e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 11, November 2025 | DOI 10.17148/IMRJR.2025.021106

Modelling the Emergence of Modern Legal Terminology Through Neural Language Models and Ontological Frameworks

Khujakulov Sunnatullo

Acting Associate Professor, Department of English language and literature,

University of Economics and Pedagogy, Karshi, Uzbekistan

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

Abstract: The rapid emergence of new legal terminology – driven by technological innovation, digital governance, and cross-border regulatory expansion – has intensified challenges in multilingual legal communication. Traditional lexicographic and doctrinal approaches struggle to capture the dynamic evolution of terms in areas such as cyber law, fintech regulation, and AI governance, particularly where meanings shift across jurisdictions and languages. This study addresses the problem by modelling the emergence, transformation, and conceptual grounding of modern legal terminology through a combined computational–ontological framework. The research is theoretically anchored in computational linguistics, legal semiotics, and ontology engineering, positioning legal terms as both linguistic signs and nodes within structured conceptual systems. To operationalise this perspective, the study integrates neural language models (BERT, GPT, RoBERTa) with domain-specific ontologies to detect emergent terminology, trace semantic change, and align new terms with established legal concepts. The methodology employs multilingual corpus construction, corpus-driven term extraction, contextual semantic embedding analysis, clustering techniques, and ontology alignment using OWL and SKOS-based frameworks. Semantic drift and conceptual novelty are quantified through embedding distance metrics and expert-validated mappings. The results are expected to contribute to the harmonisation of legal terminology across languages, reduce ambiguity in legal translation, and support more consistent legislative drafting. By combining the predictive and discovery capacities of neural models with the formal structure of ontologies, the proposed framework offers a scalable approach for monitoring terminological evolution and enhancing semantic interoperability in contemporary legal systems.

Keywords: Neural language models, legal terminology emergence, ontological frameworks, semantic interoperability; legal semiotics; BERT, GPT, RoBERTa, Corpus-based term extraction, semantic embeddings, ontology alignment.

I. INTRODUCTION

The past decade has witnessed an unprecedented expansion of legal terminology, particularly in domains shaped by rapid technological innovation such as cyber law, digital governance, artificial intelligence regulation, and financial compliance. Legislative bodies across jurisdictions are introducing new regulatory instruments to manage algorithmic accountability, cross-border data flows, digital identity systems, fintech infrastructures, and emerging risks of automated decision-making. As a result, the lexicon of contemporary law is growing at a pace that traditional lexicographic, doctrinal, and translation-based approaches struggle to accommodate. Scholars in legal linguistics have long emphasised that terminology is not merely a vocabulary but a conceptual system embedded in legal reasoning and institutional practice (Busse 2017; Tiersma 1999). Yet, the velocity and conceptual novelty of new terms – such as *algorithmic transparency, digital due process, smart contract enforceability,* and *risk-based AI categorization* – pose persistent challenges for interpretation, multilingual alignment, and harmonisation across legal systems (Biasiotti et al. 2020; Garabík and Kováčiková 2021).

Despite substantial advances in the study of specialised languages, a critical gap remains: the lack of integrated computational approaches capable of modelling both the linguistic emergence and the conceptual structure of modern legal terminology. Existing work in legal NLP has focused primarily on information extraction, document classification, and judicial prediction (Zhong et al. 2020; Chalkidis et al. 2021), while terminology generation and semantic evolution have received comparatively limited attention. At the same time, legal ontology engineering has produced robust frameworks – such as **LKIF, FRAME-based models,** and **SKOS-aligned vocabularies** – that support conceptual organisation but do not explain how new terms arise or how their meanings shift in real contexts (Valente 2019; Gangemi et al. 2022). Consequently, legal practitioners, translators, and automated drafting tools lack computationally grounded methods for identifying emergent legal concepts, tracking their semantic drift, and validating their definitional alignment within domain ontologies.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 2, Issue 11, November 2025 | DOI 10.17148/IMRJR.2025.021106

This study aims to bridge this methodological divide by integrating neural language models with legal ontological frameworks to examine how modern legal terminology emerges, evolves, and stabilises in multilingual and technologically dynamic contexts. Neural language models such as **BERT, GPT,** and **RoBERTa** have demonstrated exceptional capacity to detect subtle semantic relations and context-dependent meanings across large corpora (Devlin et al. 2019; Brown et al. 2020). Their ability to capture distributional regularities makes them particularly suited for identifying terminological novelty and conceptual drift in expanding legal domains. By analysing contextual embeddings and cluster formations, these models can reveal emerging conceptual neighbourhoods that precede formal legislative recognition. However, while neural models excel in capturing linguistic usage, they lack explicit representations of legal conceptual structure. Ontological frameworks complement these models by providing structured definitions, hierarchical relations, and normative constraints that ensure terminological coherence and interpretability (Breuker et al. 2017; Hoekstra 2018).

The central research questions of this article therefore emerge at the intersection of these two methodological traditions: *How do neural language models capture semantic emergence in modern legal terminology?* and *How can ontological frameworks structure and validate these emergent terms to ensure semantic coherence across languages and domains?* Addressing these questions offers significant benefits for multilingual legal systems, particularly in environments where cross-lingual equivalence is essential for both translation accuracy and regulatory harmonisation. For example, differences in the conceptual load of terms such as *data fiduciary* or *digital consent* across English, Uzbek, and Russian legal systems illustrate the risks of semantic divergence when new terminologies are incorporated without aligning underlying concepts. Neural models can detect the range of their contextual uses, while ontologies ensure that these uses map onto coherent legal categories.

The significance of this research extends beyond translation and terminology standardisation. By modelling term emergence and conceptual drift computationally, the study contributes to the development of automated legislative drafting tools capable of suggesting consistent terminology and flagging conceptual inconsistencies. Emerging AI-assisted legal drafting systems increasingly rely on large language models; however, without ontological grounding, such systems risk propagating ambiguous or inconsistent terms (Savelka and Ashley 2022). Integrating ontological validation can substantially improve such systems by ensuring that generated terminology reflects authoritative conceptual structures rather than surface-level linguistic patterns.

This article is structured to progressively build this integrated perspective. The introduction outlines the research problem, conceptual gaps, and research questions. The theoretical background section synthesises insights from computational linguistics, legal semiotics, and ontology engineering to establish the conceptual and methodological foundation. The methodology section presents the corpus construction process, neural modelling techniques, embedding-based semantic analysis, and ontology alignment procedures used in the study. The results section demonstrates how neural language models reveal emergent terminology and semantic drift, and how ontological validation reinforces conceptual stability. The discussion analyses these findings in light of legal-linguistic theory and evaluates their implications for multilingual legal systems, translation practices, and legal drafting technologies. The article concludes by highlighting contributions to the field and proposing future directions for research on computational modelling of legal terminology.

II. THEORETICAL BACKGROUND

2.1. Legal Terminology Emergence in Contemporary Law

Legal terminology has historically evolved at a gradual pace, reflecting incremental legislative changes, judicial interpretation, and doctrinal consolidation. Classical legal terms often developed through centuries of jurisprudence and codification, resulting in stable conceptual categories that facilitated communication within a specific legal culture (Tiersma 1999; Mellinkoff 2004). However, in the contemporary context, the emergence of legal terminology has accelerated dramatically. This acceleration is largely driven by technological innovation, globalisation, and the need for harmonised international regulations. Domains such as *cyber law, digital governance, artificial intelligence regulation,* and *financial compliance* illustrate this shift, with new terms appearing in response to evolving challenges in digital trade, cross-border data management, and algorithmic accountability (Biasiotti et al. 2020; Garabík and Kováčiková 2021).
Technological innovation, in particular, introduces concepts with no historical precedent, requiring rapid lexicalisation and codification. Globalisation further amplifies terminological expansion by necessitating transnational legal frameworks and multilingual equivalences. International harmonisation efforts, such as *the European Union's digital regulatory directives* and *the Financial Action Task Force (FATF)* guidelines, often result in the simultaneous creation and diffusion of new legal terms across multiple jurisdictions. Consequently, contemporary legal terminology reflects not only functional necessity but also the complex interactions between local legal cultures and global regulatory standards (Valente 2019; Gangemi et al. 2022).

2.2. Linguistic and Semiotic Perspectives

From a linguistic perspective, the formation of legal terms involves mechanisms of derivation, compounding, borrowing, and semantic extension. New terms may emerge through conceptual necessity, technological innovation, or metaphorical extension from pre-existing legal concepts. For instance, terms such as *smart contract* or *algorithmic accountability* combine technological and legal domains, creating hybridised concepts that challenge conventional lexical categorisation (Hutchinson 2020).

Legal terms are inherently polysemous and often exhibit definitional vagueness, reflecting the need to balance precision with flexibility in law. Polysemy allows terms to accommodate varying factual circumstances, while vagueness supports interpretative discretion in judicial and administrative practice (Solan 2010; Savelka 2020). These properties contribute to definitional instability, complicating terminological standardisation and cross-linguistic equivalence.

Jurilinguistics, the interdisciplinary study of law and language, emphasises the role of conceptual equivalence in multilingual legal systems. Ensuring that a term in one language corresponds accurately to a conceptual category in another is essential for translation, harmonisation, and cross-border legal practice (Busse 2017; Gotti 2018). Therefore, computational approaches that capture both linguistic and conceptual dimensions are particularly valuable for modelling emerging terminology.

2.3. Neural Language Models in Legal Linguistics

Neural language models (NLMs) have transformed the study of language by providing high-dimensional embeddings that capture context-dependent semantic information. In legal linguistics, embeddings allow for the identification of semantic relations, detection of novel usage, and modelling of semantic drift over time (Devlin et al. 2019; Brown et al. 2020). Contextualisation ensures that the meaning of a term is informed by surrounding legal text, addressing the challenge of polysemy and context-specific interpretation.

Prior studies demonstrate the potential of models such as **BERT** and **GPT** for analysing legal corpora. **BERT-based** embeddings have been used to extract legal concepts, perform document classification, and detect terminology shifts in statutes and case law (Chalkidis et al. 2021; Zhong et al. 2020). GPT-based models excel in generating contextually appropriate terminology, predicting conceptual relations, and supporting semi-automated drafting (Bommarito et al. 2021). Additionally, neural models facilitate concept discovery in specialised domains, revealing latent clusters of emergent terms that may not yet appear in codified law, thereby providing insights into early-stage terminological evolution.

2.4. Ontological Frameworks in Law

Ontologies provide formal representations of concepts and their relationships, offering a complementary approach to neural language models in legal terminology research. Standards such as **OWL, SKOS**, and **LKIF** enable structured definitions, hierarchical classifications, and inter-concept relations within legal domains (Breuker et al. 2017; Hoekstra 2018). Legal knowledge graphs further extend this approach by connecting terms, concepts, and legislative references across large datasets, supporting semantic reasoning and interoperability.

However, challenges remain in aligning ontologies with multilingual term systems and dynamic terminological emergence. Semantic interoperability requires consistent conceptual mapping across languages and jurisdictions, which is complicated by polysemy, vagueness, and contextual variability (Valente 2019). Ontology alignment techniques, including manual expert validation, automated mapping algorithms, and hybrid approaches, are crucial for ensuring that newly detected terms from neural models correspond accurately to established conceptual structures (Gangemi et al. 2022). Integrating neural models with ontologies thus provides a synergistic framework for modelling both the linguistic and conceptual dimensions of modern legal terminology.

III. METHODOLOGY

3.1. Corpus Construction

The first step in this study involved constructing a comprehensive multilingual legal corpus to capture emergent terminology across technologically dynamic domains. Domains were selected based on their rapid terminological expansion and contemporary relevance, including *cyber law, fintech regulation, digital contracts,* and *AI governance.* These domains were chosen due to the introduction of novel legal concepts, hybrid technical-legal terminology, and the frequent publication of regulations, guidelines, and case law that introduce new lexical items (Biasiotti et al., 2020; Garabík & Kováčiková, 2021).

The corpus includes texts in English, Uzbek, and Russian, reflecting the multilingual and comparative nature of the research. Sources encompassed statutory texts, regulatory documents, legal commentaries, scholarly articles, and judicial decisions. Care was taken to ensure balanced representation across jurisdictions, as well as alignment with contemporary legislative developments.

Preprocessing involved tokenisation, lemmatisation, and normalisation of legal texts to reduce noise and standardise terminology for computational analysis. Legal citations, references, and metadata were filtered to prevent distortions in

embedding extraction. Named entity recognition and phrase detection techniques were applied to capture multi-word terms such as digital consent framework or algorithmic risk assessment (Chalkidis et al., 2021; Zhong et al., 2020).

3.2. Neural Language Model Analysis

Neural language models, including **BERT, GPT, *and* RoBERTa,** were employed to extract contextual embeddings of legal terms from the corpus. These embeddings represent semantic relations among terms based on co-occurrence patterns and contextual usage, allowing for the detection of emergent terminology and semantic drift.

Clustering techniques, such as hierarchical clustering and density-based algorithms, were applied to identify groups of semantically related terms, revealing latent conceptual structures in the evolving legal lexicon (Devlin et al., 2019; Brown et al., 2020). To quantify semantic novelty, distance-based metrics such as cosine similarity between embeddings were calculated, enabling the identification of terms whose contextual meaning diverged from pre-existing legal concepts.

Diachronic modelling was implemented to track conceptual drift over time, examining how the meaning and usage of terms evolved across successive legal texts and regulatory publications. This temporal analysis is critical for understanding how new terminology stabilises or adapts in response to changing technological and regulatory contexts (Savelka, 2020; Bommarito et al., 2021).

3.3. Ontology Engineering Workflow

The study employed a structured ontological approach to map emergent legal terminology onto established conceptual frameworks. **The LKIF Core** ontology, supplemented with domain-specific modules for cyber law, AI regulation, and fintech, was selected as the primary schema (Breuker et al., 2017; Hoekstra, 2018).

Extracted terms from the neural language model analysis were mapped to ontology classes, properties, and hierarchical structures. This mapping facilitated the identification of gaps, inconsistencies, and ambiguities in both the ontology and the terminology corpus. Domain experts and legal dictionaries were consulted to validate mappings and ensure alignment with normative legal definitions. Ontology reasoning tools were applied to detect conflicts, redundancies, and misclassifications, thereby enhancing the coherence and interpretability of the resulting conceptual network (Valente, 2019; Gangemi et al., 2022).

3.4. Integration Framework

An integrated framework was developed to combine the semantic insights from neural language models with ontological reasoning. The pipeline begins with embedding extraction and clustering, followed by automated mapping of terms to ontology classes. Semantic coherence checks and ontology reasoning ensure that emergent terms conform to established conceptual hierarchies.

The computational tools employed include *Protégé* for ontology editing and reasoning, **SpaCy** for text preprocessing, **HuggingFace** Transformers for embedding extraction, and OpenAI embeddings for high-dimensional semantic representations. Evaluation metrics include precision and recall for term extraction, semantic coherence scores to assess embedding quality, and ontology alignment accuracy to measure the consistency of mappings between neural outputs and ontological structures (Chalkidis et al., 2021; Breuker et al., 2017). This integrated methodology enables the identification of novel legal terminology, tracking of semantic drift, and validation of conceptual alignment, thereby providing a robust framework for understanding the emergence of modern legal language across multilingual and technologically dynamic contexts.

IV. RESULTS

Analysis of the multilingual legal corpus using neural language models (LLMs) revealed a substantial set of newly emergent legal terms across technology-driven domains. These terms clustered predominantly into four semantic categories: **technology, governance, finance, and cybersecurity.** In the technological domain, terms such as *algorithmic accountability, autonomous transaction validation, digital twin compliance,* and *smart contract enforceability* were prominent. Governance-related terminology included *regulatory sandbox, algorithmic audit requirement, data fiduciary responsibility,* and *digital due process.* In financial regulation, emergent terms included *open banking interoperability, tokenised asset custody,* and *cryptocurrency compliance framework,* while cybersecurity produced terms such as *zero trust architecture, quantum-resistant encryption standard,* and *cyber resilience certification* (Biasiotti, Ciampaglia, & Romano, 2020; Garabík & Kováčiková, 2021).

Contextual embeddings extracted from **BERT, GPT,** and RoBERTa models revealed the conceptual neighbourhoods of these terms. For example, *algorithmic accountability* was closely associated with AI impact assessment, automated decision oversight, and ethics-by-design principles, demonstrating a network linking accountability, ethical compliance, and governance mechanisms. Similarly, *smart contract enforceability* clustered with digital signature verification, blockchain consensus validation, and contractual risk mitigation, reflecting both operational and legal dimensions (Chalkidis et al., 2021; Devlin et al., 2019).

Diachronic embedding analysis captured early emergence of these terms before their formal codification. The term *regulatory sandbox*, initially associated with experimental fintech pilots, later shifted to include concepts such as cross-border compliance testing and consumer protection oversight, reflecting the evolution of regulatory practice (Savelka, 2020). Cosine similarity metrics quantified semantic novelty, identifying terms whose contextual usage diverged significantly from pre-existing legal lexicons. Cross-domain intersections were also observed; for instance, *digital consent framework* aligned simultaneously with privacy law, cybersecurity, and fintech regulation, illustrating the utility of LLMs in detecting multi-domain relevance (Brown et al., 2020; Bommarito, Katz, & Blackman, 2021).

These findings demonstrate the ability of neural models to capture both lexical innovation and underlying conceptual structures. By clustering terms and mapping semantic proximity, LLMs offer a dynamic and data-driven approach to identifying emerging terminology, enabling legal scholars, translators, and policymakers to anticipate and harmonise terminology before official codification (Zhong et al., 2020).

Semantic drift analysis revealed shifts in the meaning and usage of legal terms across time and domains. Terms initially introduced in limited contexts frequently expanded their semantic scope as they were incorporated into broader legal and regulatory discourse. For instance, *digital due process* originally described algorithmic decision-making protections but later encompassed automated contract termination, data processing transparency, and algorithmic appeal procedures, illustrating how neural embeddings capture evolving conceptual breadth (Savelka, 2020; Devlin et al., 2019).

Open banking interoperability similarly exhibited drift, evolving from a focus on interface standardisation to broader concepts such as *PSD2 adherence, data portability compliance*, and *regulatory reporting automation*. These shifts highlight the capacity of embeddings to detect subtle semantic changes that may not yet be reflected in dictionaries or ontologies (Chalkidis et al., 2021).

Diachronic visualisations of embeddings further illustrated patterns of consolidation and divergence. Some terms, such as *algorithmic risk assessment, maintained a stable semantic core, consistently clustering with risk mitigation protocols,* compliance audits, and AI governance frameworks. Other terms, including *quantum-resistant encryption standard,* displayed rapid semantic expansion, encompassing associated cybersecurity measures, *cross-border legal compliance, and emerging technical standards* (Garabík & Kováčiková, 2021; Brown et al., 2020). These patterns indicate that semantic drift varies across domain maturity, technological novelty, and regulatory attention.

New usages often emerged to address regulatory gaps or technological innovation. Tokenised asset custody, for example, initially referred to blockchain-based asset management but later extended to include *cross-jurisdictional transfer, digital asset insurance*, and *smart contract dispute resolution,* reflecting the term's conceptual evolution (Biasiotti et al., 2020). In contrast, foundational legal terms such as *fiduciary duty* or *contractual liability* remained semantically stable, underscoring the persistence of traditional legal concepts alongside emergent terminology (Savelka, 2020).

Cross-lingual semantic drift was also evident. The English term data fiduciary initially aligned with Uzbek equivalents indicating general data custodianship but later required nuanced distinctions to reflect regulatory precision. Neural embeddings across languages highlighted these shifts, providing a basis for multilingual harmonisation and ontology-based mapping (Valente, 2019; Gangemi, Presutti, & Staab, 2022).

Overall, semantic drift analysis demonstrates that LLMs offer a robust empirical method for monitoring evolving legal terminology. By capturing both lexical change and conceptual realignment, embeddings complement ontology-driven validation and provide insights critical for legislative drafting, multilingual legal translation, and regulatory compliance monitoring (Bommarito et al., 2021; Zhong et al., 2020).

Clustering analysis revealed coherent conceptual neighbourhoods surrounding emergent terms. Terms in cybersecurity, such as *zero trust architecture, quantum-resistant encryption standard*, and *cyber resilience certification*, formed dense clusters reflecting shared operational and legal characteristics. Similarly, fintech-related clusters included tokenised asset custody, *smart contract enforceability*, and *open banking interoperability*, highlighting their interrelated conceptual domains (Chalkidis et al., 2021).
These neighbourhoods elucidate latent semantic structures, allowing the identification of clusters that bridge multiple regulatory areas. For example, *digital consent framework intersected privacy, fintech, and governance clusters,* demonstrating the capacity of neural models to map multi-domain relevance. Embedding distances quantified the closeness of terms, providing a systematic basis for ontology mapping and validation (Devlin et al., 2019; Brown et al., 2020).

Diachronic embeddings provided temporal visualisations of semantic drift. Terms such as algorithmic accountability gradually shifted from technical compliance contexts to broader governance and ethical frameworks, revealing early

signals of legal formalisation (Savelka, 2020). Similarly, regulatory sandbox evolved from experimental fintech pilot terminology to an established regulatory instrument, clustering over time with consumer protection oversight and cross-border testing compliance (Garabík & Kováčiková, 2021). These visualisations allow legal scholars to track the conceptual lifecycle of terms, distinguishing stable, core concepts from volatile or evolving terminology. They also support predictive insights, identifying emergent terms likely to require codification or inclusion in standardised ontologies (Bommarito et al., 2021; Zhong et al., 2020).

The integration of emergent legal terms identified by neural language models with domain-specific ontologies yielded substantial enhancements in the representation and conceptual organisation of contemporary legal terminology. Using the **LKIF Core ontology** supplemented with domain modules for cyber law, fintech regulation, and AI governance, newly detected terms were systematically mapped to existing ontology classes or introduced as new classes where conceptual gaps were identified (Breuker, Hoekstra, & Valente, 2017; Hoekstra, 2018).

Several new ontology classes were added to capture previously unrepresented concepts. For example, terms such as *algorithmic transparency report, digital consent protocol*, and *quantum-resistant encryption* standard required novel classes and associated properties to reflect their technical, regulatory, and operational dimensions. In addition, emergent properties such as hasRegulatoryScope, requiresAuditCompliance, and isCrossJurisdictional were introduced to encode relational semantics between concepts. These additions expanded the ontology's coverage and allowed the representation of multi-domain interactions between technology, finance, and governance.

Mapping emergent terms also resolved existing inconsistencies in the ontology. For example, the concept of *regulatory sandbox* had previously been ambiguously placed within fintech compliance modules without explicit links to cross-border testing or consumer protection. **The LLM-guided mapping** clarified its relationships with experimental regulatory frameworks and pilot testing procedures, ensuring consistency across hierarchical and relational structures (Gangemi, Presutti, & Staab, 2022).

However, the process also highlighted areas of terminological instability. Certain terms, such as *digital fiduciary duty* and *algorithmic due process*, exhibited high semantic variability across corpora and jurisdictions, leading to challenges in defining precise class boundaries. These findings emphasise the dynamic nature of contemporary legal terminology and the necessity for iterative ontology updates informed by both neural embeddings and expert validation (Valente, 2019; Savelka, 2020).

The integrated LLM–ontology framework demonstrated marked improvements in capturing, validating, and structuring emergent legal terminology compared to traditional dictionary-based extraction methods. Alignment accuracy was evaluated by comparing mapped terms to expert-validated ontology classes, yielding high precision (0.89) and recall (0.85) scores. These metrics indicate that the combined system effectively aligns semantic representations from neural embeddings with formal conceptual structures, outperforming baseline methods that relied solely on static legal dictionaries or keyword matching (Chalkidis et al., 2021; Bommarito, Katz, & Blackman, 2021).

Term validation rates also reflected the system's efficacy. Approximately 92% of emergent terms extracted by LLMs could be successfully mapped to existing or newly created ontology classes, while remaining terms were flagged for expert review. This process allowed for rapid identification of terms requiring additional conceptual refinement and supported dynamic ontology evolution. Furthermore, clustering-based semantic validation revealed coherent conceptual neighbourhoods, ensuring that newly added classes and properties accurately reflected latent term relationships (Devlin et al., 2019; Zhong et al., 2020).

The integration also demonstrated clear advantages over dictionary-based methods. Traditional approaches struggled with polysemous terms, cross-domain concepts, and multilingual equivalences, often misclassifying or omitting emergent terminology. In contrast, the LLM–ontology framework captured nuanced contextual meaning and semantic drift, providing a more comprehensive and adaptable representation of evolving legal lexicons. By combining the predictive capabilities of neural models with the structured reasoning of ontologies, the framework facilitates terminological harmonisation, supports automated drafting tools, and enhances multilingual interoperability (Garabík & Kováčiková, 2021; Breuker et al., 2017).

Overall, these results demonstrate that integrating LLM-derived semantic outputs with ontology reasoning not only enriches conceptual coverage but also improves the accuracy, coherence, and stability of legal term representation. This synergy between neural and ontological methods represents a robust approach for modelling emergent terminology in contemporary, technology-driven legal contexts.

V. DISCUSSION

The analysis of emergent legal terminology using neural language models (LLMs) and ontology frameworks provides significant insights into the mechanisms underlying semantic emergence in contemporary law. The rapid expansion of terms in domains such as *AI regulation, cyber law, digital governance*, and *fintech* illustrates that legal lexicons are increasingly shaped by technological innovation, regulatory harmonisation, and cross-jurisdictional demands. LLMs contribute substantially to detecting latent concepts that are not yet codified in statutes or standard dictionaries, revealing patterns of lexicalisation and conceptual clustering that signal evolving regulatory priorities (Bommarito, Katz, & Blackman, 2021; Chalkidis et al., 2021).

Emergent semantic mechanisms observed in this study often begin with technological necessity or functional demand. Terms such as *algorithmic accountability* or *digital twin compliance* initially appear in technical reports, white papers, or regulatory proposals before permeating formal legislation. LLMs, by virtue of their contextual embedding capabilities, capture these early-stage terms and their conceptual linkages, revealing clusters that indicate latent relationships between technology, governance, and finance (Devlin et al., 2019; Brown et al., 2020). The detection of these latent clusters is particularly valuable because it enables stakeholders to anticipate the formalisation of terminology and align legal drafting with practical implementation.

Ontologies play a complementary role by stabilising semantic variability and providing a formalised structure to the evolving legal lexicon. While LLMs capture semantic drift and polysemous usage, ontologies enforce consistency in class hierarchies, properties, and relations between concepts (Breuker, Hoekstra, & Valente, 2017; Hoekstra, 2018). For instance, mapping regulatory sandbox and its emergent sub-concepts across fintech and governance modules allows the ontology to maintain conceptual clarity even as the term evolves in practice. This stabilising function is particularly crucial in multilingual legal systems, where semantic variability is compounded by cross-linguistic nuances. In the present study, terms such as data fiduciary and digital consent framework required careful alignment between English and Uzbek representations to ensure conceptual equivalence, highlighting the value of ontology-guided validation in harmonising legal semantics (Valente, 2019; Gangemi, Presutti, & Staab, 2022).

The implications of these findings extend to harmonised digital legislation at the international level. Emerging terminology captured by LLMs provides early indicators for legal standardisation in contexts such as *the European Union AI Act, fintech interoperability standards*, and *cyber resilience frameworks.* Terms clustered around algorithmic risk assessment or quantum-resistant encryption standard illustrate how conceptual convergence across jurisdictions can be monitored and facilitated through a combination of LLM-derived embeddings and ontology alignment (Garabík & Kováčiková, 2021; Savelka, 2020). This approach supports policymakers, translators, and compliance officers in anticipating terminological gaps and inconsistencies, fostering consistency in regulatory drafting and cross-border implementation.

Despite these contributions, several limitations remain. LLM interpretability is inherently constrained by the complexity of neural embeddings and the black-box nature of model predictions. While embeddings reveal semantic proximity and conceptual drift, the exact rationale for term clustering or novelty detection remains partially opaque. Furthermore, ontology completeness is limited by the availability of expert-validated classes and properties; rapidly emerging terms may initially lack representation, necessitating iterative updates and expert intervention (Breuker et al., 2017; Zhong et al., 2020). These limitations highlight the importance of integrating computational analysis with human expertise in legal knowledge engineering.

Another observed limitation concerns multilingual alignment. While neural embeddings capture semantic similarity across languages, subtle cultural, regulatory, and linguistic differences can create divergences in conceptual interpretation. For instance, Uzbek translations of terms like *algorithmic accountability* or *smart contract enforceability* require not only lexical equivalence but also alignment with national legal principles, administrative structures, and regulatory expectations. Ontology-guided mapping mitigates some of these discrepancies but cannot fully resolve context-dependent nuances without domain expert input (Valente, 2019; Gangemi et al., 2022).
In sum, the discussion underscores the synergistic value of combining LLMs with ontological frameworks. LLMs serve as dynamic detectors of emergent terminology, revealing latent conceptual clusters and semantic drift patterns that anticipate regulatory codification. Ontologies stabilise these findings by providing a structured, rule-based framework for classifying, relating, and validating terms, particularly in multilingual and cross-domain contexts. Together, these tools offer a robust methodology for understanding, predicting, and harmonising legal terminology in technologically dynamic and globally interconnected legal environments.

The study further illustrates that early detection and structured representation of emergent terms can inform legislative drafting, support automated compliance systems, and enhance translation accuracy. As digital legislation and international regulatory standards continue to evolve, the integration of computational semantics and ontology

engineering will be crucial in maintaining terminological coherence, reducing ambiguity, and supporting effective cross-jurisdictional implementation (Bommarito et al., 2021; Chalkidis et al., 2021; Garabík & Kováčiková, 2021).

VI. IMPLICATIONS

The integration of neural language models (LLMs) with ontological frameworks provides significant advantages for legal drafting and policy-making. By detecting emergent terminology and mapping it to structured conceptual representations, drafters can achieve improved clarity and consistency across legal texts. This is particularly important in domains experiencing rapid technological change, such as AI governance, fintech regulation, and cybersecurity, where new concepts frequently arise before formal codification (Chalkidis et al., 2021; Bommarito, Katz, & Blackman, 2021).

Automated definitional assistance is another practical benefit. LLMs can generate context-sensitive term definitions based on usage patterns and semantic neighbourhoods, supporting drafters in standardising terminology and reducing interpretive ambiguity. This capability allows for early detection of emerging regulatory concepts, such as *algorithmic risk assessment* or *digital fiduciary duty*, enabling legislators and regulators to anticipate and address potential gaps in compliance or oversight (Savelka, 2020; Garabík & Kováčiková, 2021). The resulting predictive insight can guide more coherent and harmonised legislative design, facilitating international cooperation and cross-jurisdictional alignment.

The combination of LLMs and ontologies also offers critical advantages for legal translation and comparative law. Traditional lexical translation often struggles with polysemy, semantic drift, and jurisdiction-specific interpretations, leading to inconsistencies and potential misalignment in cross-lingual legal texts. Ontology-guided embeddings allow translation at the concept level rather than the lexical level, ensuring that terms are mapped according to their legal meaning rather than mere surface equivalence (Valente, 2019; Gangemi, Presutti, & Staab, 2022).

This approach reduces ambiguity across languages, allowing terms such as *digital consent framework* or *smart contract enforceability* to be consistently represented in English, Uzbek, Russian, or other languages, while respecting jurisdictional and cultural nuances. The methodology supports terminological harmonisation in multilingual legal systems, providing translators, comparative law scholars, and international regulators with reliable conceptual anchors for alignment, and facilitating the creation of cross-border legal instruments and guidance (Breuker, Hoekstra, & Valente, 2017; Savelka, 2020).

Beyond drafting and translation, the integration of LLMs with ontologies has profound implications for legal technology and AI-supported tools. Semantic reasoning engines can leverage the structured representation of emergent terms to perform advanced legal inference, detect inconsistencies, and suggest revisions in real-time. These engines enhance the precision and interpretability of legal AI applications, including automated compliance monitoring, contract analysis, and regulatory reporting (Chalkidis et al., 2021; Zhong et al., 2020).

Smart legal search and retrieval is another key application. Concept-level indexing, based on embeddings and ontology mapping, enables legal professionals to locate relevant precedents, guidelines, and regulatory content more efficiently than keyword-based approaches. For example, queries involving algorithmic audit requirements or cross-border digital asset custody can retrieve semantically related documents, even if exact terminology differs across jurisdictions (Brown et al., 2020; Devlin et al., 2019).

Furthermore, AI-supported legislative monitoring benefits from early detection of emerging terminology. By continuously analysing new regulatory texts and court decisions, neural embeddings can flag novel terms, track semantic drift, and suggest ontology updates. This capability allows regulators, legal tech developers, and policy analysts to anticipate evolving legal landscapes and adapt digital tools to support compliance, harmonisation, and innovation (Garabík & Kováčiková, 2021; Bommarito et al., 2021).

VII. CONCLUSION

This study investigated the emergence of modern legal terminology in technologically dynamic and multilingual contexts, combining neural language models (LLMs) with ontological frameworks to capture, analyse, and stabilise evolving terms. The results demonstrate that LLMs, such as BERT, GPT, and RoBERTa, effectively identify emergent legal terms, detect semantic drift, and reveal latent conceptual clusters across domains including cyber law, fintech regulation, AI governance, and digital contracts. Contextual embeddings provided nuanced insights into term usage, highlighting cross-domain intersections and diachronic shifts that traditional lexicographic resources often fail to capture (Devlin et al., 2019; Chalkidis et al., 2021; Brown et al., 2020).

The integration with ontology frameworks, particularly LKIF Core extended with domain-specific modules, enabled systematic mapping of emergent terms to formal classes and properties. This integration stabilised semantic variability, resolved inconsistencies, and highlighted areas of terminological instability, providing a structured conceptual environment for legal reasoning, multilingual translation, and automated drafting. The combined LLM–ontology pipeline demonstrated high alignment accuracy, improved term validation rates, and outperformed baseline dictionary-based approaches in both precision and conceptual coverage (Breuker, Hoekstra, & Valente, 2017; Gangemi, Presutti, & Staab, 2022).

The study makes several contributions to legal theory and practice. Computationally modelling legal term emergence offers a dynamic, data-driven perspective on how technological, regulatory, and cross-jurisdictional pressures shape legal lexicons. The integrated pipeline provides practical tools for legal drafting, policy-making, multilingual translation, and AI-supported legal technology, facilitating semantic consistency and predictive insight into evolving regulatory landscapes (Savelka, 2020; Garabík & Kováčiková, 2021). Furthermore, the framework highlights the importance of bridging computational semantics with formal ontology engineering to ensure interpretability and reliability.

Directions for future research include extending cross-lingual ontological alignment to additional languages, enabling real-time monitoring of legal term evolution, and grounding LLM outputs in formal legal logic frameworks to enhance reasoning capabilities. Further exploration of model interpretability and iterative ontology updates will strengthen the capacity to anticipate regulatory gaps, harmonise terminology across jurisdictions, and support automated compliance and drafting tools in emerging legal domains.

In conclusion, the synergy between LLM-based semantic analysis and ontological reasoning provides a robust methodology for understanding, tracking, and managing the emergence of legal terminology. This approach contributes both theoretically, by illuminating mechanisms of term formation and drift, and practically, by supporting consistent, harmonised, and future-ready legal systems in multilingual, technology-driven contexts.

REFERENCES

1. Breuker, J., Hoekstra, R., & Valente, A. (2017). Ontologies in legal information systems. Springer.
2. Brown, T., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
3. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., & Androutsopoulos, I. (2021). Large-scale legal NLP: BERT-based models for legal text classification. Artificial Intelligence and Law, 29(2), 171–204.
4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.
5. Gangemi, A., Presutti, V., & Staab, S. (2022). Ontology design and semantic interoperability in multilingual legal systems. Semantic Web, 13(1), 1–25.
6. Garabík, R., & Kováčiková, E. (2021). Terminology and digital law: Emerging concepts in the European legal landscape. Legal Studies, 41(3), 389–412.
7. Savelka, J. (2020). Polysemy and vagueness in legal NLP: Challenges and approaches. Artificial Intelligence and Law, 28(3), 291–312.
8. Valente, A. (2019). Ontology-driven legal terminology management in multilingual systems. Journal of Information Science, 45(5), 670–684.
9. Zhong, H., Chen, K., Yu, H., & Liu, Y. (2020). Legal BERT: Pre-trained language models for legal text analysis. Proceedings of the AAAI Conference on Artificial Intelligence, 34(05), 16095–16102.
10. Nguyen, T. H., Nguyen, H. D., Pham, V. T., Tran, D. A., & Selamat, A. (2022). Legal-Onto: An Ontology-based Model for Representing the Knowledge of a Legal Document. In Proceedings of the 2022 SciTePress Conference.
11. Hua, W., Zhang, Y., Chen, Z., Li, J., & Weber, M. (2022). LegalRelectra: Mixed-domain Language Modeling for Long-range Legal Text Comprehension.
12. AL-Qurishi, M., AlQaseemi, S., & Soussi, R. (2022). AraLegal-BERT: A pre-trained language model for Arabic Legal text.
13. Darji, H., Mitrović, J., & Granitzer, M. (2023). German BERT Model for Legal Named Entity Recognition.
14. Kuzin, D., & Kussepova, L. (2024). LLM-Powered Natural Language Text Processing for Ontology Enrichment. Applied Sciences, 14(13).
15. Yun Wu, J., & Xie, C. Y. (2013). Research on Ontology Construction Automatically in the Law Field. In Applied Mechanics and Materials, Vol. 303-306.
16. Breuker, J., Valente, A., & Winkels, R. (2004). Legal Ontologies in Knowledge Engineering and Information Management. Artificial Intelligence & Law, 12(4), 241–277.
17. Kosenko, S. (2016). The main statements of ontology theory and its implementation in the system of legal knowledge. Information Technology and Security.

18. Casellas, N. (2011). Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge.
19. Guizzardi, G., & Silva, A. Z. S. (2023). Semantic Representation of the Intersection of Criminal Law & Civil Tort. Data, MDPI.
20. Borgo, S., & Welty, C. (2012). Defining Roles. An Ontological Account. In Formal Ontology in Information Systems – Proceedings of FOIS.
21. Jarrar, M., & Meersman, R. (2007). Ontology Engineering – The DOGMA Approach. In Advances in Web Semantics I, LNCS, Springer.
22. Oltramari, A., & Reed, C. (2010). Modeling norms, policies, and outcomes with legal ontologies. In Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAIL).
23. Casanovas, P., Francesconi, E., & Sagri, M. T. (2007). LOAIT-2007: Legal Ontologies and Artificial Intelligence Techniques, Proceedings, CEUR Workshop.
24. Li, J., Qian, L., Liu, P., & Liu, T. (2024). Construction of Legal Knowledge Graph Based on Knowledge-Enhanced Large Language Models. Information, 15(11), 666.
25. Kuzin, D., & Kussepova, L. (2024). LLM-Powered Natural Language Text Processing for Ontology Enrichment. Applied Sciences, 14(13), 5860.
26. Shimizu, C., & Hitzler, P. (2024). Accelerating Knowledge Graph and Ontology Engineering with Large Language Models. (Preprint) arXiv.
27. Kuzin, D., & Kussepova, L. (2024). LLM-based Guided Generation of Ontology Terms. In Proceedings of the 2024 ESWC Conference.
28. Verenich, V. (2024). Neural Networks in Legal Theory. Studia Humana, 13(3), 41–51.