Designing a Semantic Analysis Framework for Intelligent Learning by Reading Systems Using Advanced Text Comprehension Techniques

Chinedu Okafor[1] and Aisha Bello[2]

[1]Abia State University, Department of Computer Science, Uturu-Okigwe Road, Uturu, Abia State, Nigeria

[2]Federal University of Technology Minna, School of Information and Communication Technology, Gidan Kwano Campus, Minna, Niger State, Nigeria

2024

Abstract

This paper presents a comprehensive exploration into the development of a semantic analysis framework aimed at enhancing intelligent learning by reading systems through advanced text comprehension techniques. The proposed framework seeks to integrate lexical, syntactic, and semantic representations into a cohesive model for improved natural language understanding, ultimately enabling automated systems to interpret, reason about, and respond to complex textual materials. By examining various theoretical underpinnings, including formal logic constraints and deep semantic embeddings, this study delves into the methods for capturing the layered and context-sensitive nature of human language. Multiple methods are explored, ranging from the incorporation of structured representations rooted in predicate logic to the application of high-dimensional vector space approaches that account for semantic relations between words, phrases, and larger discourse units. Furthermore, these methodologies are evaluated through implementation in prototype systems, where performance metrics highlight their potential effectiveness. Empirical findings underscore the capacity of the framework to interpret textual data with nuanced comprehension while addressing ambiguity, polysemy, and the subtleties inherent in domain-specific language usage. This contribution offers researchers and practitioners critical insights into the design of advanced text processing pipelines that promote intelligent, knowledge-driven learning experiences in computer-assisted educational environments and automated reading comprehension tasks. By unifying logical formalisms with robust statistical modeling, the presented work serves as a benchmark for future innovations in intelligent textual analysis.

1 Introduction

Over the last few decades, the need for sophisticated text comprehension systems has grown as digital content continues to expand at an exponential rate, demanding automated approaches to parse and interpret textual information [1]. Intelligent learning by reading systems, which aim to mimic or approximate the human ability to understand written material, have emerged as an essential area of inquiry within natural language processing [2]. The ultimate aspiration is not merely to retrieve relevant data from large textual corpora but to develop technology that processes, interprets, and reasons about textual content in a way that closely aligns with human comprehension. Thus, the design of a semantic analysis framework that can capture the rich layers of context in text—spanning from lexical nuances to complex logical structures—remains a pressing challenge. [3]

A robust framework for semantic analysis should account for the complex interplay of lexical and syntactic phenomena while also embracing the depth and ambiguity of meaning embedded within different forms of discourse. Approaches to text comprehension have historically ranged from rule-based systems, reliant on predefined grammars and manually curated knowledge bases, to contemporary deep learning methods that leverage large-scale corpora and distribute semantic information into high-dimensional vector spaces [4]. However, such methodologies vary widely in their handling of linguistic properties such as word order, syntactic construction, ambiguity resolution, and domain-specific terminologies. Moreover, scaling these methods to handle the ever-increasing volume and diversity of textual material requires novel strategies that maintain precision and reduce computational overhead.

One of the earliest paradigms in machine reading comprehension involved transforming text into logical predicates and employing automated theorem provers or inference engines to derive conclusions or validate statements. In such logic-based systems, a statement of the form $\forall x \in \mathcal{D}, \exists y \in \mathcal{C} : R(x, y)$ might be used to model the existence of certain relationships within text passages, enabling the system to make inferences about the content. However, purely symbolic approaches often struggle to capture the subtleties of everyday language, especially when dealing with polysemous terms or idiomatic expressions that do not conform cleanly to rigid logical structures. [5]

Statistical and machine learning methodologies emerged to address these limitations by harnessing large annotated datasets to learn models of language behavior. Concepts like distributional semantics, wherein word meaning is inferred from context, introduced the possibility of capturing lexical and semantic relationships through latent representations [6]. Techniques such as factorization of word co-occurrence matrices or the creation of embeddings in a vector space $\mathbf{w}_i \in \mathbb{R}^d$ for word w_i enabled more nuanced correlations to be discovered. These methods set the stage for modern neural architectures that rely on sequence processing, including recurrent neural networks and attention-based transformers, to derive context-sensitive embeddings that represent entire sentences or even long passages of text. Yet, while effective in many tasks, these approaches often sacrifice interpretability and struggle to integrate complex logical reasoning into their representations. [7]

As neural models continue to evolve, there is a growing recognition of the need to fuse their raw representational power with explicit mechanisms for logical inference. Advancements in knowledge representation using graph-based data structures, such as knowledge graphs, have introduced ways to encode relationships between entities, thereby facilitating more powerful reasoning engines [8], [9]. When combined with distributed word embeddings, knowledge graph representations offer an avenue to bind symbolic and sub-symbolic information into a hybrid semantic layer [10]. This hybrid approach can be formalized in part by statements of the form $\Phi(\mathbf{v}_x, \mathbf{v}_y) \to \Psi(x, y)$, where $\mathbf{v}_x$ and $\mathbf{v}_y$ are embedding vectors tied to entities x and y, respectively, and Ψ denotes an inferred relationship in the knowledge graph.

Despite these advances, creating a holistic framework for intelligent learning by reading still involves numerous challenges. Domain specificity poses one major obstacle, as medical, legal, scientific, or other specialized texts employ distinct vocabularies and syntactic constructs that may differ significantly from the general language [11]. Handling anaphoric expressions, metaphors, nested clauses, and discourse-level phenomena remains an ongoing difficulty. Moreover, ensuring that a model can handle contradictory statements or ambiguous contexts typically requires an approach that combines statistical robustness with symbolic clarity [12]. Such hybrid strategies demand carefully curated training data and, in some cases, the development of domain- or task-specific ontologies.

In the educational sphere, interactive systems that support learning by reading often require not just the ability to answer direct questions but to generate instructive feedback, highlight key ideas, summarize material, and detect inconsistencies or misinformation [13]. The assessment of user comprehension might involve the generation of follow-up questions or the provision of explanations and clarifications, features that demand nuanced semantic analyses. In these contexts, it becomes increasingly important for frameworks to provide justifications or logical derivations that humans can interpret, thereby meeting a crucial standard of transparency and trust. [14]

From a methodological standpoint, implementing a semantic analysis framework entails several layers of functionality: text ingestion and preprocessing, syntactic parsing, semantic transformation, and inference or retrieval. Specialized modules may operate on lexemes, part-of-speech tags, dependency structures, or embedding vectors [15]. Intermediate representations might include parse trees, logical forms, or adjacency matrices for graph-based encodings. Evaluating such multi-layered systems often requires metrics that go beyond simple accuracy scores or perplexity [16]. Metrics of explanatory adequacy, robustness to domain shifts, and interpretability can help illuminate strengths and weaknesses. Additionally, computational complexity and the feasibility of scaling to large document collections become salient factors in the design of practical systems. [17]

The present work seeks to merge these diverse strands into a coherent theoretical and applied framework, offering both a conceptual model and an experimental validation [18]. In the sections that follow, we explore the theoretical foundations that inform semantic analysis, present the proposed framework with details on how advanced text comprehension can be operationalized, and examine results from an implementation in a prototype reading comprehension system. By weaving together symbolic formalism and sub-symbolic learning, the framework aspires to strike a balance between depth of understanding and adaptability, reinforcing its potential for future expansions into new tasks and domains [19], [20].

2 Theoretical Foundations

The impetus behind an intelligent learning by reading system is rooted in the desire to endow machines with capabilities that parallel, in some measure, the depth and flexibility of human comprehension. To achieve this, it is instructive to anchor the framework in well-established theoretical underpinnings from multiple fields, including linguistics, cognitive science, logic, and computer science [21]. An integrated perspective draws upon syntactic theories that shape parsing strategies, lexical semantics to handle word-level meaning, and formal logic to manage inference, while also taking advantage of high-capacity statistical models that capture distributional patterns. The central premise is that semantic understanding arises from a synergy between symbolic structures and data-driven regularities, motivating a layered architecture that can accommodate varied types of information. [22], [23]

Among the earliest influences on computational linguistics is the Chomskyan tradition, which introduces generative grammars as a means to describe how sentences can be formed from smaller units. While generative grammars, such as context-free or context-sensitive grammars, offer a formal mechanism for generating syntactically valid strings of symbols, they often provide limited insight into semantic content [24]. Even extensions that integrate semantic features into syntactic rules face difficulties when attempting to handle ambiguity, idiomatic usage, or polysemy. Nonetheless, the rules-based approach presents one end of the theoretical spectrum, emphasizing explicit structures and constraints [25]. In symbolic systems, meaning is encoded through formal predicates and functions, such that a sentence like "All students read some books" might be translated into something akin to $\forall s(Student(s) \rightarrow \exists b(Book(b) \land Read(s, b)))$. This representation is powerful for reasoning tasks but can be brittle when confronting language phenomena that deviate from neat logical forms. [26]

On the other end of the spectrum lie distributional models of semantics, grounded in the principle that "you shall know a word by the company it keeps." Such approaches represent words as vectors in a high-dimensional space, derived from co-occurrence statistics within large text corpora. The geometry of this space encodes semantic relationships: words with similar contexts tend to cluster together, while antonyms or unrelated terms fall farther apart [27]. Extensions to phrases and sentences can be achieved via recurrent or recursive neural networks, which aggregate word-level vectors into larger structures [28]. In more advanced formulations, transformers use self-attention mechanisms to capture contextual relationships across entire sequences, enabling them to model word meaning in the context of surrounding tokens. Although these models demonstrate remarkable performance across tasks like question answering and text classification, they traditionally do not explicitly encode logical relations or the structure inherent in knowledge bases [29]. Instead, their strength lies in statistical generalization and flexibility, which can sometimes lead to errors when faced with tasks that require rigorous logical consistency.

Efforts to integrate logic with embedding-based approaches have led to hybrid systems where symbolic reasoning components process structured representations alongside distributed vectors [30]. One illustrative schema might link entity embeddings $\mathbf{e}_i \in \mathbb{R}^k$ with relation embeddings $\mathbf{r}_j \in \mathbb{R}^k$, and interpret logic statements as transformations or constraints in this vector space. For instance, the proposition $\forall x \, \exists y : R(x, y)$ can impose that for every entity embedding $\mathbf{e}_x$, there must be some $\mathbf{e}_y$ satisfying a certain geometric relation that reflects R. Algorithms for knowledge graph completion are an example, where the existence of edges is predicted based on learned embeddings and the geometry of the underlying space. Nevertheless, purely geometric interpretations may have difficulty with the nuanced rules often mandated by domains such as biomedical research, legal reasoning, or scientific literature, which require precise definitional constraints that are best expressed in symbolic logic. [31]

Moving beyond lexical semantics, discourse-level theories play a vital role in modeling how semantic relationships evolve across sentences and paragraphs. Phenomena like coreference, ellipsis, and bridging anaphora demand an understanding of how references to entities propagate throughout a text and how context shapes the interpretation of subsequent statements [32]. A system that identifies a pronoun "it" as referring to a particular previously mentioned object must keep track of a set of potential referents, weigh clues regarding gender or number if applicable, and integrate domain knowledge that might suggest certain plausible matches over others. Another crucial element is the recognition of discourse relations—causal, temporal, contrastive, or additive—that link sentences into a coherent narrative [33]. Such recognition can be modeled through either symbolic discourse representation structures or neural approaches that rely on attention mechanisms conditioned on inter-sentential context.

In educational settings, the capacity to ascertain the intent of an author, identify rhetorical strategies, and interpret implied meanings can be of paramount importance [34]. Models of "reading to learn" could integrate rhetorical structure theory or argumentation frameworks to evaluate how evidence is presented, how it supports or contradicts claims, and what conclusions might be derived. When the textual environment is particularly complex—such as scientific journal articles—an integrated system might combine lexical semantics, domain-specific ontologies, and a model of rhetorical structure to determine the significance of findings, data interpretations, or methodological constraints. [35]

A key dimension of the theoretical foundation is the intersection with formal knowledge representation methods, including first-order logic, description logics, and graph-based semantics. Representation languages such as $\mathcal{ALC}$ in description logics permit the expression of concepts, roles, and individuals, with inference tasks automated by reasoners that check consistency or deduce implied facts. The synergy with distributional approaches emerges when embeddings are used to approximate the extension of these concepts in large text corpora, bridging the gap between symbolic definitions and real-world usage patterns [36], [37]. One might encode domain knowledge by specifying constraints like $\forall x(CancerPatient(x) \rightarrow Person(x))$, while using learned embeddings to identify the mention of "cancer patient" in unstructured text [38]. This dual approach ensures that the system can logically verify domain constraints while still flexibly capturing lexical variability encountered in actual text passages.

Another theoretical axis is the study of cognitive models of reading comprehension, which often postulate that humans integrate multiple streams of information—linguistic, contextual, and inferential—when processing text [39]. Psycholinguistic experiments show that readers form mental models of the situations described, an approach sometimes formalized as situation semantics. These mental models are dynamic, reflecting new information introduced in subsequent sentences, and must reconcile inconsistencies or ambiguities [40]. Translating these ideas into computational practice encourages the development of incremental processing modules that update semantic representations in real time as each new portion of text is ingested. For instance, let $\mathbf{s}_t$ be the semantic state of the system at time t, then reading a sentence u_t updates $\mathbf{s}_t$ to $\mathbf{s}_{t+1} = f(\mathbf{s}_t, u_t)$, with f capturing the assimilation of new information into the existing representation.

Collectively, these theories suggest that a single representation or algorithm is unlikely to suffice for all the complex phenomena encompassed by reading comprehension [41]. Rather, a layered model appears best suited for capturing the multifaceted nature of language. This layered approach might incorporate a syntactic parsing stage to build initial structures, a semantic transformation stage to map those structures into logical forms or embedding-based representations, and an inference or retrieval stage to draw conclusions or respond to queries [42]. While many challenges remain—particularly regarding computational complexity, domain adaptation, and handling uncertain or contradictory information—the foundation laid by these theoretical considerations offers a roadmap for the design of advanced semantic analysis frameworks. In the following sections, these insights materialize in a framework that merges symbolic and sub-symbolic approaches, seeking to leverage their strengths and mitigate their individual weaknesses. [43]

3 Proposed Semantic Analysis Framework

The proposed framework integrates symbolic logic representations with distributional embedding approaches in a multi-layered system that seeks to replicate the complex cognitive processes underpinning human reading comprehension. The impetus is to capture not only the surface forms of language but also the deeper semantic relations and pragmatic cues that guide understanding [44]. The cornerstone of the framework is an interconnected pipeline composed of parsing, representation, inference, and evaluation components, each focusing on a particular dimension of the reading comprehension process. By chaining these components together, we aim to achieve a level of semantic robustness and contextual sensitivity that approximates human interpretative capabilities. [45]

The first stage of the pipeline concentrates on lexical analysis and preprocessing [46]. Here, textual input is divided into tokens, normalized, and annotated with part-of-speech tags. Named entity recognition is applied to identify references to specific entities, such as persons, locations, or specialized technical terms in scientific domains [47]. This annotated token stream then feeds into a syntactic parser trained to produce dependency or constituency trees. In domain-specialized scenarios, custom lexical resources augment this process by supplying domain lexicons or morphological variants that facilitate more accurate parsing [48]. The end product of this stage is a structured syntactic representation $\mathcal{T}$, capturing the hierarchical relationships among tokens.

Subsequent to parsing, the framework transforms syntactic structures into a form more amenable to semantic processing. One core representation is a typed logical form, reminiscent of the logical forms used in formal semantics, where predicate symbols represent relationships among entities [49]. For instance, consider the sentence "The algorithm sorts the list in ascending order." A simplified logical form might denote $\text{Algorithm}(a)$, $\text{List}(\ell)$, $\text{Sorts}(a, \ell)$, $\text{Ascending}(\ell)$. Meanwhile, named entities and references identified during lexical preprocessing are tagged with corresponding conceptual categories [50]. This transformation allows for more direct manipulation and inference, as it is now possible to apply unification or resolution-based algorithms over these logical structures.

Concurrently, a distributional embedding module encodes each token, phrase, or entity into a vector space representation [51]. For a word w in the text, an embedding $\mathbf{w} \in \mathbb{R}^d$ is retrieved from a pre-trained model, such as those based on large transformer architectures. Multi-head self-attention mechanisms can provide context-sensitive embeddings, $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n$, where each $\mathbf{h}_i \in \mathbb{R}^d$ captures not only local syntactic context but also broader discourse cues. These embeddings serve as a sub-symbolic complement to the symbolic logical form. For example, even if the sentence does not explicitly mention synonyms or hypernyms of "algorithm," the embedding captures relevant associations via learned distributional patterns. [52]

Once the logical and embedding representations are generated, the system fuses them in a conceptual integration layer. One strategy involves mapping predicates or concepts from the logical representation to embedding vectors, then applying transformations that link them to symbolically expressed constraints [53]. For a predicate $P(x, y)$, if x and y correspond to entities with embeddings $\mathbf{x}, \mathbf{y}$, a relation embedding $\mathbf{r}_P$ can be inferred to reflect the conceptual significance of P. Formally, if $\delta(\mathbf{x}, \mathbf{y}, \mathbf{r}_P)$ is a scoring function measuring how well $\mathbf{x}$ and $\mathbf{y}$ fit the relation P, then logical statements like $\forall x \exists y : P(x, y)$ can be tested by exploring the vector space for an appropriate $\mathbf{y}$. This conceptual integration layer thereby bridges the gap between symbolic expression and learned semantic nuances, allowing the framework to leverage the strengths of each representation in tandem. [54]

Inference is then conducted through a layered approach. For deductive reasoning, symbolic methods such as resolution or tableau algorithms can be employed on the logical forms to derive conclusions or identify contradictions

[55]. This might be used to check the consistency of a passage describing experimental procedures in a scientific article, ensuring that each step logically follows from stated premises. For broader, more flexible matching and retrieval, the embedding-based approach can facilitate approximate semantic searches, identifying text passages most relevant to a given query [56]. For example, if the system is asked whether the text describes a procedure akin to bubble sort, it could match distributions of terms related to sorting and comparisons, even if the term "bubble sort" is not explicitly used. Combining both symbolic and sub-symbolic approaches allows the system to both confirm logical correctness when explicit statements align with known rules and discover implicit connections buried in text. [57]

Evaluation metrics within this framework encompass a range of tasks indicative of reading comprehension. These may include question answering, where the system must supply correct responses based on textual evidence; textual entailment, which checks if a given hypothesis logically follows from a text; and contradiction detection, testing whether any statements conflict with one another [58]. More advanced tasks might require summarization or explanation generation, demanding that the system not only produce answers but articulate the chain of reasoning that led to them. In an educational context, such capacities enable the creation of interactive modules that guide learners, diagnose misunderstandings, or provide justifications for extracted conclusions. [59], [60]

One important aspect is the handling of discourse phenomena. In many real-world texts, crucial information unfolds over multiple sentences or paragraphs, requiring the system to track references and evolving contexts [61]. This can be achieved by incorporating methods for anaphora resolution and referencing chains, coupling them with both the logical representation and distributional embeddings. For example, when encountering a pronoun "they" in a subsequent sentence, the system consults the set of previously mentioned entities to determine the most probable referent, using both syntactic constraints and embedding similarities to disambiguate potential candidates [62]. In logic form, references to these entities can be updated by extending or modifying the predicate set, while embedding vectors for each candidate entity guide the resolution process through semantic similarity. [63]

Additionally, the framework acknowledges the importance of domain adaptation and the presence of specialized ontologies or terminologies in fields like biomedicine, law, or finance. Thus, an optional domain knowledge layer may incorporate specialized concept definitions, hierarchical relationships, or axioms that refine the semantic analysis [64]. For instance, a knowledge base might specify the domain rule $\forall x(Gene(x) \rightarrow BiologicalEntity(x))$, so when the text references a "gene," the system knows it belongs to a larger class of biological entities and can apply specialized inference rules that apply to that domain. If embeddings are also trained or fine-tuned on domain-specific corpora, the distributional aspect better reflects the specialized usage of terminology, facilitating improved disambiguation of terms that might have different meanings in a general corpus. [65]

Scalability is addressed through modular design. Each stage—parsing, logical transformation, embedding generation, integration, and inference—can be parallelized or distributed across computational nodes [66]. For very large document collections, indexing structures can be built on top of the embedded vectors, facilitating approximate nearest-neighbor queries for retrieval tasks. The symbolic component can be scaled by employing incremental reasoners or logic-based indexing strategies [67]. System resource usage can be managed by restricting the complexity of the generated logical forms to only the elements necessary for the target inference tasks, mitigating the overhead associated with fully detailed representations.

In summary, the proposed semantic analysis framework provides a cohesive pipeline that unites symbolic and sub-symbolic methodologies to navigate the complexity of human language [68]. It begins by parsing input into structured forms, proceeds by generating both logical and embedding-based representations, and culminates in a hybrid inference module that merges deductive logic with statistically driven retrieval. By providing mechanisms for discourse handling, domain adaptation, and robust evaluation, the framework aspires to facilitate intelligent reading systems capable of deep, context-sensitive comprehension across a range of textual domains. [69]

4 Implementation and Experimental Analysis

In order to validate the efficacy of the proposed semantic analysis framework, a prototype system was implemented and subjected to a series of evaluation tasks designed to test different facets of reading comprehension. This section provides an overview of the practical realization of the framework's components, details concerning the experimental setup and datasets, and a discussion of the performance outcomes and key insights gleaned from these experiments. [70]

The prototype system is built around a modular architecture that translates textual documents into a pipeline of representations, culminating in a hybrid inference layer. At the outset, a robust tokenizer and part-of-speech tagger processes the input [71]. Several publicly available resources, such as pretrained lexical embeddings, assist in the annotation of named entities [72]. A neural network parser then constructs syntactic dependency graphs, ensuring that the system can systematically retrieve subject-verb-object relationships and other syntactic configurations from a given sentence. The logical form generator simultaneously traverses these dependency graphs to build an internal representation, including predicate symbols for verbs and relational links between recognized entities. [73]

These symbolic representations are stored in a graph-based repository, utilizing a structure in which nodes correspond to entities or predicates and edges denote relations inferred by the parser. For example, if the system parses "The new method improves accuracy," it might store Method(m), Improves(m, a), Accuracy(a), along with typed edges connecting the nodes [74]. This symbolic graph also includes cross-references to distributional embeddings. Each node in the symbolic graph links to an embedding vector, enabling the system to leverage context-sensitive semantics [75]. For instance, the word "method" might have an embedding that clusters with related procedural terms, while "accuracy" might cluster near measures of performance or evaluation metrics.

In parallel, a transformer-based model is used to generate contextual embeddings for each token in the sentence [76]. These embeddings are pooled for each recognized entity, effectively integrating local syntactic cues from the immediate sentence with broader discourse features from surrounding text. The system thus maintains a dual representation—a structured symbolic layer and a learned embedding layer—for each piece of information parsed from the input. [77]

Once these representations are generated, the inference engine operates in one of two modes. In symbolic inference mode, the system queries the logical graph to test for the presence or absence of specific predicates, relationships, or constraints [78]. Queries like "Does the text state that the new method reduces error?" translate into a search for a predicate Reduces(m, e) or an equivalent paraphrase. If found, the system then checks whether the relevant arguments correspond to recognized references (e.g., "method," "error") [79]. If the system lacks direct evidence, it can inspect transitive relationships, or attempt unification-based resolution when multiple synonyms or morphological variants appear [80], [81]. This symbolic reasoning portion is implemented using a forward-chaining engine adapted from a knowledge-base reasoner capable of incremental updates.

In embedding-based inference mode, the system retrieves relevant text passages through similarity matching [82]. A query or target statement is converted into an embedding representation, which is then compared to stored embeddings via cosine similarity or a more sophisticated metric. If textual segments exist that are semantically similar to the query, the system retrieves them and checks whether any direct symbolic links can be established [83]. This dual approach ensures that the system can handle direct logical entailments while also uncovering implicit connections that are only apparent from distributional regularities.

Evaluation encompassed multiple datasets reflecting diverse domains and task types [84]. For question-answering tasks, standard reading comprehension benchmarks were used, including a corpus of Wikipedia articles paired with queries requiring short-form answers. Another dataset consisted of scientific abstracts where factual questions were posed regarding experimental findings or methodological details [85]. In a third evaluation, domain-specific corpora in biomedical research were tested, focusing on detecting statements about gene-disease associations, treatment effects, and other structured knowledge that can be naturally expressed in logical forms.

Quantitative metrics included accuracy, precision, recall, and F1 scores for tasks such as fact retrieval, contradiction detection, and textual entailment [86]. The system demonstrated high accuracy in retrieving correct answers to factual questions when the source text unambiguously contained the relevant information in a relatively direct manner. For instance, in paragraphs discussing the application of a novel method, symbolic queries about improvements in performance metrics were accurately answered by identifying and linking the relevant predicates (Improves(m, a)) in the logical graph [87]. Embedding-based retrieval also facilitated near-synonym matching, allowing the system to respond accurately even if the text used synonyms like "enhances" or "boosts" instead of "improves." The integration of domain knowledge in specialized tasks, such as biomedicine, allowed for improved handling of technical expressions and abbreviations, which might otherwise have posed challenges for purely distributional methods.

Concerning contradiction detection, the system leveraged logical forms to identify direct negations or inconsistent statements [88]. An example occurs when the text claims both "Method m reduces error e" and "Method m has no effect on error e." Symbolic reasoning recognized these statements as contradictory because it assigned conflicting truth values to the same predicate [89]. Embedding-based analysis complemented this process by revealing paraphrases in conflicting assertions that used different words but carried the same essential meaning. This fusion of symbolic and semantic approaches led to a robust contradiction detection mechanism, although the performance decreased in cases involving subtle rhetorical devices or implicit negations that were not syntactically explicit. [90]

A more nuanced evaluation tested the system's ability to track referential expressions across multiple sentences, particularly those involving pronouns, elliptical constructs, or nominal anaphora. In texts where a concept introduced as "the proposed algorithm" was later referenced as "it" or "this approach," the system's discourse handling component effectively reconciled these references by unifying them under the same entity node in the symbolic graph [91]. Embedding similarity also aided this process, reinforcing the link between "algorithm," "approach," and "it" when the local context strongly suggested the same referent. This synergy was critical for long-form reading comprehension tasks in which essential information was scattered across paragraphs, requiring a continuous updating of the knowledge graph as each new piece of text was processed. [92]

However, experiments uncovered some limitations. The system occasionally struggled with figurative language or highly ambiguous statements that lacked clear logical anchors [93]. Metaphors or idiomatic usages, such as "breaking new ground," did not always map well to either symbolic representations or standard embedding spaces, leading to errors or missed inferences. Additionally, while domain adaptation improved performance in specialized areas, the cost of constructing domain-specific resources (ontologies, specialized embeddings) could be prohibitive, and the system's coverage was limited by the scope of these resources [94]. Scaling to extremely large text collections also demanded efficient indexing strategies and computational resources, as the integration layer needed to cross-reference a vast volume of symbolic and embedded data.

Despite these challenges, the experimental results affirm that a hybrid semantic framework grounded in both symbolic and embedding-based methodologies can deliver strong performance in tasks demanding sophisticated text comprehension [95]. From a design standpoint, the ability to pivot between formal logical inference and fuzzy semantic similarity confers flexibility in handling the varied nature of real-world language data. The system's interpretability benefits from the availability of explicit logical forms, allowing developers and end-users to trace the rationale behind certain inferences or detect misinterpretations [96]. At the same time, the distributional embeddings provide a measure of resilience to lexical variations or gaps in the symbolic knowledge base. [97]

Beyond performance metrics, the qualitative inspection of inference traces and error cases offered valuable insights for future enhancements. In certain instances, deeper integration of pragmatic and world knowledge is needed to resolve ambiguities that remain even after parsing and embedding alignment [98]. For example, "The new technique substantially lowered the rate of false positives," without specifying which algorithm was replaced, can lead to incomplete reasoning about the comparative advantage of the technique. Implementing more refined discourse models that track rhetorical relations could help the system connect method evaluations across multiple paragraphs. [99]

In conclusion, the experimental analysis shows that the prototype successfully addresses many of the foundational problems in reading comprehension by unifying symbolic and distributional approaches. While open challenges persist—particularly in handling non-literal language, scaling to massive corpora, and constructing domain-specific resources—this hybrid strategy presents a promising route toward building robust, generalizable, and interpretable intelligent learning by reading systems. [100]

5 Conclusion

The research presented here has laid out an extensive, integrated framework designed to advance the capabilities of intelligent learning by reading systems through the synthesis of symbolic and embedding-based semantic analysis. Drawing on insights from formal logic, distributional semantics, syntax, and domain-specific knowledge bases, the proposed framework addresses the multifaceted challenges inherent in human language comprehension [101]. By merging structured logical representations with high-dimensional vector embeddings, the system achieves a level of inference that can detect explicit statements and uncover implicit relationships, thereby offering enhanced robustness in the face of linguistic variability.

Experimental evaluation across diverse datasets revealed that this hybrid approach yields accurate and interpretable results for tasks such as factual question answering, contradiction detection, and discourse-level comprehension [102]. The symbolic layer's clarity in explicit reasoning is bolstered by the distributional layer's capacity for approximate matching and contextual nuance. Limitations remain in the form of handling figurative language, domain coverage, and scalability, but these challenges suggest further opportunities for refinement and extension [103], [104]. They also underline the crucial role of additional knowledge resources and more advanced discourse modeling techniques in achieving deeper semantic understanding.

In sum, the work undertaken here serves as a benchmark for researchers and developers seeking to build sophisticated reading comprehension systems capable of both rigorous logical inference and flexible interpretation of natural language [105]. Future directions could involve extending the discourse analysis component to fully capture rhetorical strategies, refining methods for domain transfer, and implementing pragmatic reasoning modules that take into account speaker intent and conversational context. Ultimately, these efforts promise to advance the field toward highly adaptive, context-aware reading systems that can engage with textual content in a manner increasingly reminiscent of human understanding. [106]

References

[1] R. Muthuri, S. Capecchi, E. Sulis, I. A. Amantea, and G. Boella, "Integrating value modeling and legal risk management: An it case study," *Information Systems and e-Business Management*, vol. 20, no. 1, pp. 27–55, Nov. 9, 2021. DOI: 10.1007/s10257-021-00543-2.

[2] A. Elsheikh, S. Yacout, M.-S. Ouali, and Y. Shaban, "Failure time prediction using adaptive logical analysis of survival curves and multiple machining signals," *Journal of Intelligent Manufacturing*, vol. 31, no. 2, pp. 403–415, Oct. 30, 2018. DOI: 10.1007/s10845-018-1453-4.

[3] Y. Ming, D. Pelusi, C.-N. Fang, *et al.*, "Eeg data analysis with stacked differentiable neural computers," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7611–7621, Nov. 20, 2018. DOI: 10.1007/s00521-018-3879-1.

[4] C. B. Şahin, Ö. B. Dinler, and L. Abualigah, "Prediction of software vulnerability based deep symbiotic genetic algorithms: Phenotyping of dominant-features," *Applied Intelligence*, vol. 51, no. 11, pp. 8271–8287, Mar. 31, 2021. DOI: 10.1007/s10489-021-02324-3.

[5] N. A. Bizami, Z. Tasir, and S. N. Kew, "Innovative pedagogical principles and technological tools capabilities for immersive blended learning: A systematic literature review.," *Education and information technologies*, vol. 28, no. 2, pp. 1373–1425, Jul. 29, 2022. DOI: 10.1007/s10639-022-11243-w.

[6] L. Pascarella, M. Bruntink, and A. Bacchelli, "Classifying code comments in java software systems," *Empirical Software Engineering*, vol. 24, no. 3, pp. 1499–1537, Apr. 26, 2019. DOI: 10.1007/s10664-019-09694-w.

[7] A. Türkyılmaz, O. Senvar, İ. Ünal, and S. Bulkan, "A research survey: Heuristic approaches for solving multi objective flexible job shop problems," *Journal of Intelligent Manufacturing*, vol. 31, no. 8, pp. 1949–1983, Feb. 19, 2020. DOI: 10.1007/s10845-020-01547-4.

[8] J. A. Marmolejo-Saucedo, "Design and development of digital twins: A case study in supply chains," *Mobile Networks and Applications*, vol. 25, no. 6, pp. 2141–2160, Jun. 6, 2020. DOI: 10.1007/s11036-020-01557-9.

[9] A. Sharma and K. D. Forbus, "Graph-based reasoning and reinforcement learning for improving q/a performance in large knowledge-based systems," in *2010 AAAI Fall Symposium Series*, 2010.

[10] M. Ribera, R. Pozzobon, and S. Sayago, "Publishing accessible proceedings: The dsai 2016 case study," *Universal Access in the Information Society*, vol. 19, no. 3, pp. 557–569, Jun. 18, 2019. DOI: 10.1007/s10209-019-00660-3.

[11] A. Kader, K. Z. Zamli, and B. S. Ahmed, "A systematic review on emperor penguin optimizer," *Neural Computing and Applications*, vol. 33, no. 23, pp. 15 933–15 953, Sep. 1, 2021. DOI: 10.1007/s00521-021-06442-4.

[12] M. Nowicki, "Comparison of sort algorithms in hadoop and pcj," *Journal of Big Data*, vol. 7, no. 1, pp. 1–28, Nov. 16, 2020. DOI: 10.1186/s40537-020-00376-9.

[13] *Improved Architecture of Artificial Neural Network for Secondary Structure Analysis*, vol. 20, Suppl 17, United Kingdom: Springer Science and Business Media LLC, Nov. 8, 2019. DOI: 10.1186/s12859-019-3122-9.

[14] K. Sandkuhl and U. Seigerroth, "Method engineering in information systems analysis and design: A balanced scorecard approach for method improvement," *Software & Systems Modeling*, vol. 18, no. 3, pp. 1833–1857, Aug. 27, 2018. DOI: 10.1007/s10270-018-0692-3.

[15] L. Pan, Y. Xiong, B. Li, T. Huang, and W. Wan, "Feature attenuation reinforced recurrent neural network for diffusion prediction," *Applied Intelligence*, vol. 53, no. 2, pp. 1855–1869, May 3, 2022. DOI: 10.1007/s10489-022-03413-7.

[16] R. Mercado, V. Munoz-Jiménez, M. Ramos, and F. Ramos, "Generation of virtual creatures under multidisciplinary biological premises," *Artificial Life and Robotics*, vol. 27, no. 3, pp. 495–505, Jun. 20, 2022. DOI: 10.1007/s10015-022-00767-6.

[17] H. Song, R. Dautov, N. Ferry, A. Solberg, and F. Fleurey, "Model-based fleet deployment in the iot–edge–cloud continuum," *Software and Systems Modeling*, vol. 21, no. 5, pp. 1931–1956, May 3, 2022. DOI: 10.1007/s10270-022-01006-z.

[18] B. A. Muse, C. Nagy, A. Cleve, F. Khomh, and G. Antoniol, "Fixme: Synchronize with database! an empirical study of data access self-admitted technical debt," *Empirical Software Engineering*, vol. 27, no. 6, Jul. 8, 2022. DOI: 10.1007/s10664-022-10119-4.

[19] S. Nikpour and S. Asadi, "A dynamic hierarchical incremental learning-based supervised clustering for data stream with considering concept drift," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 6, pp. 2983–3003, Jan. 4, 2022. DOI: 10.1007/s12652-021-03673-0.

[20] M. Kansara, "Advancements in cloud database migration: Current innovations and future prospects for scalable and secure transitions," *Sage Science Review of Applied Machine Learning*, vol. 7, no. 1, pp. 127–143, 2024. DOI: 10.17613/zvsjp-kzr29.

[21] K. S. Kiangala and Z. Wang, "Initiating predictive maintenance for a conveyor motor in a bottling plant using industry 4.0 concepts," *The International Journal of Advanced Manufacturing Technology*, vol. 97, no. 9, pp. 3251–3271, May 24, 2018. DOI: 10.1007/s00170-018-2093-8.

[22] L. F. C. S. Durão, M. M. de Carvalho, S. M. Takey, P. A. Cauchick-Miguel, and E. de Senzi Zancul, “Internet of things process selection: Ahp selection method,” *The International Journal of Advanced Manufacturing Technology*, vol. 99, no. 9, pp. 2623–2634, Sep. 10, 2018. DOI: 10.1007/s00170-018-2617-2.

[23] A. Sharma, “Structural and network-based methods for knowledge-based systems,” Ph.D. dissertation, Northwestern University, 2011.

[24] Y. Wang, C. Zhang, and K. Li, “A review on method entities in the academic literature: Extraction, evaluation, and application,” *Scientometrics*, vol. 127, no. 5, pp. 2479–2520, Mar. 12, 2022. DOI: 10.1007/s11192-022-04332-7.

[25] H. Ergin, J. Gray, B. Rumpe, and M. Schindler, “Sosym reflections: The 2021 ”state of the journal” report.,” *Software and systems modeling*, vol. 21, no. 1, pp. 1–7, Feb. 7, 2022. DOI: 10.1007/s10270-022-00979-1.

[26] P. Rani, S. Panichella, M. Leuenberger, M. Ghafari, and O. Nierstrasz, “What do class comments tell us? an investigation of comment evolution and practices in pharo smalltalk,” *Empirical Software Engineering*, vol. 26, no. 6, pp. 112–, Aug. 18, 2021. DOI: 10.1007/s10664-021-09981-5.

[27] A. M. Fernández-Sáez, M. R. V. Chaudron, and M. Genero, “An industrial case study on the use of uml in software maintenance and its perceived benefits and hurdles,” *Empirical Software Engineering*, vol. 23, no. 6, pp. 3281–3345, Mar. 16, 2018. DOI: 10.1007/s10664-018-9599-4.

[28] R. Rajesh, “Network design for resilience in supply chains using novel crazy elitist tlbo,” *Neural Computing and Applications*, vol. 32, no. 11, pp. 7421–7437, May 23, 2019. DOI: 10.1007/s00521-019-04260-3.

[29] M. H. Alsulami, “Challenges facing the implementation of 5g,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 5, pp. 6213–6226, Apr. 7, 2022. DOI: 10.1007/s12652-021-03397-1.

[30] Y. Zhang, Y. Li, and H. Wang, “Mobile neural intelligent information system based on edge computing with interactive data,” *Neural Computing and Applications*, vol. 33, no. 9, pp. 4329–4341, Aug. 5, 2020. DOI: 10.1007/s00521-020-05269-9.

[31] F. de A. T. de Carvalho, A. Irpino, R. Verde, and A. Balzanella, “Batch self-organizing maps for distributional data with an automatic weighting of variables and components,” *Journal of Classification*, vol. 39, no. 2, pp. 343–375, Mar. 18, 2022. DOI: 10.1007/s00357-022-09411-1.

[32] L. Li, “Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond.,” *Information systems frontiers : a journal of research and innovation*, vol. 26, no. 5, pp. 1–1712, Jul. 13, 2022. DOI: 10.1007/s10796-022-10308-y.

[33] F. Lopes, C. Teixeira, and H. G. Oliveira, “Comparing different methods for named entity recognition in portuguese neurology text,” *Journal of medical systems*, vol. 44, no. 4, pp. 77–77, Feb. 28, 2020. DOI: 10.1007/s10916-020-1542-8.

[34] X. Lei, P. Wu, J. Zhu, and J. Wang, “Ontology-based information integration: A state-of-the-art review in road asset management,” *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2601–2619, Oct. 21, 2021. DOI: 10.1007/s11831-021-09668-6.

[35] Q. Hu, B. Liu, J. Gao, K. L. Nielbo, and M. R. Thomsen, “Fractal scaling laws for the dynamic evolution of sentiments in never let me go and their implications for writing, adaptation and reading of novels,” *World Wide Web*, vol. 24, no. 4, pp. 1147–1164, May 25, 2021. DOI: 10.1007/s11280-021-00892-5.

[36] S. E. Bibri, “On the sustainability of smart and smarter cities in the era of big data: An interdisciplinary and transdisciplinary literature review,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–64, Mar. 15, 2019. DOI: 10.1186/s40537-019-0182-7.

[37] A. Sharma and K. Forbus, “Automatic extraction of efficient axiom sets from large knowledge bases,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 27, 2013, pp. 1248–1254.

[38] A. Cartas, J. Marín, P. Radeva, and M. Dimiccoli, “Batch-based activity recognition from egocentric photostreams revisited,” *Pattern Analysis and Applications*, vol. 21, no. 4, pp. 953–965, May 5, 2018. DOI: 10.1007/s10044-018-0708-1.

[39] M. K. Martin, C. Lebiere, M. Fields, and C. Lennon, “Learning features while learning to classify: A cognitive model for autonomous systems,” *Computational and Mathematical Organization Theory*, vol. 26, no. 1, pp. 23–54, Jul. 17, 2018. DOI: 10.1007/s10588-018-9279-3.

[40] P. Monnin, J. Legrand, G. Husson, *et al.*, “Pgxo and pgxlod: A reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison.,” *BMC bioinformatics*, vol. 20, no. 4, pp. 1–16, Apr. 18, 2019. DOI: 10.1186/s12859-019-2693-9.

[41] J. Montalvo, Á. García-Martín, and J. Bescós, "Exploiting semantic segmentation to boost reinforcement learning in video game environments," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 10 961–10 979, Sep. 15, 2022. DOI: 10.1007/s11042-022-13695-1.

[42] D. Sas and P. Avgeriou, "Quality attribute trade-offs in the embedded systems industry: An exploratory case study," *Software Quality Journal*, vol. 28, no. 2, pp. 505–534, Dec. 4, 2019. DOI: 10.1007/s11219-019-09478-x.

[43] A. R. Florea and M. Roman, "Artificial neural networks applied for predicting and explaining the education level of twitter users.," *Social network analysis and mining*, vol. 11, no. 1, pp. 112–112, Nov. 1, 2021. DOI: 10.1007/s13278-021-00832-1.

[44] F. Jáñez-Martino, R. Alaiz-Rodríguez, V. González-Castro, E. Fidalgo, and E. Alegre, "A review of spam email detection: Analysis of spammer strategies and the dataset shift problem," *Artificial Intelligence Review*, vol. 56, no. 2, pp. 1145–1173, May 11, 2022. DOI: 10.1007/s10462-022-10195-4.

[45] A. Brogi, S. Forti, C. Guerrero, and I. Lera, "Declarative application management in the fog," *Journal of Grid Computing*, vol. 19, no. 4, pp. 45–, Oct. 27, 2021. DOI: 10.1007/s10723-021-09582-y.

[46] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, and M. Bilal, "Smart home security: Challenges, issues and solutions at different iot layers," *The Journal of Supercomputing*, vol. 77, no. 12, pp. 14 053–14 089, May 10, 2021. DOI: 10.1007/s11227-021-03825-1.

[47] N. Maleki, A. M. Rahmani, and M. Conti, "Mapreduce: An infrastructure review and research insights," *The Journal of Supercomputing*, vol. 75, no. 10, pp. 6934–7002, Jun. 8, 2019. DOI: 10.1007/s11227-019-02907-5.

[48] V. Engerer, "Temporality revisited: Dynamicity issues in collaborative digital writing research," *Education and Information Technologies*, vol. 26, no. 1, pp. 339–370, Jul. 6, 2020. DOI: 10.1007/s10639-020-10262-9.

[49] I. Ruiz-Rube, T. Person, J. M. Dodero, J. M. Mota, and J. M. Sánchez-Jara, "Applying static code analysis for domain-specific languages," *Software and Systems Modeling*, vol. 19, no. 1, pp. 95–110, Apr. 6, 2019. DOI: 10.1007/s10270-019-00729-w.

[50] E. Fourati, W. Elloumi, and A. Chetouani, "Anti-spoofing in face recognition-based biometric authentication using image quality assessment," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 865–889, Oct. 10, 2019. DOI: 10.1007/s11042-019-08115-w.

[51] J. Torregrosa, G. Bello-Orgaz, E. Martínez-Cámara, J. D. Ser, and D. Camacho, "A survey on extremism analysis using natural language processing: Definitions, literature review, trends and challenges.," *Journal of ambient intelligence and humanized computing*, vol. 14, no. 8, pp. 1–9905, Jan. 12, 2022. DOI: 10.1007/s12652-021-03658-z.

[52] S. Kanza, N. Gibbins, and J. G. Frey, "Too many tags spoil the metadata: Investigating the knowledge management of scientific research with semantic web technologies.," *Journal of cheminformatics*, vol. 11, no. 1, pp. 1–23, Mar. 21, 2019. DOI: 10.1186/s13321-019-0345-8.

[53] A. Bakhshi, S. K. Chalup, A. Harimi, and S. M. Mirhassani, "Recognition of emotion from speech using evolutionary cepstral coefficients," *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35 739–35 759, Sep. 18, 2020. DOI: 10.1007/s11042-020-09591-1.

[54] T. Si, P. B. C. de Miranda, J. V. Galdino, and A. L. B. do Nascimento, "Grammar-based automatic programming for medical data classification: An experimental study," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4097–4135, Feb. 17, 2021. DOI: 10.1007/s10462-020-09949-9.

[55] J. Yuan, "An anomaly data mining method for mass sensor networks using improved pso algorithm based on spark parallel framework," *Journal of Grid Computing*, vol. 18, no. 2, pp. 251–261, Jan. 17, 2020. DOI: 10.1007/s10723-020-09505-3.

[56] M. Singh, "Scalability and sparsity issues in recommender datasets: A survey," *Knowledge and Information Systems*, vol. 62, no. 1, pp. 1–43, Oct. 16, 2018. DOI: 10.1007/s10115-018-1254-2.

[57] J. M. Palomares-Pecho, G. F. M. Silva-Calpa, and A. Raposo, "End-user adaptable technologies for rehabilitation: A systematic literature review," *Universal Access in the Information Society*, vol. 20, no. 2, pp. 299–319, Jun. 6, 2020. DOI: 10.1007/s10209-020-00720-z.

[58] F. T. Giuntini, M. T. Cazzolato, M. de Jesus Dutra dos Reis, A. T. Campbell, A. J. M. Traina, and J. Ueyama, "A review on recognizing depression in social networks: Challenges and opportunities," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 4713–4729, Jan. 24, 2020. DOI: 10.1007/s12652-020-01726-4.

[59] L. Pedro, C. Barbosa, and C. Santos, "A critical review of mobile learning integration in formal educational contexts," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 1, pp. 1–15, Mar. 15, 2018. DOI: 10.1186/s41239-018-0091-4.

[60] A. Sharma and K. Forbus, "Graph traversal methods for reasoning in large knowledge-based systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 27, 2013, pp. 1255–1261.

[61] M. Abreu, T. H. Silva, H. Teixeira, L. P. Reis, and N. Lau, "6d localization and kicking for humanoid robotic soccer," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 2, pp. 1–25, May 12, 2021. DOI: 10.1007/s10846-021-01385-3.

[62] Z. Li, "Visual analysis framework for network abnormal data based on multi-agent model," *Soft Computing*, vol. 25, no. 3, pp. 1833–1845, Aug. 19, 2020. DOI: 10.1007/s00500-020-05257-0.

[63] Z. Ali, G. Qi, P. Kefalas, W. A. Abro, and B. Ali, "A graph-based taxonomy of citation recommendation models," *Artificial Intelligence Review*, vol. 53, no. 7, pp. 5217–5260, Feb. 21, 2020. DOI: 10.1007/s10462-020-09819-4.

[64] K. Khanafer, B. Alshuraiaan, A. Al-Masri, S. Aithal, and I. Deiab, "Finite element analysis of thermo-mechanical behavior of a multi-layer laser additive manufacturing process," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 16, no. 3, pp. 893–911, Jun. 1, 2022. DOI: 10.1007/s12008-022-00916-y.

[65] J. J. L. Escobar, R. P. D. Redondo, and F. Gil-Castiñeira, "In-depth analysis and open challenges of mist computing," *Journal of Cloud Computing*, vol. 11, no. 1, Nov. 19, 2022. DOI: 10.1186/s13677-022-00354-x.

[66] N. A. M. Mokmin and R. P. Rassy, "Review of the trends in the use of augmented reality technology for students with disabilities when learning physical education," *Education and Information Technologies*, vol. 29, no. 2, pp. 1251–1277, Dec. 26, 2022. DOI: 10.1007/s10639-022-11550-2.

[67] M. P. Ramos, P. M. Tasinaffo, A. M. Cunha, D. A. Silva, G. S. Gonçalves, and L. A. V. Dias, "A canonical model for seasonal climate prediction using big data," *Journal of Big Data*, vol. 9, no. 1, Mar. 3, 2022. DOI: 10.1186/s40537-022-00580-9.

[68] J. Afonso, M. M. Saraiva, J. P. S. Ferreira, *et al.*, "Automated detection of ulcers and erosions in capsule endoscopy images using a convolutional neural network.," *Medical & biological engineering & computing*, vol. 60, no. 3, pp. 719–725, Jan. 17, 2022. DOI: 10.1007/s11517-021-02486-9.

[69] B. C. Stöver, S. Wiechers, and K. F. Müller, "Jphyloio : A java library for event-based reading and writing of different phylogenetic file formats through a common interface," *BMC bioinformatics*, vol. 20, no. 1, pp. 402–402, Jul. 22, 2019. DOI: 10.1186/s12859-019-2982-3.

[70] W. Zang, F. Miao, R. Gravina, F. Sun, G. Fortino, and Y. Li, "Cmdp-based intelligent transmission for wireless body area network in remote health monitoring," *Neural Computing and Applications*, vol. 32, no. 3, pp. 829–837, Jan. 29, 2019. DOI: 10.1007/s00521-019-04034-x.

[71] V. T. Rathod, P. K. Jha, and N. M. Sawai, "Optical cad modelling and designing of compound die using the python scripting language," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 17, no. 2, pp. 981–991, Jun. 15, 2022. DOI: 10.1007/s12008-022-00922-0.

[72] D. Cortés, J. Ramírez, J. Medina, and A. Molina, "Integrated product, process and manufacturing system development for multifunctional micromachine tool," *The International Journal of Advanced Manufacturing Technology*, vol. 120, no. 11-12, pp. 7477–7521, Apr. 29, 2022. DOI: 10.1007/s00170-022-09060-z.

[73] H. Mohapatra and A. K. Rath, "A fault tolerant routing scheme for advanced metering infrastructure: An approach towards smart grid," *Cluster Computing*, vol. 24, no. 3, pp. 2193–2211, Mar. 7, 2021. DOI: 10.1007/s10586-021-03255-x.

[74] P. Nimbe, B. A. Weyori, and A. F. Adekoya, "Models in quantum computing: A systematic review," *Quantum Information Processing*, vol. 20, no. 2, pp. 1–61, Feb. 24, 2021. DOI: 10.1007/s11128-021-03021-3.

[75] C. Corbane, V. Syrris, F. Sabo, *et al.*, "Convolutional neural networks for global human settlements mapping from sentinel-2 satellite imagery," *Neural Computing and Applications*, vol. 33, no. 12, pp. 6697–6720, Oct. 27, 2020. DOI: 10.1007/s00521-020-05449-7.

[76] G. Z. del Balzo, "Statistical field theory of the transmission of nerve impulses.," *Theoretical biology & medical modelling*, vol. 18, no. 1, pp. 1–10, Jan. 6, 2021. DOI: 10.1186/s12976-020-00132-9.

[77] T. H. Nguyen, L. Waizenegger, and A. A. Techatassanasoontorn, ""don't neglect the user!" – identifying types of human-chatbot interactions and their associated characteristics," *Information Systems Frontiers*, vol. 24, no. 3, pp. 797–838, Dec. 10, 2021. DOI: 10.1007/s10796-021-10212-x.

[78] S. A. Safdar, H. Lu, T. Yue, S. Ali, and K. Nie, “A framework for automated multi-stage and multi-step product configuration of cyber-physical systems,” *Software and Systems Modeling*, vol. 20, no. 1, pp. 211–265, Jun. 13, 2020. DOI: 10.1007/s10270-020-00803-8.

[79] M. Paredes-Velasco, J. Á. Velázquez-Iturbide, and M. Gómez-Ríos, “Augmented reality with algorithm animation and their effect on students' emotions.,” *Multimedia tools and applications*, vol. 82, no. 8, pp. 11 819–11 845, Sep. 7, 2022. DOI: 10.1007/s11042-022-13679-1.

[80] M. Sridharan and T. Mota, “Towards combining commonsense reasoning and knowledge acquisition to guide deep learning,” *Autonomous Agents and Multi-Agent Systems*, vol. 37, no. 1, Nov. 1, 2022. DOI: 10.1007/s10458-022-09584-4.

[81] A. Sharma and K. Goolsbey, “Identifying useful inference paths in large commonsense knowledge bases by retrograde analysis,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017.

[82] V. Döller, D. Karagiannis, and W. Utz, “Metamorph: Formalization of domain-specific conceptual modeling methods—an evaluative case study, juxtaposition and empirical assessment,” *Software and Systems Modeling*, vol. 22, no. 1, pp. 75–110, Oct. 6, 2022. DOI: 10.1007/s10270-022-01047-4.

[83] G. Uddin, F. Sabir, Y.-G. Guéhéneuc, O. Alam, and F. Khomh, “An empirical study of iot topics in iot developer discussions on stack overflow,” *Empirical Software Engineering*, vol. 26, no. 6, pp. 1–45, Sep. 6, 2021. DOI: 10.1007/s10664-021-10021-5.

[84] M. I. Naas, L. Lemarchand, P. Raipin, and J. Boukhobza, “Iot data replication and consistency management in fog computing,” *Journal of Grid Computing*, vol. 19, no. 3, pp. 33–, Jul. 21, 2021. DOI: 10.1007/s10723-021-09571-1.

[85] N. Gadipudi, I. Elamvazuthi, L. I. Izhar, *et al.*, “A review on monocular tracking and mapping: From model-based to data-driven methods,” *The Visual Computer*, vol. 39, no. 11, pp. 5897–5924, Nov. 17, 2022. DOI: 10.1007/s00371-022-02702-z.

[86] A. Guarino, L. Grilli, D. Santoro, F. Messina, and R. Zaccagnino, “To learn or not to learn? evaluating autonomous, adaptive, automated traders in cryptocurrencies financial bubbles,” *Neural Computing and Applications*, vol. 34, no. 23, pp. 20 715–20 756, Jul. 27, 2022. DOI: 10.1007/s00521-022-07543-4.

[87] S. Singh, J. Singh, and S. Singh, “Mitigating spoofed gnss trajectories through nature inspired algorithm,” *GeoInformatica*, vol. 25, no. 3, pp. 581–600, May 22, 2020. DOI: 10.1007/s10707-020-00412-z.

[88] J. P. dos Reis, F. B. e Abreu, G. de Figueiredo Carneiro, and C. Anslow, “Code smells detection and visualization: A systematic literature review,” *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 1–48, Mar. 10, 2021. DOI: 10.1007/s11831-021-09566-x.

[89] C. Guitton, A. Tamò-Larrieux, and S. Mayer, “Mapping the issues of automated legal systems: Why worry about automatically processable regulation?” *Artificial Intelligence and Law*, vol. 31, no. 3, pp. 571–599, Aug. 4, 2022. DOI: 10.1007/s10506-022-09323-w.

[90] S. Huang, Y. Guo, N. Yang, S. Zha, D. Liu, and W. Fang, “A weighted fuzzy c-means clustering method with density peak for anomaly detection in iot-enabled manufacturing process,” *Journal of Intelligent Manufacturing*, vol. 32, no. 7, pp. 1845–1861, Oct. 22, 2020. DOI: 10.1007/s10845-020-01690-y.

[91] A. Polyviou and E. D. Zamani, “Are we nearly there yet? a desires & realities framework for europe's ai strategy,” *Information Systems Frontiers*, vol. 25, no. 1, pp. 143–159, Jun. 8, 2022. DOI: 10.1007/s10796-022-10285-2.

[92] J. Samhi, K. Allix, T. F. Bissyandé, and J. Klein, “A first look at android applications in google play related to covid-19,” *Empirical software engineering*, vol. 26, no. 4, pp. 57–57, Apr. 21, 2021. DOI: 10.1007/s10664-021-09943-x.

[93] E. C. Groen, R. Harrison, P. K. Murukannaiah, and A. Vogelsang, “Guest editorial: Special section on artificial intelligence for requirements engineering,” *Automated Software Engineering*, vol. 26, no. 3, pp. 511–512, Jul. 31, 2019. DOI: 10.1007/s10515-019-00262-6.

[94] O. Levy and D. G. Feitelson, “Understanding large-scale software systems – structure and flows,” *Empirical Software Engineering*, vol. 26, no. 3, pp. 1–39, Mar. 31, 2021. DOI: 10.1007/s10664-021-09938-8.

[95] R. Beamonte, N. Ezzati-Jivan, and M. Dagenais, “Automated generation of model-based constraints for common multi-core and real-time applications using execution tracing,” *International Journal of Parallel Programming*, vol. 49, no. 1, pp. 104–134, Jan. 1, 2021. DOI: 10.1007/s10766-020-00689-5.

[96] S. S. Kolesnikov, N. Siegmund, C. Kästner, A. Grebhahn, and S. Apel, “Tradeoffs in modeling performance of highly configurable software systems,” *Software & Systems Modeling*, vol. 18, no. 3, pp. 2265–2283, Feb. 8, 2018. DOI: 10.1007/s10270-018-0662-9.

[97] R. Alt, "Electronic markets on digital transformation methodologies," *Electronic Markets*, vol. 29, no. 3, pp. 307–313, Sep. 16, 2019. DOI: 10.1007/s12525-019-00370-x.

[98] B. A. Sanchez, A. Zolotas, H. H. Rodriguez, *et al.*, "Runtime translation of ocl-like statements on simulink models: Expanding domains and optimising queries," *Software and systems modeling*, vol. 20, no. 6, pp. 1–30, Aug. 23, 2021. DOI: 10.1007/s10270-021-00910-0.

[99] F. Hou and S. Jansen, "A systematic literature review on trust in the software ecosystem," *Empirical Software Engineering*, vol. 28, no. 1, Nov. 23, 2022. DOI: 10.1007/s10664-022-10238-y.

[100] J. Xu and Y. Liu, "Cet-4 score analysis based on data mining technology," *Cluster Computing*, vol. 22, no. 2, pp. 3583–3593, Mar. 19, 2018. DOI: 10.1007/s10586-018-2208-x.

[101] E. S. Vieira, "International research collaboration in africa: A bibliometric and thematic analysis," *Scientometrics*, vol. 127, no. 5, pp. 2747–2772, Mar. 25, 2022. DOI: 10.1007/s11192-022-04349-y.

[102] E. Hogg, S. Hauert, D. Harvey, and A. Richards, "Evolving behaviour trees for supervisory control of robot swarms," *Artificial Life and Robotics*, vol. 25, no. 4, pp. 569–577, Oct. 18, 2020. DOI: 10.1007/s10015-020-00650-2.

[103] F. D. B. S. Praciano, P. R. P. Amora, I. C. Abreu, F. L. F. Pereira, and J. C. Machado, "Robust cardinality: A novel approach for cardinality prediction in sql queries," *Journal of the Brazilian Computer Society*, vol. 27, no. 1, pp. 11–, Sep. 1, 2021. DOI: 10.1186/s13173-021-00115-9.

[104] A. Sharma and K. M. Goolsbey, "Simulation-based approach to efficient commonsense reasoning in very large knowledge bases," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 1360–1367.

[105] Z. Falomir, R. Tarin, A. Puerta, and P. Garcia-Segarra, "An interactive game for training reasoning about paper folding," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6535–6566, Oct. 18, 2020. DOI: 10.1007/s11042-020-09830-5.

[106] M. Nadin, "Machine intelligence: A chimera," *AI & SOCIETY*, vol. 34, no. 2, pp. 215–242, Apr. 21, 2018. DOI: 10.1007/s00146-018-0842-8.