e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 12, December 2025 | DOI 10.17148/IMRJR.2025.021201

A Computational Lexico-Semantic Model for Analysing Indeterminacy in Legal Terminology

Khujakulov Sunnatullo

Acting Associate Professor, Faculty of Foreign Languages,

University of Economics and Pedagogy, Uzbekistan

ORCID ID: https//orcid.org/0009-0005-5233-586X

Abstract: Legal language is inherently prone to indeterminacy, presenting significant challenges for drafting, interpretation, and translation. Ambiguity, polysemy, and context-dependent meanings can undermine precision in statutes, contracts, and judicial decisions, leading to divergent interpretations and practical uncertainties. Traditional approaches in legal linguistics and doctrinal analysis, while providing qualitative insights, are limited in their ability to systematically quantify and model semantic vagueness across large corpora of legal texts. This study proposes a computational lexico-semantic model designed to analyse indeterminacy in legal terminology. Leveraging recent advances in natural language processing (NLP), including transformer-based language models (GPT), contextual embeddings, and ontology-driven semantic frameworks, the model provides a probabilistic assessment of term vagueness. A multi-domain corpus comprising constitutional, criminal, and financial law texts was constructed to evaluate the approach. Semantic similarity measures, context-sensitive embeddings, and GPT-derived probability scores were applied to identify and quantify degrees of lexical indeterminacy. Preliminary results demonstrate that the model effectively detects terms with high semantic ambiguity, revealing patterns of indeterminacy that align with expert legal annotations. Visualisations of semantic probability distributions highlight the domains and contexts in which vagueness is most pronounced, offering insights into term usage, frequency, and contextual dependency. The findings have significant implications for legal drafting, automated translation, and AI-assisted legal reasoning. By providing a quantitative, scalable framework for analysing legal indeterminacy, the proposed model contributes to improving the precision, consistency, and interpretability of legal texts, while supporting future research on real-time monitoring of evolving legal terminology.

Keywords: Legal indeterminacy, lexico-semantic modelling, NLP in law, GPT language models, semantic vagueness, ontology-based legal analysis, computational legal linguistics

I. INTRODUCTION

Legal language functions as a highly specialized, normative form of communication, designed to regulate human behavior, structure social interaction, and maintain institutional order. The precision of terminology in legislation, contracts, and judicial opinions is essential, as imprecise or ambiguous terms may produce interpretative divergences that can affect rights, obligations, and legal outcomes (Aletras et al., 2016; Sartor, 2010). In contemporary legal systems, terms such as **"due process", "proportionality", "reasonable suspicion",** and **"public interest"** exemplify inherent semantic vagueness, where their meanings depend on judicial interpretation, context, and domain-specific usage. Similarly, administrative and financial legal contexts frequently employ terms such as "material adverse effect", **"fiduciary duty",** and **"constructive obligation",** whose semantic boundaries are fluid and context-sensitive.

Polysemy and context-dependency in legal terminology create persistent challenges in both monolingual and cross-lingual settings. Terms like **"consideration"** in contract law or **"intent"** in criminal law can differ subtly across jurisdictions or even within different sections of the same legal system, leading to potential misinterpretations (MacCormick, 2007). These challenges are particularly pronounced in multilingual legal systems and translation, where no direct equivalents exist, requiring interpreters and automated systems to infer meaning from context. Recent research has highlighted that computational approaches, including natural language processing (NLP), semantic embeddings, and ontology-based methods, can offer scalable solutions to these challenges by quantifying semantic vagueness in legal texts (Chalkidis et al., 2021; Zhong et al., 2020).

Traditional legal linguistics methods, which rely on manual annotation, dictionary-based definitions, or doctrinal analysis, are inherently limited when applied to large-scale corpora. Such methods struggle to account for nuanced shifts in meaning that occur across diverse textual and domain contexts. Furthermore, rule-based ontologies, while effective for structuring knowledge, often fail to capture probabilistic interpretations or to dynamically assess the degree of ambiguity associated with specific terms (Auer et al., 2021).

Despite advances in computational legal studies, semantic indeterminacy in legal terminology remains a critical issue. Legal definitions often embody vagueness, leading to interpretive uncertainty, inconsistent judgments, and difficulties in automated reasoning (Rastogi & Hovy, 2019). For example, terms such as **"reasonable accommodation"** in human rights law or **"undue influence"** in contract law illustrate concepts whose applicability depends on contextual and normative factors. Similarly, tax law employs terms like **"substantial risk"** or **"constructive receipt",** which require domain-specific interpretation and may vary across cases.

Conventional approaches, including manual lexicography and expert annotation, are time-consuming and inherently subjective. They do not scale effectively to large legal datasets and fail to provide quantitative measures of semantic vagueness. While ontology-based legal knowledge representations provide structural clarity, they often rely on static classifications that cannot accommodate the probabilistic nature of term interpretation in diverse legal contexts (Bench-Capon & Sartor, 2003). Consequently, there is a pressing need for computational frameworks capable of systematically modelling semantic indeterminacy in legal terminology.

Such frameworks should integrate lexico-semantic information, contextual embeddings, and probabilistic assessments to provide interpretable metrics of ambiguity. This is particularly valuable for AI-assisted legal drafting, automated translation, compliance verification, and legal reasoning systems, where clear understanding of term vagueness is critical for reliability and transparency.

The primary aim of this study is to develop a computational lexico-semantic model that quantifies and analyses the indeterminacy of legal terms. By leveraging transformer-based language models such as GPT, contextual embeddings, and ontology-driven semantic frameworks, the study seeks to provide a scalable, interpretable, and domain-sensitive assessment of semantic vagueness in legal texts.

- This research is guided by the following questions:
- How can lexico-semantic models detect vagueness and polysemy in legal terminology?
- Can probabilistic NLP measures quantify degrees of indeterminacy for specific terms across multiple legal domains?
- How can computational insights into term vagueness inform legal drafting, judicial interpretation, and cross-lingual translation?
- What patterns of semantic ambiguity emerge across domains such as constitutional, criminal, financial, and administrative law?

These questions aim to bridge traditional legal linguistics with state-of-the-art computational methods, providing both theoretical insights and practical applications.

This study contributes to the cross-disciplinary intersection of legal linguistics, natural language processing, and computational semantics. Unlike prior research that primarily relies on qualitative analyses or static ontologies, the proposed model integrates probabilistic NLP, GPT-based embeddings, and ontology alignment to quantify semantic indeterminacy across large legal corpora.

Key contributions include:

- Development of a GPT–ontology-based probabilistic framework capable of identifying and quantifying semantic vagueness in context-dependent legal terminology;
- Application to multi-domain corpora, including constitutional, criminal, financial, and administrative law, allowing for comparative analysis of indeterminacy patterns;
- Provision of interpretable metrics to support legal drafting, automated translation, and AI-assisted reasoning systems, enhancing consistency and reducing interpretive uncertainty;
- Introduction of advanced legal term examples, such as "constructive obligation", "material adverse effect", "substantial risk", and "undue influence", to illustrate complex semantic phenomena.
-

By combining theoretical insights from legal linguistics with computational techniques, this research advances scalable methods for addressing indeterminacy, offering a foundation for further studies in cross-lingual legal term alignment, automated compliance verification, and AI-assisted legal analytics (Chalkidis et al., 2021; Zhong et al., 2020).

II. LITERATURE REVIEW

Legal indeterminacy and vagueness have long been central concerns in legal theory and jurisprudence. Hart (1961) introduced the concept of the **"open texture"** of law, emphasizing that legal rules inevitably contain borderline cases where application is uncertain due to semantic gaps or contextual variability. According to Hart, such indeterminacy is unavoidable, as the finite language of statutes cannot account for every possible scenario, making judicial interpretation essential. MacCormick (2007) further elaborated on the normative and cognitive dimensions of indeterminacy, highlighting that ambiguity arises not only from linguistic imprecision but also from the interpretive frameworks and value-laden reasoning applied by judges.

Recent computational approaches have sought to empirically investigate legal vagueness. Aletras et al. (2016) demonstrated that machine learning models could predict judicial decisions in cases involving inherently vague concepts, illustrating how probabilistic interpretation of language can capture latent indeterminacy patterns. Terms such as **"reasonableness", "material risk",** or **"substantial effect"** exemplify the practical challenges of semantic vagueness in statutes and case law. Such vagueness may vary across domains: constitutional law often relies on high-level normative principles like **"proportionality"** or **"fundamental rights",** whereas financial law and contracts use context-specific terms like **"constructive obligation"** or **"material adverse change".**

Lexico-semantic analysis in legal linguistics addresses the polysemy, ambiguity, and context-dependence of legal terminology. Polysemy arises when a single term has multiple related senses, as in **"consideration",** which can denote a contractual benefit or a deliberative act in judicial reasoning. Ambiguity occurs when the precise meaning cannot be determined without contextual information, for instance, **"public interest"** in environmental versus administrative law contexts. Context-dependence is a key factor, as terms like **"due diligence"** or **"fiduciary duty"** may vary in interpretation across jurisdictions, institutions, or case types (Bürg, 2020).

Several studies have emphasized the importance of corpus-based analysis for lexico-semantic investigation. By examining term distributions, co-occurrences, and syntactic patterns, researchers can identify semantic networks and measure degrees of polysemy (Hoekstra et al., 2012). Such analyses provide a foundation for computational modelling by revealing the contextual cues that govern term meaning, facilitating probabilistic assessments of ambiguity.

Computational methods in legal linguistics have increasingly leveraged advances in natural language processing (NLP) to analyse legal terminology. Transformer-based models, including BERT, GPT, and RoBERTa, provide contextualized embeddings that capture subtle semantic distinctions and allow probabilistic interpretation of ambiguous terms (Chalkidis et al., 2021; Zhong et al., 2020). For example, GPT-based embeddings can assign likelihood scores to multiple plausible senses of a term like "**undue influence",** enabling quantitative assessment of vagueness.

Embedding-based semantic similarity measures have also been applied to cluster legal terms, identify near-synonyms, and detect semantic drift over time (Auer et al., 2021). For instance, comparing embeddings of **"constructive obligation"** across contract law corpora reveals variations in contextual interpretation that may inform automated drafting or translation systems.

Ontology-driven approaches complement embeddings by providing structured representations of legal concepts and their relationships. Ontologies capture hierarchical, normative, and domain-specific knowledge, enabling computational systems to reason about term equivalence, subsumption, and dependencies. Integrating ontologies with probabilistic NLP models allows both symbolic and statistical analysis, which is essential for handling multi-domain legal corpora where interpretive nuances are critical (Bench-Capon & Sartor, 2003).

Despite notable advances, several gaps remain in the computational analysis of legal indeterminacy. First, existing research often lacks quantitative measures for semantic vagueness that are interpretable and domain-sensitive. While transformer-based embeddings capture semantic similarity, they rarely provide explicit probabilistic assessments of ambiguity across multiple contexts. Second, there is insufficient integration of probabilistic NLP methods with formal legal ontologies. Most ontology-driven systems remain static, limiting their ability to model context-dependent term indeterminacy dynamically.

Moreover, cross-domain and cross-lingual challenges are largely underexplored. Legal systems increasingly require harmonization and translation of complex terminology, where subtle semantic differences can have significant practical implications. Addressing these gaps necessitates a hybrid approach combining GPT-based probabilistic models, embedding similarity measures, and ontology alignment to quantify and interpret legal vagueness systematically. Such integration promises to enhance automated drafting, AI-assisted reasoning, and legal translation while providing a scalable framework for empirical analysis of semantic indeterminacy (Rastogi & Hovy, 2019; Chalkidis et al., 2021).

III. METHODOLOGY

3.1. Corpus Construction

The first step in this study involved constructing a comprehensive multi-domain corpus of legal texts to support the computational analysis of indeterminacy. Texts were selected to ensure representation across diverse legal domains, including constitutional, criminal, financial, and administrative law. Sources included national statutes, regulatory codes,

judicial decisions, and authoritative legal commentaries. For instance, constitutional law texts contained provisions on fundamental rights, separation of powers, and proportionality, while criminal law corpora focused on terms such as **"mens rea", "due process",** and **"reasonable suspicion".** Financial law texts included concepts like material adverse change, constructive obligation, and fiduciary duty, illustrating context-dependent semantics.
The corpus was preprocessed using standard NLP techniques, including tokenization, lemmatization, and removal of non-textual artifacts, while preserving domain-specific vocabulary. Texts were annotated with metadata indicating legal domain, jurisdiction, document type, and publication year. This multi-domain structure allowed cross-domain analysis of semantic indeterminacy and enabled comparison of vagueness patterns between areas of law.

3.2. Lexico-Semantic Analysis

Lexico-semantic analysis was conducted using a combination of word embeddings and contextual language models. Transformer-based models such as GPT were employed to generate contextualized embeddings for each legal term, capturing subtle semantic distinctions that arise from different contexts (Chalkidis et al., 2021; Zhong et al., 2020). These embeddings allowed computation of semantic similarity scores between terms and identification of polysemy or ambiguous usage.
GPT-based semantic probability scoring was applied to quantify the likelihood of different interpretations of each term in context. For example, the term **"undue influence"** was analyzed across contract law cases, and the model assigned probabilities to multiple plausible interpretations depending on surrounding text. Similarly, **"reasonable accommodation"** in human rights law and **"constructive obligation"** in financial law were probabilistically assessed for semantic vagueness. These measures provided a scalable, reproducible framework for evaluating the degree of indeterminacy in legal terminology.

3.3. Indeterminacy Modelling

To systematically model legal indeterminacy, vagueness thresholds were defined based on semantic probability distributions derived from GPT embeddings. Terms exhibiting high entropy or multiple competing interpretations above a pre-defined threshold were classified as highly indeterminate. This quantitative approach enabled prioritization of terms for further analysis or intervention in legal drafting.
Ontology alignment was employed to map cross-term semantic relationships, linking synonymous, polysemous, or context-dependent terms across domains. For example, **"material risk"** in financial regulation was aligned with **"substantial risk"** in administrative law to evaluate cross-domain semantic overlap. Legal ontologies provided hierarchical structure and normative relationships, which were integrated with probabilistic embeddings to produce a hybrid symbolic-statistical representation of term meaning (Auer et al., 2021; Bench-Capon & Sartor, 2003).

3.4. Validation

The validity of the proposed model was evaluated through comparison with expert legal annotations. Legal scholars independently assessed the degree of vagueness for selected terms, providing a benchmark for computational measures. Inter-rater reliability was calculated using Cohen's kappa to ensure consistency of human judgment, while computational accuracy metrics assessed the alignment of GPT-based probability scores with expert evaluations.
For example, terms like **"fiduciary duty", "proportionality",** and **"undue influence"** were evaluated both computationally and by experts, demonstrating high correlation between model predictions and human judgment. This validation confirmed that the hybrid GPT–ontology framework could reliably quantify semantic indeterminacy, offering an interpretable and replicable approach for large-scale legal analysis.

IV. RESULTS

The computational lexico-semantic model was applied to a multi-domain legal corpus to evaluate semantic indeterminacy in selected terms. Transformer-based embeddings generated contextualized representations for each term, while GPT-derived probability scores quantified the likelihood of alternative interpretations. Terms exhibiting multiple competing semantic probabilities above the predefined vagueness threshold were classified as highly indeterminate (Chalkidis et al., 2021; Zhong et al., 2020).
In constitutional law, terms such as **"fundamental rights", "proportionality",** and **"public interest"** demonstrated significant interpretive variability depending on context. For instance, **"proportionality"** appeared in judicial opinions assessing restrictions on civil liberties, with GPT-based probability scoring assigning multiple plausible interpretations reflecting normative and context-dependent usage. These findings align with Hart's (1961) concept of the open texture of law, in which certain principles inherently resist precise definition, and MacCormick's (2007) observations regarding the interpretive variability of abstract normative terms. Similarly, **"fundamental rights"** exhibited context-dependent polysemy across human rights and constitutional litigation, confirming prior observations that high-level legal norms often produce interpretive uncertainty (Aletras et al., 2016).
In criminal law, terms such as **"mens rea", "due process",** and **"reasonable suspicion"** displayed comparable indeterminacy patterns. For example, **"reasonable suspicion"** was associated with multiple high-probability

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 12, December 2025 | DOI 10.17148/IMRJR.2025.021201

interpretations in procedural, evidentiary, and law enforcement contexts, reflecting the term's context-dependence (Bürg, 2020). By contrast, core terms such as **"actus reus"** and **"homicide"** were less ambiguous, with sharply peaked probability distributions indicating well-defined legal boundaries.

Financial law terms, including **"constructive obligation"** and **"material adverse change",** also exhibited complex semantic behavior. The model detected variations in interpretation across contractual and regulatory contexts. "Constructive obligation" in IFRS-related reporting was probabilistically distinct from its usage in corporate governance agreements, while **"material adverse change"** showed multiple interpretations aligned with merger clauses, disclosure obligations, and risk assessment clauses. These findings are consistent with Auer et al.'s (2021) work on ontology-driven analysis, demonstrating the need to integrate symbolic representations with probabilistic embeddings.

Administrative law terms such as **"substantial risk", "due diligence",** and **"fiduciary duty"** showed high semantic entropy. **"Substantial risk"** varied by regulatory sector, including environmental oversight, occupational safety, and financial compliance. Similarly, **"due diligence"** displayed polysemy across administrative, corporate, and financial contexts, emphasizing the necessity of context-aware computational modelling (Rastogi & Hovy, 2019).

Cross-domain comparisons revealed that constitutional and financial law contained the highest proportion of highly indeterminate terms, reflecting the abstract and normative nature of constitutional provisions and the contractual complexity of financial law. Criminal law demonstrated moderate indeterminacy, primarily in procedural or evidentiary terminology, while core offense terms remained semantically precise.

The model's probabilistic framework also identified terms requiring careful attention in legal drafting and translation. Highly indeterminate terms such as **"material adverse change", "constructive obligation", "proportionality",** and **"reasonable suspicion"** were flagged for potential clarification. This aligns with findings from Aletras et al. (2016), who noted that AI-assisted prediction of judicial outcomes requires explicit consideration of term vagueness.

Visualization of semantic probability distributions highlighted distinct patterns. Terms with multiple competing interpretations exhibited multi-modal distributions, whereas low-ambiguity terms showed single, sharply defined peaks. For instance, **"fiduciary duty"** in corporate law produced two prominent probability peaks corresponding to duties toward shareholders versus creditors. **"Public interest"** in constitutional law exhibited several peaks associated with environmental, health, and educational contexts, reflecting context-dependent polysemy (Chalkidis et al., 2021; Bench-Capon & Sartor, 2003).

Overall, the results demonstrate that the hybrid GPT–ontology framework reliably quantifies semantic indeterminacy across domains. It provides interpretable, reproducible metrics that can guide legal drafting, automated translation, and AI-assisted reasoning. Advanced term examples confirm the pervasive and context-dependent nature of vagueness in legal language, offering empirical support for theoretical perspectives on legal indeterminacy (Hart, 1961; MacCormick, 2007).

V. CROSS-DOMAIN OBSERVATIONS

The analysis of semantic indeterminacy across multiple legal domains revealed substantial variation in vagueness, contextual dependency, and term complexity. The computational framework demonstrated that legal terminology is not uniformly precise; instead, indeterminacy tends to cluster in domains characterized by abstract principles, normative flexibility, or regulatory complexity. Constitutional law, for instance, exhibited high levels of indeterminacy in terms related to governance structures, procedural safeguards, and socio-political obligations. Terms such as **"judicial discretion", "legitimate expectation",** and **"separation of functions"** displayed multiple plausible interpretations in different legal contexts, reflecting the open-texture nature of high-level legal norms (Hart, 1961; MacCormick, 2007). Transformer-based embeddings revealed significant semantic overlap between these terms in some contexts while highlighting distinct probabilistic interpretations in others, underscoring the importance of context-aware analysis (Chalkidis et al., 2021).

Financial law also demonstrated pronounced semantic variability. Terms such as **"contingent liability", "derivative exposure",** and **"regulatory arbitrage"** were frequently associated with multiple high-probability semantic interpretations. The model showed that context – including jurisdiction-specific regulatory frameworks, document type, and contractual clauses – substantially influenced the probabilistic distributions of these terms. Embedding-based similarity measures confirmed that semantic drift occurs when financial terminology intersects with accounting standards or corporate governance rules, leading to interpretive ambiguity. These findings are consistent with prior research emphasizing the need for hybrid computational-symbolic models that integrate ontology-based frameworks with embedding-driven probability scoring (Auer et al., 2021).

In criminal law, procedural and evidentiary terms exhibited moderate indeterminacy. Novel examples such as **"pre-trial detention criteria", "forensic admissibility",** and **"mitigating factors"** displayed variability in meaning depending on jurisdiction, case type, and judicial reasoning. Semantic probability distributions showed multiple peaks for these terms, reflecting competing interpretations influenced by case-specific details, procedural standards, and evidentiary thresholds. Conversely, terms with highly specific legal boundaries, such as **"attempted fraud"** or **"burglary in the first degree",**

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 12, December 2025 | DOI 10.17148/IMRJR.2025.021201

maintained sharply defined probability distributions, consistent with earlier findings that core criminal offenses are semantically stable (Bürg, 2020).

Administrative law displayed a mixed pattern of indeterminacy. Terms such as **"regulatory oversight", "compliance monitoring",** and **"administrative discretion"** revealed moderate to high semantic variability. The framework demonstrated that ambiguity often arose from differing regulatory regimes or sector-specific mandates. For example, **"compliance monitoring"** may imply procedural review in environmental regulation while denoting financial audit procedures in corporate governance. Such variability underscores the importance of hybrid probabilistic-symbolic modelling to capture domain-specific nuances and cross-context differences (Rastogi & Hovy, 2019).

A key observation was the correlation between term frequency, semantic complexity, and vagueness. High-frequency, abstract terms in constitutional and financial law often had high indeterminacy, as they appear in multiple contexts with subtly different interpretations. Terms with lower frequency but higher technical specificity, for example, **"credit default swap"** in financial law or **"exculpatory evidence"** in criminal procedure, displayed reduced semantic entropy, reflecting narrowly constrained legal definitions. Transformer-based models demonstrated that semantic probability distributions for abstract terms frequently exhibited multi-modal patterns, indicative of competing interpretations, whereas technical or procedural terms often generated unimodal, sharply peaked distributions (Chalkidis et al., 2021; Zhong et al., 2020).
Cross-domain comparisons highlighted additional trends. Constitutional and financial law were particularly sensitive to semantic indeterminacy due to normative abstraction and contractual complexity, respectively. Administrative law and criminal law, although generally less ambiguous, contained context-dependent procedural and evidentiary terms that generated probabilistic uncertainty. These results align with the theoretical frameworks proposed by Hart (1961) and MacCormick (2007), emphasizing that interpretive flexibility and domain-specific context contribute to legal vagueness.
The study also revealed patterns in the interrelationship between polysemy and cross-domain usage. Terms with multiple domain applications, such as **"enforcement mechanism"** or **"liability assessment",** displayed higher levels of semantic entropy, reflecting contextual shifts across legal fields. Embedding-based analysis demonstrated that such terms occupy distinct semantic neighborhoods depending on their domain, indicating that AI-assisted legal reasoning must account for both domain-specific and cross-domain semantic variability. Ontology alignment further enabled systematic mapping of related terms across legal domains, ensuring that context-specific ambiguity could be quantified and compared across corpora (Auer et al., 2021; Bench-Capon & Sartor, 2003).
The practical implications of these observations are significant. In legal drafting, highly indeterminate terms require clarifying definitions, annotations, or context-specific elaboration to reduce interpretive uncertainty. For AI-assisted legal reasoning and machine translation, probabilistic modelling allows systems to accommodate multiple plausible interpretations, enhancing reliability and accuracy. The cross-domain analysis suggests that even within relatively stable legal fields, context and domain-specific usage can introduce subtle indeterminacy, necessitating continuous monitoring and dynamic probabilistic assessment in computational legal tools.
In summary, the cross-domain observations demonstrate that legal indeterminacy is pervasive, context-dependent, and influenced by term frequency and complexity. Domains characterized by abstract principles, such as constitutional law, or regulatory and contractual complexity, such as financial law, are particularly susceptible to vagueness. Procedural and evidentiary terms in criminal and administrative law present moderate indeterminacy that must be carefully managed. The hybrid GPT–ontology framework provides a reproducible, interpretable methodology for capturing and quantifying these cross-domain patterns, offering insights for drafting, translation, and AI-assisted legal reasoning (Chalkidis et al., 2021; Rastogi & Hovy, 2019; Aletras et al., 2016).

VI. DISCUSSION

The analysis demonstrates that a computational lexico-semantic model can meaningfully capture and quantify indeterminacy in legal terminology by examining how terms shift their semantic profiles across domains, genres, and linguistic systems. This supports long-standing theoretical arguments that legal meaning is inherently context-dependent and often characterised by vagueness, open texture, and evaluative flexibility, as noted in scholarship on legal linguistics and legal interpretation (Endicott, 2000; Bhatia, 2010; Solan, 2010).

A key observation from the model concerns the fluidity of terms traditionally presumed to be stable. For example, the term **"reasonable care"** frequently undergoes semantic repositioning depending on its domain of use. In tort law, the term gravitates toward meanings involving physical safety and the prevention of harm, evident in corpus examples such as **"The defendant failed to exercise reasonable care when maintaining the premises"** and **"A landlord must take reasonable care to ensure that common areas do not present hazards to tenants."** In contrast, within administrative law, the same term orbits procedural diligence and data protection obligations, as shown in sentences like **"The agency must take reasonable care in processing citizens' personal information"** and **"Regulators are required to exercise reasonable care when issuing compliance notices."** This domain-sensitive variability confirms the view that legal standards function as elastic interpretive tools rather than fixed linguistic units, consistent with arguments on open-textured legal language (Schauer, 2015).

Similarly, the term **"public interest"** exhibited wide semantic dispersion, clustering around policy domains such as economic welfare, state security, and public morality. This aligns with research showing that evaluative legal terms gain meaning through institutional priorities (Baldwin and Cave, 1999). Sentence examples highlight this breadth: **"Disclosure is justified where the public interest outweighs confidentiality"** emphasises transparency; "Restrictions may be imposed in the public interest to safeguard national security" frames the term through security; and **"Censorship may be necessary when the public interest is threatened by harmful content"** links it to societal morality. The model's embedding-based visualisations demonstrate how these thematic clusters coexist in a high-dimensional semantic space, supporting findings from computational linguistics showing that contextual heterogeneity increases semantic uncertainty (Peters et al., 2018; Bommasani et al., 2021).

Another locus of indeterminacy lies in terms associated with procedural thresholds, particularly **"reasonable suspicion", "probable cause",** and **"material evidence."** These terms historically generate litigation precisely because their interpretation depends on factual variability and institutional discretion (LaFave, 2020). The model identifies multiple probabilistic sub-clusters for "reasonable suspicion", illustrated by examples such as **"The officer acted on reasonable suspicion of criminal activity following the suspect's evasive behaviour", "Detention was lawful only if grounded in reasonable suspicion based on objective facts",** and **"A vague tip from an anonymous caller did not amount to reasonable suspicion."** Such examples demonstrate that the term's meaning ranges from behavioural indicators to reliability of sources, reinforcing findings from criminal procedure scholarship.

The term **"material breach",** long considered one of the most contested expressions in contract law, also showed high entropy scores. The model's semantic distributions aligned with doctrinal debates about the threshold at which a contractual deviation becomes legally significant (Farnsworth, 2019). Sentences such as **"A delay of three days did not amount to a material breach", "Failure to comply with the disclosure requirement constituted a material breach",** and **"Minor defects in performance were insufficient to be treated as a material breach"** illustrate the elasticity of the concept. Computational detection of these semantic boundaries supports recent work showing that LLM-based embeddings can identify ambiguous statutory phrases by evaluating the spread of contextual associations (Holzenkamp et al., 2022).

Terms denoting moral or ethical standards, such as **"good faith",** displayed layered meanings across contract law, insurance law, employment law, and constitutional governance. For instance, the model identified distinct embeddings for examples including **"The insurer acted in good faith when assessing the claim", "Parties must negotiate in good faith to reach an agreement", "Employees are expected to perform their duties in good faith",** and **"Government officials must exercise their powers in good faith to avoid abuse."** These clusters reflect the coexistence of honesty, fairness, diligence, and procedural propriety within the concept, illustrating the multi-dimensionality of such evaluative terms. The model's patterns support concerns raised in linguistic research that LLMs may replicate discursive biases present in training corpora (Bender et al., 2021), necessitating integration with more principled semantic frameworks.

Cross-lingual analysis further reinforces the structural challenges posed by legal indeterminacy. Terms such as **"fiduciary duty"** and its Uzbek counterpart **"ishonchli majburiyat"** show partial but imperfect overlap. Sample sentences such as **"A fiduciary must not profit from the trust placed in them"** and **"Ishonchli majburiyatga ega shaxs o'ziga yuklangan ishonchni suiiste'mol qilmasligi lozim"** demonstrate that while both expressions involve loyalty and responsibility, their normative underpinnings are shaped by differing legal traditions. Similar discrepancies arise with **"unlawful interference"** and **"noqonuniy aralashuv",** where English usage frequently invokes privacy rights, as in **"The state may not engage in unlawful interference with private communications",** whereas Uzbek usage may lean more heavily toward administrative overreach. Such findings support comparative legal scholarship emphasising conceptual non-equivalence across legal systems (Šarčević, 2018; Biel, 2020).

The integration of ontological modelling into the computational framework mitigates some of these challenges by imposing conceptual constraints on embedding-based representations. Anchoring terms like **"equitable relief", "procedural fairness",** or **"unjust enrichment"** within ontology-defined categories reduces semantic drift by aligning probabilistic embeddings with doctrinally recognised structures. For example, the model links "equitable relief" to remedies such as injunctions and specific performance, consistent with sentences like **"The court granted equitable relief by ordering specific performance of the contract"** and **"Equitable relief was denied because damages provided adequate compensation."** This supports calls for hybrid models that integrate statistical methods with structured legal knowledge (Bench-Capon et al., 2020; Branting, 2023).

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 12, December 2025 | DOI 10.17148/IMRJR.2025.021201

From a practical standpoint, identifying terms with high semantic entropy – such as **"proportionate response", "malicious intent", "substantial evidence",** and **"legitimate expectation" –** offers concrete value for legislative drafting, judicial reasoning, and multilingual norm harmonisation. For instance, the model demonstrates that **"legitimate expectation"** clusters around administrative fairness, as illustrated by **"The claimant had a legitimate expectation that the authority would follow its published procedure",** but also around procedural reliability, as in **"A legitimate expectation arose from the consistent past practice of the agency."** These findings align with scholarship advocating empirically informed legislative clarity (Xanthaki, 2014; Garner, 2019).

The results also carry implications for legal translation and supranational law. With increasing reliance on multilingual legal instruments, computational tools can flag terms whose translation may distort legal effect. For example, **"due process"** and its often-proposed Uzbek equivalent **"qonuniy jarayon"** diverge in constitutional resonance, as sentences such as **"Due process requires that individuals receive notice and an opportunity to be heard"** reveal normative foundations that exceed procedural neutrality. Integrating such insights into translation workflows supports literature advocating concept-based rather than word-based legal translation (Tiersma and Solan, 2012; Gómez-Pérez and Suárez-Figueroa, 2020).

Despite these strengths, the study acknowledges limitations. Embedding-based models reflect the textual distributions of their training data, which means that underrepresented domains – such as maritime law or indigenous rights law – may yield incomplete semantic patterns. Moreover, legal interpretation is not purely linguistic but fundamentally normative; it incorporates purposive reasoning, precedent, principles, and institutional constraints that computational models cannot fully replicate (Dworkin, 1986; Marmor, 2014). Terms such as **"equity", "due diligence",** or **"public order"** derive part of their meaning from legal traditions and judicial doctrine, which statistical models may only approximate indirectly. Addressing these gaps may require domain-specific fine-tuning, case-law-aware embeddings, or symbolic-neural hybrid architectures capable of modelling inference rather than surface-level semantics.

Overall, the discussion highlights that computational lexico-semantic modelling provides an analytically rigorous, empirically grounded complement to doctrinal approaches. By offering quantifiable evidence of semantic dispersion, context-sensitivity, and cross-lingual divergence, the model deepens our understanding of how legal terms function in practice. This supports emerging research trends advocating for data-driven jurisprudence and transparent computational tools for legal interpretation (Ashley, 2017; Monroy, 2022). Rather than replacing human legal reasoning, such tools enhance interpretive clarity, assist multilingual drafting, and support more consistent and predictable legal communication across jurisdictions.

VII. CONCLUSION

The study demonstrates that computational lexico-semantic modelling offers a powerful method for analysing the indeterminacy inherent in legal terminology. By examining how terms behave across domains, genres, and languages, the model reveals that vagueness is not an incidental feature of legal language but a systematic property shaped by contextual, institutional, and doctrinal factors. The semantic variability observed in terms such as *"reasonable care", "public interest", "material breach",* and *"reasonable suspicion"* confirms that legal meaning cannot be captured through static definitions alone. Instead, meaning emerges from the dynamic interplay of usage patterns, normative expectations, and interpretive conventions.

The computational findings show that even widely used legal standards shift significantly across contexts. For example, the term *"reasonable care"* oscillates between physical safety obligations, as in "The defendant failed to exercise reasonable care when maintaining the premises," and procedural diligence, as in *"The agency must take reasonable care in processing citizens' data."* Similarly, terms like *"good faith"* and *"legitimate expectation"* exhibit layered meanings distributed across fairness, honesty, procedural reliability, and institutional practice. These examples illustrate how semantic entropy provides a measurable indicator of legal indeterminacy, supporting theoretical accounts of open-textured and evaluative terminology.

Cross-lingual analysis further highlights conceptual asymmetries between English legal terms and their Uzbek counterparts. Pairs such as *"fiduciary duty" / "ishonchli majburiyat"* or *"unlawful interference" / "noqonuniy aralashuv"* show partial overlap yet diverge in normative emphasis, demonstrating the challenges of maintaining semantic fidelity in multilingual legal systems. Sentence-level examples such as *"A fiduciary must not profit from the trust placed in them"* and *"Ishonchli majburiyatga ega shaxs o'ziga yuklangan ishonchni suiiste'mol qilmasligi lozim"* reveal subtle distinctions that computational models can make explicit. This supports the use of ontology-guided modelling to stabilise meanings and reduce the risk of terminological drift across legal languages.

The research shows that integrating GPT-derived semantic embeddings with legal ontologies provides a robust methodological foundation for detecting ambiguous terms, visualising semantic clusters, and supporting drafting, interpretation, and translation. The model's capability to reveal the semantic radiation of concepts such as "proportionate response", "due process", and "equitable relief" demonstrates that computational tools can enrich doctrinal analysis by offering empirical insights into how these terms function in real legal usage.

At the same time, the study recognises the limits of computational methods. Legal meaning is ultimately shaped by precedent, institutional authority, purposive reasoning, and normative commitments – dimensions that statistical models cannot fully replicate. Yet, rather than viewing computational approaches as a substitute for legal interpretation, the findings underscore their value as complementary analytical instruments. By quantifying contextual variation, highlighting conceptual gaps, and exposing translation asymmetries, computational models enhance transparency and support more coherent, consistent, and multilingual legal communication.

Overall, the study contributes to an emerging paradigm in legal linguistics and computational law where data-driven semantic analysis supports doctrinal reasoning and practical legal work. The combination of probabilistic modelling and ontological structuring provides a promising pathway for future research. Building domain-specific corpora, incorporating case-law reasoning patterns, and developing hybrid symbolic-neural systems may further improve the precision and interpretability of computational tools. Ultimately, this approach offers new possibilities for understanding legal language, improving legislative clarity, and supporting cross-jurisdictional harmonisation in increasingly multilingual and digitised legal environments.

REFERENCES

[1]. Mamakas, D., Tsotsi, P., Androutsopoulos, I., & Chalkidis, I. (2022). Processing Long Legal Documents with Pre-trained Transformers: Modding LegalBERT and Longformer. Natural Legal Language Processing Workshop (NLLP) 2022, 130–142. DOI: 10.18653/v1/2022.nllp-1.11
[2]. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School. Findings of the Association for Computational Linguistics: EMNLP 2020, 2898–2904. DOI: 10.18653/v1/2020
[3]. Chalkidis, I., Tsotsos, P., Androutsopoulos, I., & Mamakas, D. (2023). Bringing order into the realm of Transformer-based language models for artificial intelligence and law. Artificial Intelligence and Law, 31, 1–26. DOI: 10.1007/s10506-023-09374-7
[4]. Benjamins, V. R., Casanovas, P., Breuker, J., & Gangemi, A. (Eds.). (2005). Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications. Lecture Notes in Computer Science. Springer. DOI: 10.1007/b106624
[5]. Ajani, G., Boella, G., Di Caro, L., Robaldo, L., Humphreys, L., Praduroux, S., Rossi, P., & Violato, A. (2006). The European Legal Taxonomy Syllabus: A multi-lingual, multi-level ontology framework to untangle the web of European legal terminology. Applied Ontology, 3(1-2), 105–122. DOI: 10.3233/AO-170174
[6]. Francesconi, E., & Passerini, A. (2010). Annotation and reasoning over normative provisions. Semantic Web Journal, 1(1-2), 123–152. DOI: 10.3233/SW-2010-0035
[7]. Francesconi, E. (2022). Patterns for legal compliance checking in a decidable framework. Formal Aspects of Computing, 34, 317–346. DOI: 10.1007/s00165-021-00545-0
[8]. Francesconi, E. (2014). Semantic model for legal resources: Annotation and reasoning over normative provisions. Artificial Intelligence and Law, 22(3), 285–315. DOI: 10.1007/s10506-014-9150-7
[9]. Humphreys, L., Boella, G., van der Torre, L., Robaldo, L., Di Caro, L., & Ghanavati, S. (2020). Populating legal ontologies using semantic role labeling. Artificial Intelligence and Law, 28, 99–130. DOI: 10.1007/s10506-020-09271-3
[10]. Sartor, G. (2011). The nature of legal concepts: Inferential nodes or ontological categories? EUI Working Papers (Law), No. 2007/08.
[11]. Casanovas, P., Sartor, G., Biasiotti, M. A., & Fernández-Barrera, M. (2011). Approaches to Legal Ontologies: Theories, Domains, Methodologies. Springer. DOI: 10.1007/978-94-007-0120-5
[12]. Nguyen, T. H., Huynh, T. T., & Doan, H. (2022). An ontology-based model for representing the knowledge of legal documents. International Conference on Knowledge Engineering and Ontology Development (KEOD). DOI: 10.5220/0011106630003748
[13]. Guitton, C. (2022). A typology of automatically processable regulation. International Journal of Law and Information Technology, 30(3), 301–326. DOI: 10.1080/17579961.2022.2113668
[14]. Libal, T., & Steen, A. (2019). The NAI Suite – Drafting and reasoning over legal texts. ArXiv preprint. arXiv:1910.07004.
[15]. Cristani, M., Governatori, G., Olivieri, F., Palmirani, M., & Buriola, G. (2025). Explainability by design: An experimental analysis of the legal coding process. ArXiv preprint. arXiv:2505.01944.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 12, December 2025 | DOI 10.17148/IMRJR.2025.021201

[16]. Crafa, S., Laneve, C., & Sartor, G. (2021). Pacta sunt servanda: Legal contracts in Stipula. ArXiv preprint. arXiv:2110.11069.
[17]. Nay, J. J. (2022). Law Informs Code: A legal informatics approach to aligning artificial intelligence with humans. ArXiv preprint. arXiv:2209.13020.
[18]. Filtz, E., Kirrane, S., & Polleres, A. (2019). The linked legal data landscape. AI and Law, 27(4), 371–408. DOI: 10.1007/s10506-019-09240-6
[19]. European Parliament, Council & Commission. (2015). Joint Practical Guide for persons involved in the drafting of European Union legislation. Publications Office of the European Union. DOI: 10.2880/5575
[20]. Slootweg, P., Rutledge, L., Wedemeijer, L., & Joosten, S. (2016). The implementation of Hohfeldian legal concepts with Semantic Web technologies. In AI for Justice, 22nd European Conference on Artificial Intelligence (ECAI 2016). IOS Press. ISBN: 978-1-61499-610-4
[21]. Bench-Capon, T. (1989). Knowledge Representation: Ontologies and Legal Reasoning. In Proceedings of the 2nd International Conference on Artificial Intelligence and Law (ICAIL 1989), 225–232. ACM. DOI: 10.1145/74539.74567
[22]. Bourcier, D. (2011). A complex-system approach: Legal knowledge, ontology, information, and networks. In P. Casanovas, G. Sartor, M. A. Biasiotti, & G. Ajani (Eds.), Approaches to Legal Ontologies. Springer, 117–132. DOI: 10.1007/978-94-007-0120-5_9
[23]. Garfinkel, R., & Mahoney, M. S. (2005). Law and the Digital Age: Legal Implications of Computing, Communications, and the Internet. Cambridge University Press. (classic reference in legal informatics)
[24]. Ashley, K. D. (2017). Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press.
[25]. Xanthaki, H. (2014). Quality of Legislation: Selected Comparisons. Oxford University Press.
[26]. Mellinkoff, D. (1963). The Language of the Law. Little, Brown & Co.
[27]. Tiersma, P. (1999). Legal Language. University of Chicago Press.
[28]. Solan, L. M. (2010). The Language of Judges. University of Chicago Press.