Deep Learning-Driven Ontology Learning: A Systematic Mapping Study

Asma Amalki
Image and Pattern Recognition – Intelligent and Communicating Systems Laboratory (IRF-SIC), Faculty of Science, Ibn Zohr University, Agadir, Morocco
asma.amalki@edu.uiz.ac.ma (corresponding author)

Khalid Tatane
ESTIDMA Research Team, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco
k.tatane@uiz.ac.ma

Ali Bouzit
Image and Pattern Recognition – Intelligent and Communicating Systems Laboratory (IRF-SIC), Faculty of Science, Ibn Zohr University, Agadir, Morocco
a.bouzit@uiz.ac.ma

Received: 28 October 2024 | Revised: 18 November 2024 and 26 November 2024 | Accepted: 27 November 2024

 DOI: https://doi.org/10.48084/etasr.9431

ABSTRACT

Today, ontologies are the widely accepted framework for managing knowledge in a manner that supports sharing, reuse, and automatic interpretation. Ontologies are fundamental to various Artificial Intelligence (AI) applications, including smart information retrieval, knowledge management, and contextual organization. However, the rapid growth of data in various domains has made ontology acquisition and enrichment, time-consuming, labor-intensive, and expensive. Consequently, there is a need for automated methods for this task, commonly referred to as ontology learning. Deep learning models have made significant advancements in this field, as they can extract concepts from vast corpora and infer semantic relationships from wide-ranging datasets. This paper aims to explore and synthesize existing research on the application of deep learning techniques to ontology learning. To achieve this, a Systematic Mapping Study (SMS) was conducted, encompassing 2765 papers published between 2015 and September 2024, from which 47 research papers were selected for review and analysis. The studies were systematically categorized according to eight refined criteria: publication year, type of contribution, empirical study design, type of data used, deep learning techniques implemented, domain of application, focused ontology learning tasks, and evaluation metrics and benchmarks.

Keywords-ontology; ontology learning; deep learning; systematic mapping study; knowledge representation

I. INTRODUCTION

The purely philosophical concept of ontology, which began in the 1990s, has evolved into one of the key paradigms within computer science and information systems. Nowadays, ontologies can also serve as a kind of domain knowledge representation, that is, a way to systemize and harmonize concepts across many disciplines [1]. This evolution has enabled syntactic interoperability to evolve into semantic interoperability, giving systems a way to consistently find meaning in data [2]. As a result, ontologies have become instrumental in enhancing data integration, information retrieval, and decision making in areas as diverse as healthcare [3], knowledge management [4], and semantic web technologies [5]. For example, recent works point out their role in integrating complex datasets in personalized medicine and other rapidly evolving domains [6]. Ontology development and maintenance activities remain resource-intensive and time-consuming. Traditional ontology development relies heavily on domain-driven experts, leading to scalability bottlenecks and human errors [7]. These challenges are further compounded when heterogeneous data sources need to be integrated, as inconsistencies can arise, leading to incompleteness and biased representations of knowledge [8]. The dynamic nature of most fields, such as healthcare or technology, also poses challenges, as many ontology-based systems lack the flexibility to quickly adapt to new information [9]. Such limitations signify the need for automated and intelligent techniques to facilitate the process of ontology development and maintenance [10]. In this context, ontology learning has emerged as a promising solution. Ontology learning enables the automated extraction of terms, relations, and axioms from textual data [9]. By reducing the reliance on manual input, ontology learning significantly enhances the efficiency and scalability of ontology development. Recent research in deep learning has further improved the accuracy and flexibility of these approaches [11]. These models are capable of handling complex and high-dimensional data, making them suitable for automating and refining the ontology learning process [12]. Algorithms such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformer Models, Graph Neural Networks (GNNs), and Large Language Models (LLMs), including BERT and GPT, excel at capturing deep semantic structures and contextual information, providing new avenues toward enhanced, accurate, and complete ontology development, thereby enhancing the overall quality and scalability of knowledge representation systems.

The scope of this Systematic Mapping Study (SMS) is to explore the intersection of ontology learning and deep learning techniques. It aims to categorize and analyze the existing body of research based on eight criteria: publication year, type of contribution, empirical study design, type of data used, deep learning techniques implemented, domain of application, focused ontology learning tasks, and evaluation metrics and benchmarks. This categorization provides a comprehensive overview of the current state of knowledge and highlights emerging trends and research gaps in the field.

II. RESEARCH METHODOLOGY

In this research, a systematic mapping was conducted to find the application of deep learning models or techniques in ontology learning. According to the authors in [13], a SMS is a method that systematically provides a high-level, organized overview of a field of study by categorizing and synthesizing the existing literature. In contrast to the detailed systematic review, SMS classifies the included studies according to study dimensions such as research topics, methodology, and application domain, thus providing an overview of the area under study. It involves defining research questions, developing a search strategy, selecting studies using predefined criteria, and extracting data to effectively categorize these studies.

A. Research Questions

Table I outlines the key Research Questions (RQs) that have structured and guided the SMS on deep learning-based ontology learning. These are aimed at contributions, methods, data, techniques, tasks, and metrics in selected studies, with motivations focused on tracking research evolution, identifying trends, and evaluating methodologies.

B. Search Strategy

This SMS implemented an extensive search strategy to identify relevant literature related to deep learning-based ontology learning. Various databases known for their extensive coverage of computer science and Artificial Intelligence (AI) research were included: IEEE Xplore, Springer Link, Elsevier-ScienceDirect, Google Scholar, and ACM Digital Library. The target databases were selected based on their ability to provide access to peer-reviewed journals, conference proceedings, and technical reports to ensure a broad and comprehensive search. The search was restricted to peer-reviewed articles, conference proceedings, and high-impact book chapters published between 2015 and September 2024 to maintain a focus on recent and high-quality studies.

The main search string used was: ("Ontology" AND ("Ontology Development" OR "Ontology Construction" OR "Ontology Acquisition" OR "Ontology Enrichment" OR "Ontology Population" OR "Ontology Extension" OR "Ontology Learning")) AND ("Deep Learning" OR "Neural Networks" OR "GNN" OR "Transformer" OR "CNN" OR "RNN" OR "Long Short-Time Memory (LSTM)" OR "LLMs"). This query targeted abstract, title, and keyword fields to capture papers related to ontology work such as development, enrichment, and population associated with deep learning methods such as Neural Networks (NNs), GNNs, or LLMs. This search, with appropriate modifications to individual databases, ensured that the recent advances in the field were covered for each relevant term.

TABLE I. KEY RESEARCH QUESTIONS AND MOTIVATIONS

ID	Research question	Motivation
RQ1	What types of publications are present in selected studies, and how are they distributed over time?	To trace the evolution of publications and highlight innovations
RQ2	What types of empirical studies and contribution are commonly used in selected studies on deep learning for ontology learning?	To identify common methodologies and their reliability in the field
RQ3	What data types are utilized in selected studies for deep learning-based ontology learning?	To understand data diversity and its impact on deep learning models
RQ4	Which deep learning techniques are predominantly applied in selected studies for ontology learning tasks?	To discover effective models and techniques for ontology learning
RQ5	What ontology learning tasks and domains are addressed by selected studies using deep learning?	To map tasks and domains addressed by deep learning methods
RQ6	What evaluation metrics and benchmarks are employed in selected studies to assess deep learning models in ontology learning?	To standardize and improve performance evaluation practices

C. Selection Criteria

1) Inclusion and Exclusion Criteria

The definition of inclusion and exclusion criteria is one of the most important steps in the selection of studies within an SMS, as it guarantees a better quality and relevance of the studies considered suitable for inclusion. These criteria are used to filter out poor or irrelevant papers so that the review can focus on only the most valuable and credible research. The criteria used in the present study are presented in Table II.

2) Study Selection Summary

The initial search in five databases, including IEEE Xplore, Springer Link, Elsevier-ScienceDirect, Google Scholar, and ACM Digital Library, yielded a total of 2765 studies. The selection process included several filtering stages, as shown in the PRISMA flow diagram in Figure 1. First, duplicate studies were identified and removed, further refining the pool to 2164 papers. Second, titles, abstracts, and keywords were screened, reducing the pool to 286 publications. A subsequent full-text review narrowed the selection to 85 studies that met all criteria. The final dataset consisted of 20 peer-reviewed journal articles and 27 conference proceedings.

TABLE II. SUMMARY OF INCLUSION AND EXCLUSION CRITERIA FOR THE SMS

Criteria	Details
Inclusion	Automated/semi-automated ontology learning based on deep learning methods
	Publications that present empirical findings or evaluations
	Publications from 2015 to September 2024
	Peer-reviewed articles, conference papers, and high-impact book chapters
	Highly indexed sources
	Only articles published in English
Exclusion	Manual ontology construction without automation
	Studies that do not focus on deep learning in the context of ontology learning
	Studies lacking empirical findings or evaluations
	Systematic literature reviews or systematic mapping studies
	Non-peer-reviewed content
	Brief studies (less than four pages)
	Non-English publications

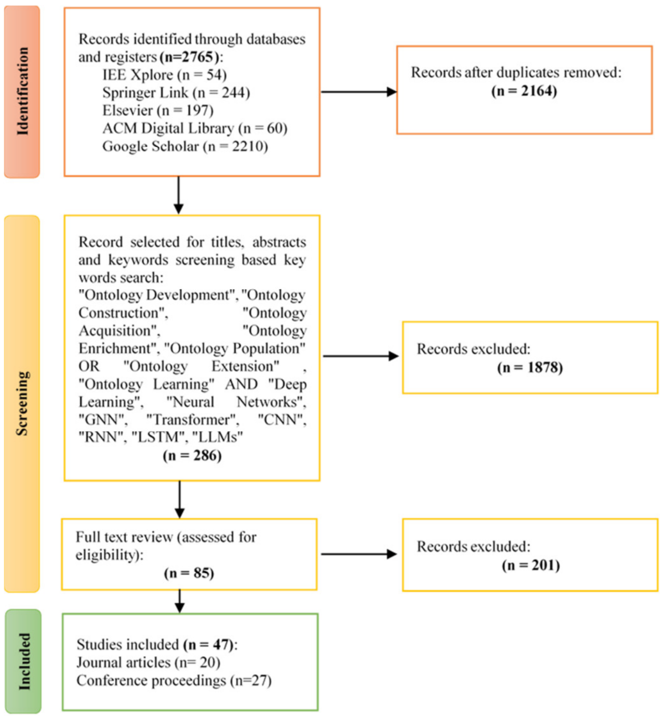

Fig. 1. PRISMA flow diagram illustrating the study selection process.

D. Data Extraction and Analysis

The data extraction process was designed to directly address the RQs posed in the study. Each parameter was carefully selected to provide the necessary information to comprehensively address these questions, as shown in Table III. This approach allowed for a detailed analysis and categorization of trends, methodologies, and applications in deep learning-based ontology learning, supporting the motivations behind each research question.

TABLE III. DATA EXTRACTION AND ANALYSIS PARAMETERS

Parameters	Research Question Addressed	Purpose
Title, DOI, and author(s)	All RQs	Provide bibliographic information for reference purposes
Paper type, source, and year of publication	RQ1	Identify the distribution of publications by time and type
Contribution type	RQ1	Identify the nature and distribution of a contribution, such as frameworks, algorithms
Empirical study type	RQ2	Understand common and frequent approaches and assess their reliability
Type of data used	RQ3	Identify data diversity and how it applies to different models
Deep learning methods/techniques	RQ4	Discover predominant techniques and their effectiveness.
Ontology learning tasks	RQ5	Map tasks and domains addressed by deep learning methods and techniques
Domain/field	RQ5	Describe the application areas of the ontology learning methods
Evaluation metrics/benchmarks	RQ6	Standardize and improve performance evaluation practices

III. RESULTS AND ANALYSIS

A. Publication Trends and Sources (RQ1)

The publication trends of the selected studies, shown in Figure 2, show a gradual increase in research activity and a sharper increase in 2019, when the applied developments in AI technologies made the deep learning applications of ontology learning more interventional. The spike in 2023 marks a peak in innovation, probably due to the emergence of advanced models such as transformers and LLMs. The small dip in 2024 may be due to incomplete access to all publications. Overall, this trend shows the growing maturity and importance of deep learning in ontology learning, with the field gaining momentum and focus.

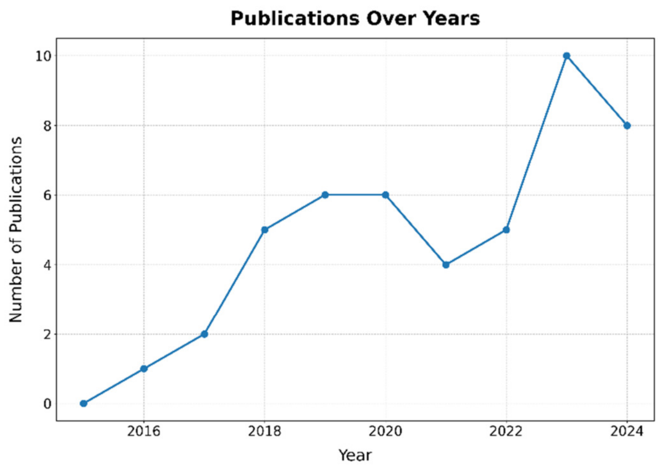

Fig. 2. Evolution of publications in the selected studies.

The chart in Figure 3 shows that Springer Link and ACM Digital Library have the most conference paper sources, accounting for over 60% of the sources, whereas Google Scholar has a diverse mix, including journal articles (41,7%) and conference papers (58,3%). IEEE Xplore has the most conference papers (75%), highlighting its emphasis on conference proceedings. In contrast, Elsevier has the most journal articles (60%), reflecting its focus on more established, peer-reviewed content. Overall, conference papers are the most common publication type across these sources, reflecting their importance in disseminating research in ontology learning and deep learning applications.

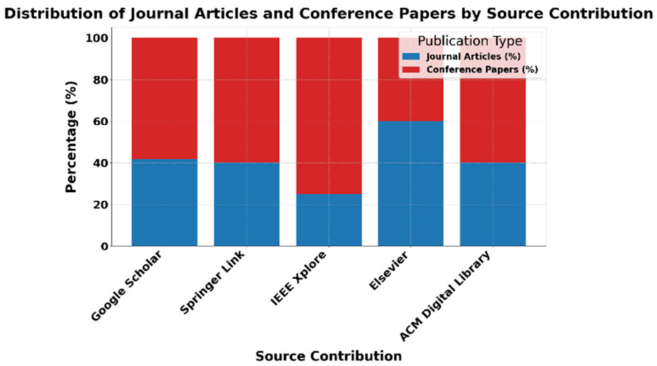

Fig. 3. Publication types and sources.

B. *Contribution Types and Research Approaches (RQ2)*

As illustrated in Figure 4, the distribution of contributions by type is as follows: 53% focus on methods/methodology, making it the most dominant category, followed by frameworks at 22%, reflecting their importance in structuring solutions. Contributions related to algorithms account for 12%, whereas tools/software represent 8% of the total. Case studies/applications represent 5%, emphasizing the application of methods in specific contexts. In terms of empirical study methodologies, 76% of the studies are experimental studies, demonstrating the field's strong reliance on empirical testing to validate new methods and frameworks. Evaluation studies account for 12% and focus on assessing the clinical utility of specific contributions, particularly frameworks. Case studies account for 7%, highlighting the limited application-based research, whereas comparative studies also represent only 4%, highlighting an opportunity for future research to emphasize comparative analyses and practical implementation in real-world contexts.

C. *Data Types Utilized (RQ3)*

Table IV indicates that most works focus on biomedical and clinical data (27%), as in [14] and [15], underlining the critical importance of high-quality ontology learning in the health and life sciences, where its impact can be significant. Textual corpora (18%), as in [16], and knowledge graphs (9%), as in [17], are also prominent, reflecting the reliance on structured and unstructured textual data for extracting and linking concepts. Multimedia data is another emerging trend (13%), showing that deep learning models can be extended to multimedia and other diverse data types, as in [18]. Similarly, domain-specific ontologies (13%) reflect the shift towards addressing more specific and different types of datasets, such as the studies in [19] and [20]. However, scholarly research data (7%) and geographic/environmental data (2%), such as the works in [21] and [22], are less frequently used, indicating opportunities for further exploration in these areas to broaden the application of ontology learning techniques. This distribution highlights the current areas of concentration and also indicates the possibility of increasing the breadth of the studies towards the areas that are currently minimal.

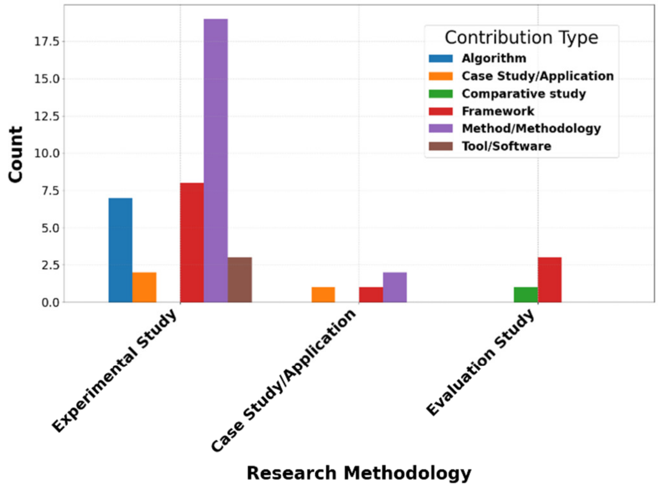

Fig. 4. Research methodologies and contribution types used in the selected studies.

TABLE IV. DATA TYPES USED IN THE SELECTED STUDIES

Category	Description	Count
Textual corpora	Text data from various domains, including academic publications, textbooks, specialized professional materials, and read-write news. This category includes general text mining sources for Natural Language Processing (NLP) and ontology learning	10
Biomedical and clinical data	Data specific to the biomedical and clinical domains, including gene expression data, drug interactions, clinical text, and biomedical ontologies. These datasets are primarily used for medical and biological ontology learning and classification tasks	14
Knowledge graphs	Graph-based datasets such as WordNet, Freebase, and ConceptNet, that represent structured knowledge and relationships between entities, often used for ontology population and extension	4
Multimedia data	Non-textual data such as images, social media content, and tabular data. This includes data such as gene expression images, social images from Facebook, and event logs	7
Domain-specific ontologies	Customized ontologies for specific domains, such as chemistry, agriculture, finance, and information security, that form the foundation for domain-specific ontology enhancement and population	7
Scholarly and research data	Academic publications and research articles used to extract concepts, classify information, and build scholarly ontologies	4
Geographic and environmental data	Geographic information and tabular information about environmental events, ontologies, and HTML pages containing geographic information	1

D. Deep Learning Techniques Applied to Ontology Learning (RQ4)

Figure 5 shows that most studies favor transformer-based models (12) and CNNs (11), highlighting their dominance in deep learning for ontology learning. These models' versatility and advanced capabilities in NLP, as in [23] and [24], and image-based tasks, as in [18], make them essential tools in the field. RNN and its types, including LSTM and Gated Recurrent Unit (GRU), also show promising use (9 studies), as the studies in [25] and [26], which can be related to the applicability of the technique to capture sequential data and text-related tasks. Graph-based models (5 studies) show a converging trend towards incorporating structured knowledge into deep learning techniques, mainly GNNs and knowledge graph embedding, as in the studies in [19] and [27]. Meanwhile, ensemble & hybrid models, as in [28], and LLMs, as in [9], used in 5 studies each, highlight their emerging importance in merging different architectures and leveraging large-scale language models for improved performance. Contextualized embedding models, such as Embeddings from Language Models (ELMo) and Universal Sentence Encoder (USE), appear less frequently (2 studies, [14] and [29]), as newer transformer models may be replacing them. Other techniques, including specialized models such as Generative Adversarial Networks (GANs), as in [30], and contrastive learning, as in [31], account for 2 studies, suggesting that while less common, they offer unique capabilities that complement mainstream methods.

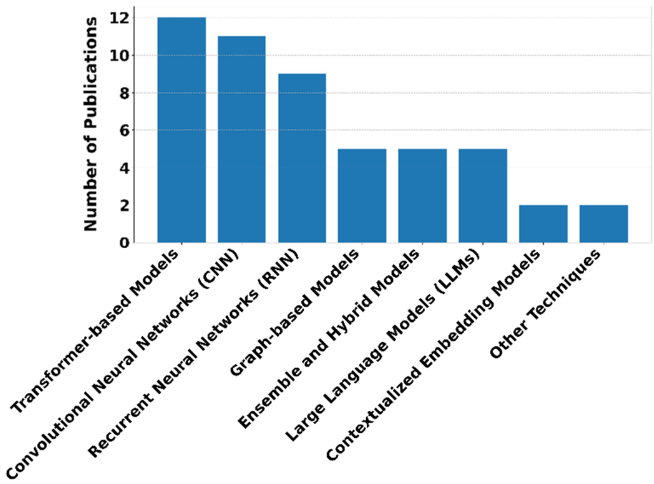

Fig. 5. Deep learning techniques used in the selected studies.

Likewise, Table V shows that the most commonly used RNN techniques are Bi-LSTM and LSTM, indicating a preference for using models that can handle dependencies and long contexts. Other uses of RNNs, such as GRU, Bi-GRU, and the encoder-decoder architecture are evident, but their use is not as common. Table VI shows that CNNs and their variants such as Convolutional Denoising Autoencoders (CDAE), Graph Convolutional Networks (GCN), and particular structures like AlexNet, LeNet, Google Net, and VGG19 are widely used. This is due to the versatility and resilience of CNNs in a variety of problems, including data in image and graph forms. The even distribution among different models implies the application of the convolutional operation regardless of the problem type as structures (GCNs) or visions.

TABLE V. RNN MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
Bi-LSTM	4
LSTM	2
Bi-GRU	1
GRU	2
RNN	1
Encoder-Decoder Architecture	1

TABLE VI. CNN MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
CNN	5
CDAE	2
GCN	2
AlexNet, LeNet, GoogleNet, VGG19	2

Transformer-based models are the most common, as shown in Table VII, with models such as BERT, the GPT series, and others such as RoBERTa and ChatGPT being frequently used. The dominance of these models indicates a trend toward leveraging transformer architectures for their ability to efficiently handle complex NLP tasks. The variety of models suggests that both encoder-only (BERT) and decoder-only (GPT series) architectures are crucial for ontology learning and NLP applications.

TABLE VII. TRANSFORMER BASED MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
BERT	5
GPT series (GPT-3, GPT-4)	3
RoBERTa	2
ELECTRA	1
Transformer-based architectures	1
Triplet-BERT	1
FLAN-T5	1
BLOOM	1
ChatGPT	1
Llama 2	1

As shown in Table VIII, the use of graph-based models is moderate, as GNN, GCN, and Knowledge Graph Embeddings (KGE) imply the importance of structural and relational knowledge for deep learning. This has been done through models like Node2Vec and Neural Tensor Networks (NTNs), highlighting the efforts made in placing graph-based relationships and modeling them appropriately. The diversity in the use of graph models presented in this study reveals that structure and relationships are such features that are important for the task, even though graphs are used less frequently than text and images. As indicated in Table IX, contextualized embedding models are the least used with only two techniques, ELMo and USE. This may be partly due to the fact that newer transformer models, which offer better contextualized embeddings, are used more frequently than these earlier deep learning methods. Nevertheless, these models are still valued for their ability to generate word and sentence embeddings in specific contexts.

TABLE VIII. GRAPH BASED MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
GNN	2
GCN	2
KGE	2
NTN	1
Node2Vec	1

TABLE IX. CONTEXTUALIZED EMBEDDING MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
ELMo	2
USE	1

Table X shows that techniques such as Multilayer Perceptron (MLP), autoencoders, and Siamese Neural Networks (SNNs) suggest a combination of approaches to improve performance. The use of the ensemble learning methods indicates that the application of multiple models can be considered as a technique to achieve high accuracy and adaptability in deep learning applications. The use of hybrid models highlights efforts to integrate diverse architectures for specialized tasks. Table XI demonstrates that models such as GPT-3.5, GPT-4, and others indicate an increasing trend in the application of LLMs in the field. Ontology development and NLP are some of the works to which these models are applied. Their relatively high count suggests that LLMs are becoming increasingly relevant for complex text generation and classification tasks, highlighting a shift towards the use of more generalizable, large-scale models. As shown in Table XII, features specialized models such as GANs and non-standard methods such as contrastive learning may have their advantages, but are less widely used than standard deep learning.

TABLE X. ENSEMBLE AND HYBRID MODELS USED IN THE SELECTED STUDIES

Technique used	Count of studies
Stacking Ensemble Learning	1
Neural Ranker	1
MLP	2
Autoencoders	1
SNN	1

TABLE XI. LLMS USED IN THE SELECTED STUDIES

Technique used	Count of studies
GPT-3.5	2
GPT-4	1
Claude 3 Sonnet	1
Retrieval-Augmented Generation (RAG)	1
Mixtral 8x7B	1
LLMs	1
Ontology Development Kit (ODK)	1

TABLE XII. OTHER DEEP LEARNING TECHNIQUES USED IN THE SELECTED STUDIES

Technique used	Count of studies
GAN	1
Contrastive Learning	1

E. Ontology Learning Tasks and Domains (RQ5)

Figure 6 shows that the most frequent task with 31.3% is ontology construction, as in [32] and [33], which would mean that in this dataset most of the works are focused on building the structure of the ontologies. It is followed by the task of ontology enrichment, as in [31] and [28], with 20.8%, which puts emphasis on the refinement and extension processes of already developed ontologies to make them more functional. Ontology population comes in third place with 12.5%, as in [26] and [34], which emphasizes that such structures actually need to be filled with information if they are ever to become useful. Less frequent activities are ontology annotation (8.3%) as in [35], ontology extension (6.3%) as in [36], and ontology matching (6.3%) as in [37], indicating that these are complementary and specialized activities. The lower figures for ontology embedding as in [30], knowledge graph completion as in [17], and ontology classification as in [15], all at 4.2%, indicate a representative emergent intersection with advanced AI methods. The least frequent activity is ontology subsumption prediction as in [24], at 2.1%, suggesting that this is a niche area. Overall, the distribution emphasizes the importance of building and enriching ontological frameworks, with other tasks supporting the development and application of these structures.

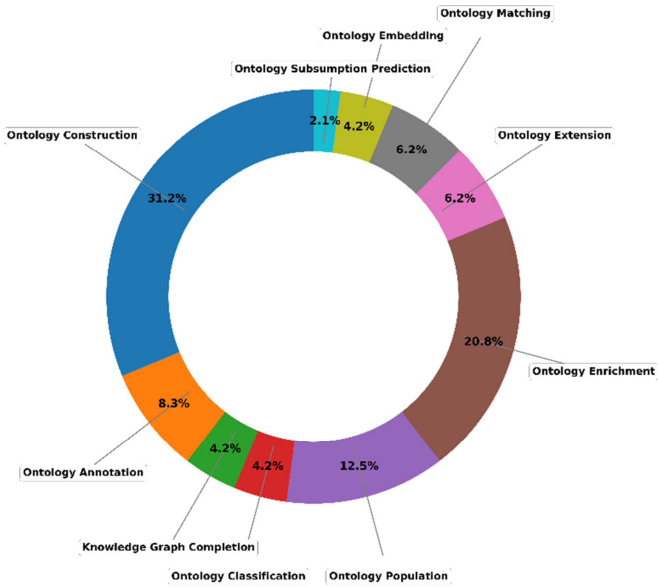

Fig. 6. Ontology learning tasks in the selected studies.

Similarly, Figure 7 shows that the most frequent subtasks of all ontology learning tasks are term extraction, relation extraction, and ontology refinement, representing their core position in the ontology learning process. For instance, term extraction is especially dominant in ontology construction (25.5%) and ontology annotation (40.0%) tasks, confirming its crucial role in the development of ontological frameworks. On the other hand, there are refined tasks, namely ontology matching and ontology extension, which contain subtasks, namely similarity feature construction and axiom learning, suggesting a higher degree of specialization of certain tasks.

Overall, the distribution shows a strong focus on basic activities that form the basis for ontology creation and refinement, with specialized higher-order techniques used in a subset of tasks.

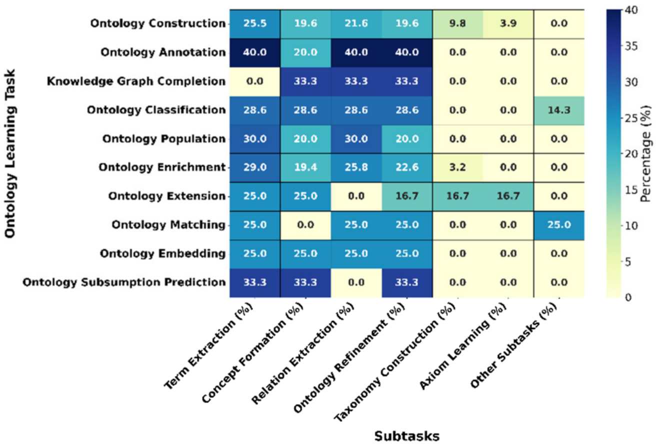

Fig. 7. Ontology learning tasks and subtasks used in the selected studies.

Figure 8 shows that the most common domain where the ontology learning applications have been applied is the biomedical and healthcare domain, as in the works in [38] and [14], since 25.0% of the studies belong to this domain. This indicates a critical need for better structured understanding and improved ontological content in the medical and healthcare context, in line with the previously discussed ontology creation and enhancement.

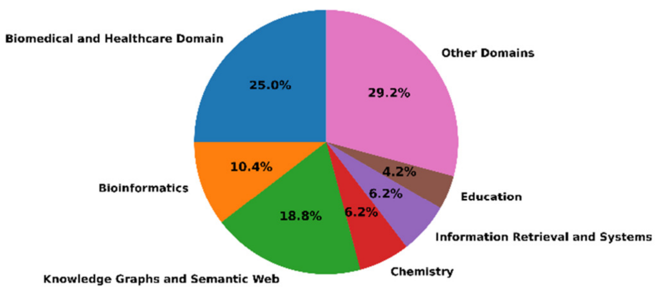

Fig. 8. Distribution of domains in the studied ontology learning applications.

The category of knowledge graphs and semantic web such as the studies in [17] and [37], with 18.8%, underlines the importance of developing and enriching general and multi-domain knowledge structures, according to tasks such as ontology population and enrichment, intended for building and managing various information sources. The field of bioinformatics, with 10.4%, supports this trend of focusing on biological and life sciences, and represents a substantial connection of ontology learning with state-of-the-art AI techniques within these domains, as in the study in [39]. Smaller categories, such as chemistry with 6.3% as in [23], information retrieval and systems also with 6.3% as in [40], and education with 4.2% as in [33] emphasize the narrower but emerging applications in these domains. The category of other domains (9 domains), with a total of 29.2%, really demonstrates not only the breadth but also the versatility of ontology learning in capturing the diverse range of niche and emerging fields like cybersecurity as in [34], finance as in [20], and legal as in [41], where structured knowledge representation is increasingly valuable. Overall, this distribution underlines the predominance of biomedical applications, but also shows the penetration of ontology learning techniques beyond this domain into various other domains.

F. Evaluation Metrics and Benchmarks Employed (RQ6)

Precision, recall, and F1 score remain among the most popular evaluation metrics across a range of ontology learning tasks, including ontology construction, ontology population, and ontology enrichment, as shown in Table XIII. This not only indicates their central role in evaluating the accuracy and relevance of ontology elements, but also probably points to serious limitations of related research work. Specific tasks like ontology classification and knowledge graph completion involve more metrics such as Accuracy, Mean Reciprocal Rank (MRR) as in the work in [14], and Hits@K as in [24], since ranking performance and classification accuracy need to be evaluated in these contexts. Specific tasks also involve special, task-specific measures, for example, ontology construction involves the structural measures of Relationship Richness (RR) and Link Richness (LR) as in [16], while ontology matching involves statistical hypothesis testing and similarity feature construction to address the complexity of matching processes as in [42]. This indicates that while there is a core set of common metrics to ensure consistency, some tasks require specialized evaluation measures to capture the distinct aspects of their processes, suggesting a balance between standardization and task-specific adaptability in ontology learning evaluation practices.

TABLE XIII. EVALUATION METRICS USED IN THE SELECTED STUDIES

Task	Common Metrics	Task-Specific Metrics
Ontology construction	Precision, Recall, F1	RR, AR, LR, Formula correctness
Ontology annotation	Precision, Recall, F1	Jaccard similarity
Knowledge graph completion	Accuracy, MRR, Hits@K	Triple accuracy, Improvement %
Ontology classification	Accuracy, AUC	-
Ontology population	Precision, Recall, F1, mAP	Link classification
Ontology enrichment	Precision, Recall, F1	Insertion rate, Top-k edges
Ontology extension	F1, Precision, Recall, AUC	ROC-AUC
Ontology matching	Precision, Recall, Accuracy	Hypothesis testing, Similarity features
Ontology embedding	Accuracy (Cosine, Euclidean)	Harmonic mean, MRR, Hits@K
Ontology subsumption	MRR, Hits@K	-

Regarding the use of benchmarks within ontology learning, as shown in Table XIV, there is a sharp division: tasks in the biomedical and bioinformatics domains, such as ontology annotation and enrichment, uniformly depend on established benchmarks, like Gene Ontology (GO) as in [43] and SNOMED CT as in [31]. This shows that mature evaluation practices are in place. Other tasks, such as ontology construction and ontology population, do not have standardized benchmarks, so variability can be seen. Various studies use multiple benchmarks due to the maturity of the domain, so that the evaluation can be comprehensive. Whereas some of these fields have standardized standards, others require further development for evaluation protocols to be consistently effective.

TABLE XIV. USE OF BENCHMARKS AND LEVEL OF STANDARDIZATION IN THE SELECTED STUDIES

Task	**Domain**	**Benchmarks**	**Standardization**
Ontology annotation	Biomedical	CRAFT, GO	High
Ontology enrichment	Biomedical	SNOMED CT, MedDRA	High
Knowledge graph completion	General	WordNet, Freebase	High
Ontology matching	General	OAEI	High
Ontology construction	Various	Custom / Not established	Low
Ontology population	Various	Custom / Not established	Low
Ontology embedding	General	BioPortal, WordNet	Medium
Ontology extension	Biomedical/ chemical	ChEBI, Disease ontology	Medium

IV. DISCUSSION

The results of this study show a clear evolution in the field of ontology learning, driven by the advancement of AI technologies, especially since 2019 with the emergence of complex models such as transformers, GNNs and LLMs. At the same time, this growth highlights a number of challenges: first, the growing computational complexity that increases with the power of such models; second, the demand for domain-specific competency, especially in biomedical applications that dominate the field. Data sparsity remains an issue, as evidenced by the limited use of some domains, such as geographic and environmental data. The predominance of experimental studies and the reliance on empirical testing highlight the field's emphasis on practical validation. However, the absence of any kind of comparative and real-world case studies points to a gap that future research should fill in order to increase the applicability of the field. Expanding the application to more underrepresented domains such as finance, legal, and environmental sectors would serve to further increase the impact and applicability of ontology learning.

Although the core metrics of precision, recall, and F1 score are the most important, especially for tasks like ontology construction and population, most of these research areas don't have established benchmarks. For instance, construction and population require standardization efforts. Additionally, as model complexity increases, the need for explainable AI approaches becomes crucial to ensure transparency and interpretability of results. Future work should focus on creating comprehensive benchmarks and providing clear evaluation protocols, especially for new and less standardized domains, so that applications of ontology learning remain reliable, explainable, and adaptable.

V. CONCLUSION

This study systematically maps the application of deep learning in ontology learning, exploring research trends, contribution types, and evaluation methods. It shows a rise in publications since 2019, reflecting the impact of the advent of sophisticated models such as transformers, Graph Neural Networks (GNNs) and Large Language Models (LLMs). The dominance of experimental and empirical approaches shows a primary concern in method development, whereas the prevalence of biomedical and healthcare domains underlines their priority in the field. However, the limited investigations observed in areas such as finance, legal, and the environmental contexts, as well as the minimal use of comparative studies, expose clear research gaps and areas that need to be addressed in future works. The research highlights the lack of standardization in ontology learning, especially in the areas of evaluation metrics and benchmarks, as inconsistencies were observed in tasks such as ontology construction and population. Future research should focus on creating more extensive and field-specific benchmarks, expanding applications to the lacking domains, and highlighting the explainability to improve trust and transparency. Practitioners and researchers can leverage these insights to refine their methodologies and practices to make ontology learning approaches flexible, reliable, and applicable in various domains.

REFERENCES

[1] V. T. da Silva, J. S. dos Santos, R. Thiago, E. Soares, and L. G. Azevedo, "OWL ontology evolution: understanding and unifying the complex changes," *The Knowledge Engineering Review*, vol. 37, 2022, Art. no. e10, https://doi.org/10.1017/S0269888922000066.

[2] Z. A. Mora-Alvarez, O. Hernandez-Uribe, R. A. Luque-Morales, and L. A. Cardenas-Robledo, "Modular Ontology to Support Manufacturing SMEs Toward Industry 4.0," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12271–12277, Dec. 2023, https://doi.org/10.48084/etasr.6454.

[3] L. E. Chan *et al.*, "The Environmental Conditions, Treatments, and Exposures Ontology (ECTO): connecting toxicology and exposure to human health and beyond," *Journal of Biomedical Semantics*, vol. 14, no. 1, Feb. 2023, Art. no. 3, https://doi.org/10.1186/s13326-023-00283-x.

[4] A. Haller, J. D. Fernández, M. R. Kamdar, and A. Polleres, "What Are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web," *Journal of Data and Information Quality*, vol. 12, no. 2, pp. 1–34, May 2020, https://doi.org/10.1145/3369875.

[5] A. Donkers, D. Yang, B. de Vries, and N. Baken, "Semantic Web Technologies for Indoor Environmental Quality: A Review and Ontology Design," *Buildings*, vol. 12, no. 10, Oct. 2022, Art. no. 1522, https://doi.org/10.3390/buildings12101522.

[6] R. Thirumahal, G. Sudha Sadasivam, and P. Shruti, "Semantic Integration of Heterogeneous Data Sources Using Ontology-Based Domain Knowledge Modeling for Early Detection of COVID-19," *SN Computer Science*, vol. 3, no. 6, Aug. 2022, Art. no. 428, https://doi.org/10.1007/s42979-022-01298-4.

[7] Y. He, J. Chen, D. Antonyrajah, and I. Horrocks, "BERTMap: A BERT-Based Ontology Alignment System," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 5, pp. 5684–5691, Jun. 2022, https://doi.org/10.1609/aaai.v36i5.20510.

[8] M. Wang *et al.*, "Big Data Health Care Platform With Multisource Heterogeneous Data Integration and Massive High-Dimensional Data Governance for Large Hospitals: Design, Development, and Application," *JMIR Medical Informatics*, vol. 10, no. 4, Apr. 2022, Art. no. e36481, https://doi.org/10.2196/36481.

[9] H. Babaei Giglou, J. D'Souza, and S. Auer, "LLMs4OL: Large Language Models for Ontology Learning, " in *Proceedings of the 22nd International Semantic Web Conference, Part I,* Athens, Greece, 2023, pp. 408–427, https://doi.org/10.1007/978-3-031-47240-4_22.

[10] P. Rathee and S. K. Malik, "Merging Operation for Domain Ontologies in Semantic Web: Some Issues," in *Proceedings of Third Doctoral Symposium on Computational Intelligence*, Lucknow, India, 2022, pp. 551–560, https://doi.org/10.1007/978-981-19-3148-2_47.

[11] B. G. Humm *et al.*, "New directions for applied knowledge-based AI and machine learning," *Informatik Spektrum*, vol. 46, no. 2, pp. 65–78, Apr. 2023, https://doi.org/10.1007/s00287-022-01513-9.

[12] L. Zemmouchi-Ghomari, "Ontology and Machine Learning: A Two-Way Street to Improved Knowledge Representation and Algorithm Accuracy," in *Proceedings of International Conference on Paradigms of Communication, Computing and Data Analytics*, New Delhi, India, 2023, pp. 181–189, https://doi.org/10.1007/978-981-99-4626-6_15.

[13] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009, https://doi.org/10.1016/j.infsof.2008.09.009.

[14] E. Schumacher, A. Mulyar, and M. Dredze, "Clinical Concept Linking with Contextualized Neural Representations," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 2020, pp. 8585–8592, https://doi.org/10.18653/v1/2020.acl-main.760.

[15] I. Cohen, E. David, and N. S. Netanyahu, "Supervised and Unsupervised End-to-End Deep Learning for Gene Ontology Classification of Neural In Situ Hybridization Images," *Entropy*, vol. 21, no. 3, Mar. 2019, Art. no. 221, https://doi.org/10.3390/e21030221.

[16] J. Chen and J. Gu, "ADOL: a novel framework for automatic domain ontology learning," *The Journal of Supercomputing*, vol. 77, no. 1, pp. 152–169, Jan. 2021, https://doi.org/10.1007/s11227-020-03261-7.

[17] E. Amador-Domínguez, P. Hohenecker, T. Lukasiewicz, D. Manrique, and E. Serrano, "An Ontology-Based Deep Learning Approach for Knowledge Graph Completion with Fresh Entities," in *16th International Conference on Distributed Computing and Artificial Intelligence*, Ávila, Spain, 2019, pp. 125–133, https://doi.org/10.1007/978-3-030-23887-2_15.

[18] M. Muscetti, A. M. Rinaldi, C. Russo, and C. Tommasino, "Multimedia ontology population through semantic analysis and hierarchical deep features extraction techniques," *Knowledge and Information Systems*, vol. 64, no. 5, pp. 1283–1303, May 2022, https://doi.org/10.1007/s10115-022-01669-6.

[19] K. Sanju Saravanan and V. Bhagavathiappan, "Innovative agricultural ontology construction using NLP methodologies and graph neural network," *Engineering Science and Technology, an International Journal*, vol. 52, Apr. 2024, Art. no. 101675, https://doi.org/10.1016/j.jestch.2024.101675.

[20] C. Saetia, J. Phruetthiset, T. Chalothorn, M. Lertsutthiwong, S. Taerungruang, and P. Buabthong, "Financial Product Ontology Population with Large Language Models," in *Proceedings of TextGraphs-17: Graph-based Methods for Natural Language Processing*, Bangkok, Thailand, 2024, pp. 53–60.

[21] V. K. Kommineni, B. König-Ries, and S. Samuel, "From human experts to machines: An LLM supported approach to ontology and knowledge graph construction." arXiv, Mar. 13, 2024, https://doi.org/10.48550/arXiv.2403.08345.

[22] A. Mukanova *et al.*, "LLM-Powered Natural Language Text Processing for Ontology Enrichment," *Applied Sciences*, vol. 14, no. 13, Jun. 2024, Art. no. 5860, https://doi.org/10.3390/app14135860.

[23] M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski, and J. Hastings, "Interpretable ontology extension in chemistry," *Semantic Web*, vol. 15, no. 4, pp. 937–958, Oct. 2024, https://doi.org/10.3233/SW-233183.

[24] J. Chen, Y. He, Y. Geng, E. Jiménez-Ruiz, H. Dong, and I. Horrocks, "Contextual semantic embeddings for ontology subsumption prediction," *World Wide Web*, vol. 26, no. 5, pp. 2569–2591, Sep. 2023, https://doi.org/10.1007/s11280-023-01169-9.

[25] P. Manda, S. SayedAhmed, and S. D. Mohanty, "Automated ontology-based annotation of scientific literature using deep learning," in *Proceedings of The International Workshop on Semantic Big Data*, Portland, OR, USA, 2020, pp. 1–6, https://doi.org/10.1145/3391274.3393636.

[26] M.-H. Su, C.-H. Wu, and P.-C. Shih, "Automatic Ontology Population Using Deep Learning for Triple Extraction," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Lanzhou, China, 2019, pp. 262–267, https://doi.org/10.1109/APSIPAASC47483.2019.9023113.

[27] H. M. Zahera *et al.*, "Tab2Onto: Unsupervised Semantification with Knowledge Graph Embeddings," in *Proceedings of the 19th Extended Semantic Web Conference, ESWC 2022 Satellite Events,* Hersonissos, Greece, 2022, pp. 47–51, https://doi.org/10.1007/978-3-031-11609-4_9.

[28] A. V. Semenova and V. M. Kureichik, "Ensemble of classifiers for ontology enrichment," *Journal of Physics: Conference Series*, vol. 1015, no. 3, May 2018, Art. no. 032123, https://doi.org/10.1088/1742-6596/1015/3/032123.

[29] L. M. Sanagavarapu, V. Iyer, and R. Reddy, "A Deep Learning Approach for Ontology Enrichment from Unstructured Text." arXiv, Dec. 16, 2021, https://doi.org/10.48550/arXiv.2112.08554.

[30] Y. Geng *et al.*, "Disentangled Ontology Embedding for Zero-shot Learning," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 443–453, https://doi.org/10.1145/3534678.3539453.

[31] H. Dong, J. Chen, Y. He, Y. Gao, and I. Horrocks, "A Language Model Based Framework for New Concept Placement in Ontologies," in *Proceedings of the 21st International Conference on The Semantic Web, ESWC 2024*, *Part I,* Hersonissos, Greece, 2024, pp. 79–99, https://doi.org/10.1007/978-3-031-60626-7_5.

[32] J. Wang, J. Liu, and L. Kong, "Ontology Construction Based on Deep Learning," in *Proceedings of the 12th KIPS International Conference on Ubiquitous Information Technologies and Applications and the 9th International Conference on Computer Science and its Applications*, Taichung, Taiwan, 2017, pp. 505–510, https://doi.org/10.1007/978-981-10-7605-3_83.

[33] C.-S. Lee *et al.*, "A Study on AI-FML Robotic Agent for Student Learning Behavior Ontology Construction," in *2020 International Symposium on Community-centric Systems*, Tokyo, Japan, 2020, pp. 1–6, https://doi.org/10.1109/CcS49175.2020.9231339.

[34] H. Gasmi, J. Laval, and A. Bouras, "Cold-start cybersecurity ontology population using information extraction with LSTM," in *2019 International Conference on Cyber Security for Emerging Technologies*, Doha, Qatar, 2019, pp. 1–6, https://doi.org/10.1109/CSET.2019.8904905.

[35] P. Manda, L. Beasley, and S. D. Mohanty, "Taking a Dive: Experiments in Deep Learning for Automatic Ontology-based Annotation of Scientific Literature." bioRxiv, Jul. 10, 2018, Art. no. 365874, https://doi.org/10.1101/365874.

[36] S. Althubaiti, Ş. Kafkas, M. Abdelhakim, and R. Hoehndorf, "Combining lexical and context features for automatic ontology extension," *Journal of Biomedical Semantics*, vol. 11, no. 1, Jan. 2020, Art. no. 1, https://doi.org/10.1186/s13326-019-0218-0.

[37] X. Xue, J. Guo, M. Ye, and J. Lv, "Similarity Feature Construction for Matching Ontologies through Adaptively Aggregating Artificial Neural Networks," *Mathematics*, vol. 11, no. 2, Jan. 2023, Art. no. 485, https://doi.org/10.3390/math11020485.

[38] W. H. Bangyal *et al.*, "Constructing Domain Ontology for Alzheimer Disease Using Deep Learning Based Approach," *Electronics*, vol. 11, no. 12, Jun. 2022, Art. no. 1890, https://doi.org/10.3390/electronics11121890.

[39] F. Zhang *et al.*, "A Deep Learning Framework for Gene Ontology Annotations With Sequence- and Network-Based Information," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*,

vol. 18, no. 6, pp. 2208–2217, Nov. 2021, https://doi.org/10.1109/TCBB.2020.2968882.

[40] A. Balali, M. Asadpour, and S. H. Jafari, "COfEE: A comprehensive ontology for event extraction from text," *Computer Speech & Language*, vol. 89, Jan. 2025, Art. no. 101702, https://doi.org/10.1016/j.csl.2024.101702.

[41] Y. Ren, J. Han, Y. Lin, X. Mei, and L. Zhang, "An Ontology-Based and Deep Learning-Driven Method for Extracting Legal Facts from Chinese Legal Texts," *Electronics*, vol. 11, no. 12, Jun. 2022, Art. no. 1821, https://doi.org/10.3390/electronics11121821.

[42] L. Korel, U. Yorsh, A. S. Behr, N. Kockmann, and M. Holeňa, "Text-to-Ontology Mapping via Natural Language Processing with Application to Search for Relevant Ontologies in Catalysis," *Computers*, vol. 12, no. 1, Jan. 2023, Art. no. 14, https://doi.org/10.3390/computers12010014.

[43] C. Chen and L. Luo, "GOGCN: using deep learning to support insertion of new concepts into gene ontology," in *5th International Conference on Information Science, Electrical, and Automation Engineering*, Wuhan, China, 2023, pp. 641–647, https://doi.org/10.1117/12.2689526.