abstractive summarization of indian court decisions," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 2241-2250.

[34] M. Narzary, P. K. Singh, and M. Brahma, "Legal NLP in India: a comprehensive survey of tasks, challenges, and future directions," *AI & SOCIETY,* pp. 1-30, 2025.

[35] S. Sharma and P. P. Singh, "Advancing legal document summarization: introducing an approach using a recursive summarization algorithm," *SN Computer Science,* vol. 5, p. 927, 2024.

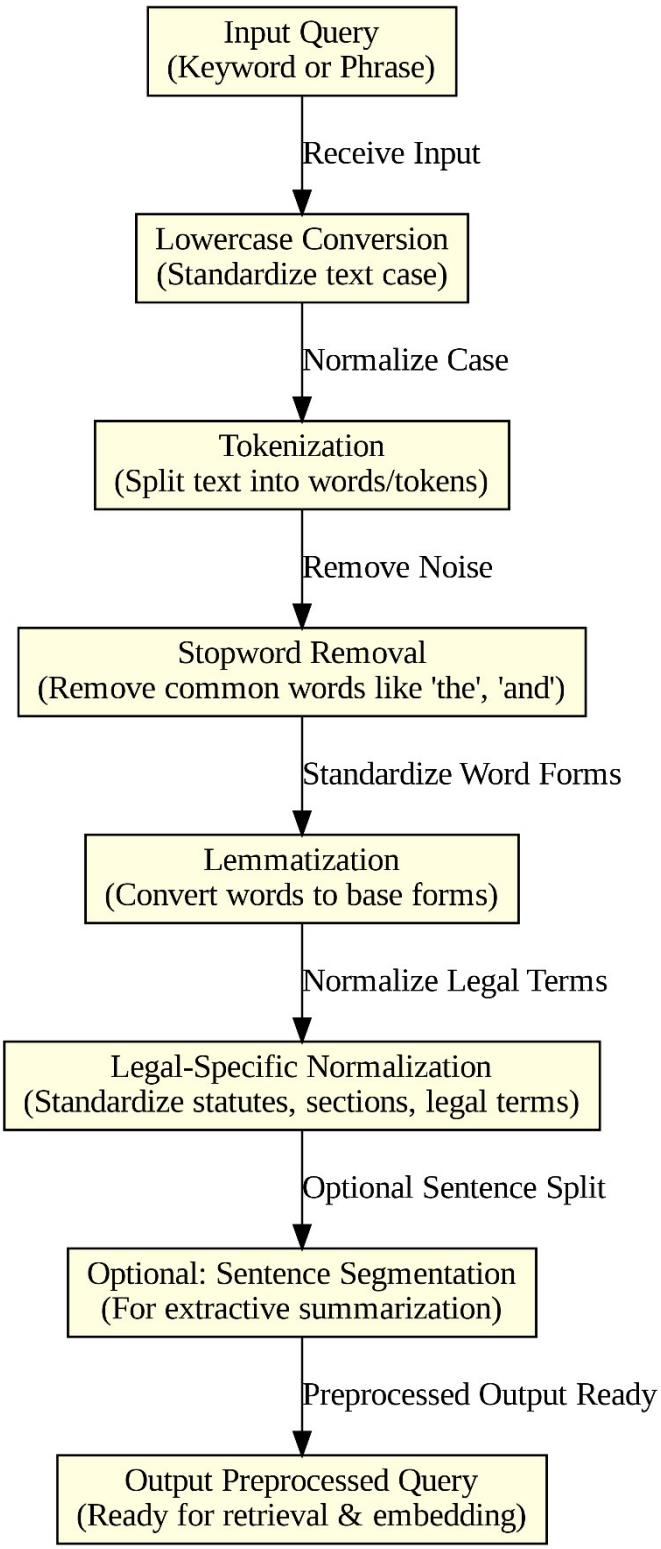

Figure 1: Steps involved in preprocessing

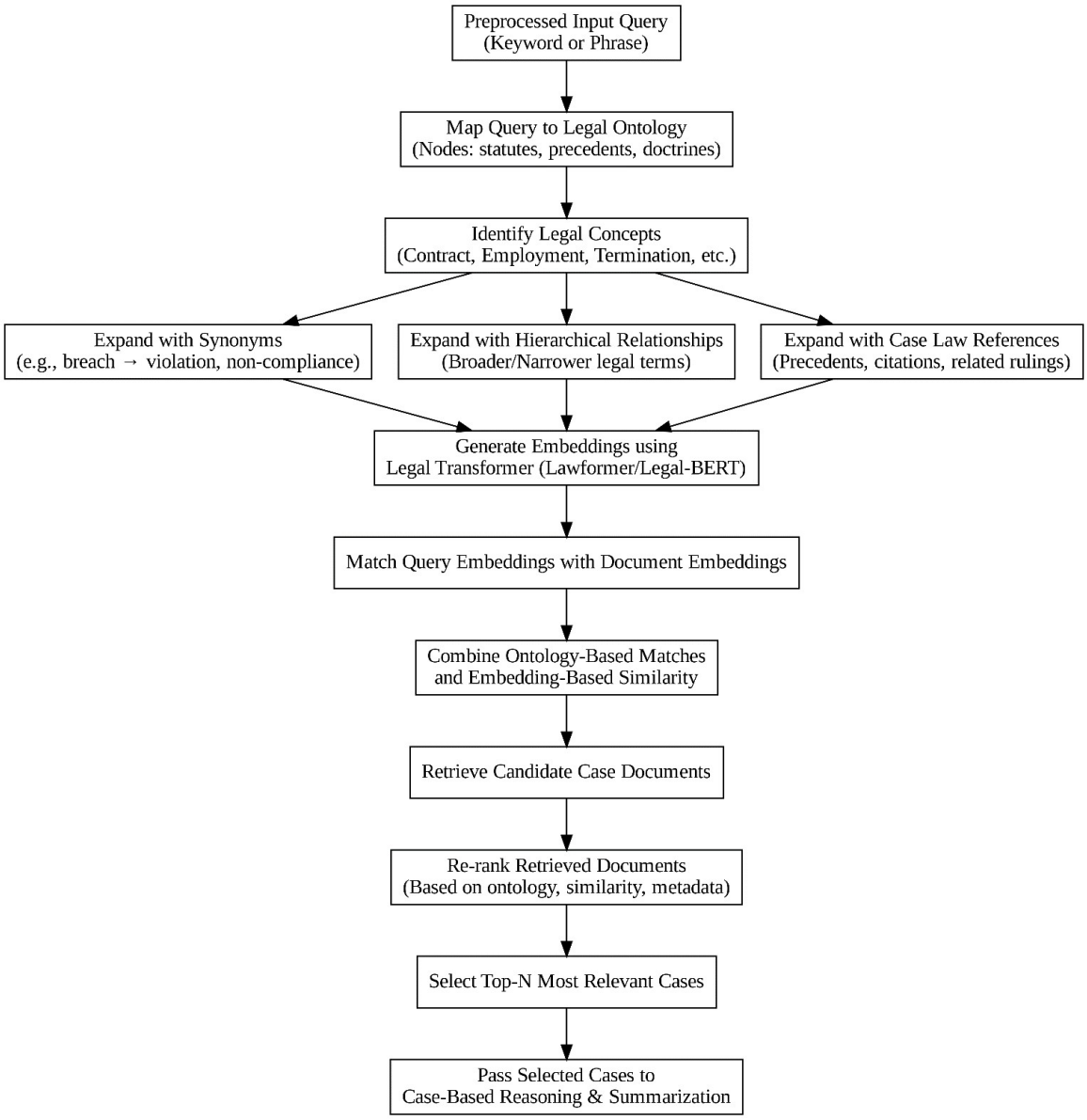

Figure 2: Flow of semantic ontology-based retrieval process

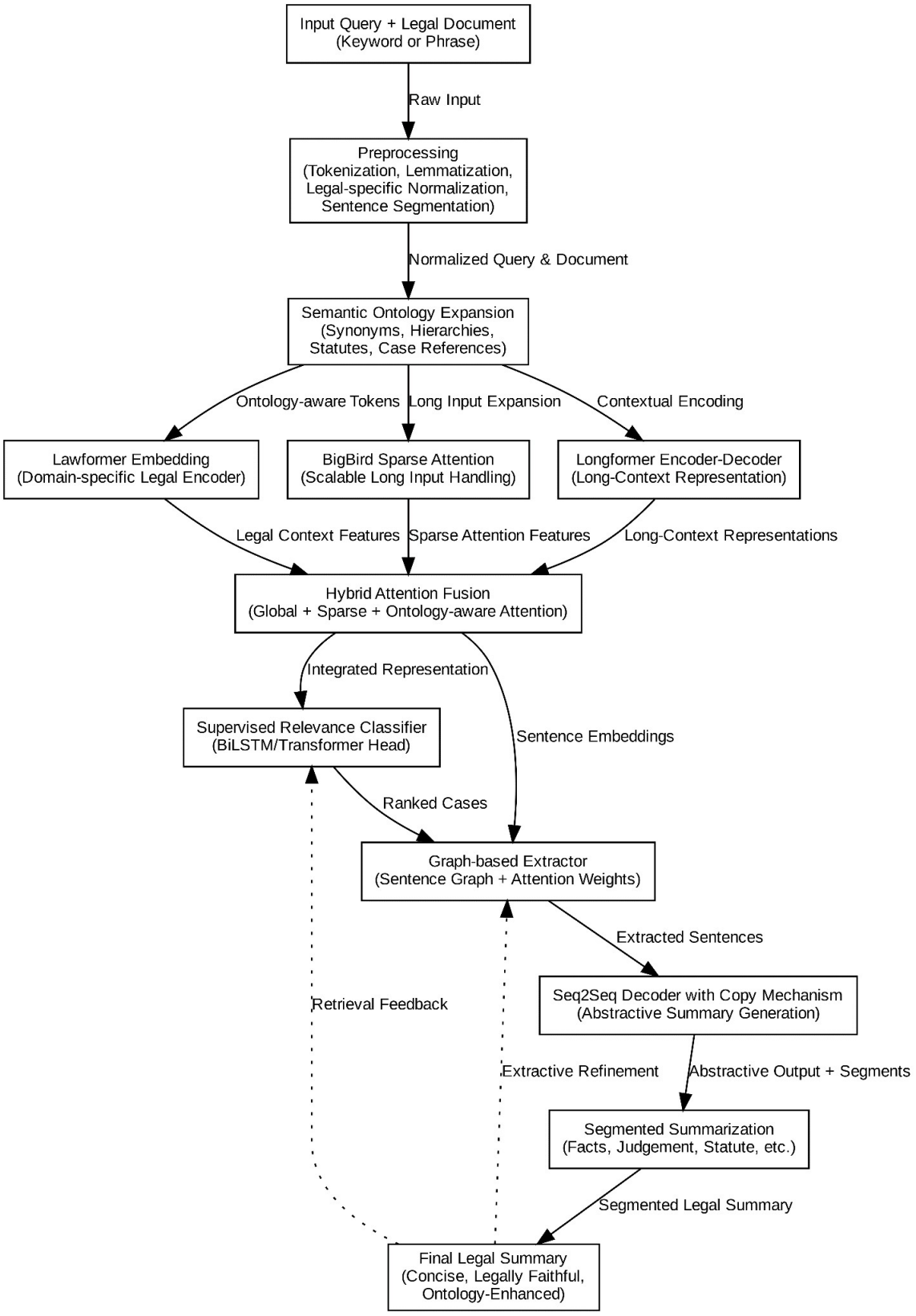

Figure 3: Flow of the proposed LawBird-LED

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. Qualitative Analysis of The Complete Lawbird-LED Retrieval–Summarization Pipeline

Analysis stages	Case -1	Case -1	Case - 3
Legal domain	Employment/Contract	Criminal	Constitutional/Administrative
User query	Breach of contract in employment	actual query	actual query
Ontology-expanded concepts	Wrongful termination; contractual violation; employment contract	actual concepts	actual concepts
Top retrieved precedent	actual case	actual case	actual case
Relevant legal concepts preserved	actual concepts	actual concepts	actual concepts
Baseline summary issue	actual observation	actual observation	actual observation
LawBird-LED summary advantage	actual observation	actual observation	actual observation