# Synergizing Machine Learning & Symbolic Methods: A Survey on Hybrid Approaches to Natural Language Processing

Rrubaa Panchendrarajan[a,*], Arkaitz Zubiaga[a]

[a] *School of Electronic Engineering and Computer Science, Queen Mary University of London, 327 Mile End Rd, Bethnal Green, London, E1 4NS, United Kingdom*

**Abstract**

The advancement of machine learning and symbolic approaches have underscored their strengths and weaknesses in Natural Language Processing (NLP). While machine learning approaches are powerful in identifying patterns in data, they often fall short in learning commonsense and the factual knowledge required for the NLP tasks. Meanwhile, the symbolic methods excel in representing knowledge-rich data. However, they struggle to adapt dynamic data and generalize the knowledge. Bridging these two paradigms through hybrid approaches enables the alleviation of weaknesses in both while preserving their strengths. Recent studies extol the virtues of this union, showcasing promising results in a wide range of NLP tasks. In this paper, we present an overview of hybrid approaches used for NLP. Specifically, we delve into the state-of-the-art hybrid approaches used for a broad spectrum of NLP tasks requiring natural language understanding, generation, and reasoning. Furthermore, we discuss the existing resources available for hybrid approaches for NLP along with the challenges and future directions, offering a roadmap for future research avenues.

*Keywords:* Hybrid NLP, Machine Learning, Symbolic Methods, Hybrid Approaches, Natural Language Processing

*Corresponding author

*Email addresses:* r.panchendrarajan@qmul.ac.uk (Rrubaa Panchendrarajan), a.zubiaga@qmul.ac.uk (Arkaitz Zubiaga)

*URL:* www.zubiaga.org (Arkaitz Zubiaga)

## 1. Introduction

The field of machine learning has witnessed remarkable progress over the past few decades achieving human-level performance in various Natural Language Processing (NLP) tasks. Large language models, in particular, have garnered global attention in the last couple of years, captivating not only the NLP community but also integrating into the daily routines of numerous professionals (Min et al., 2023). Meanwhile, symbolic methods have seen significant advancements in representing human cognitive capabilities and knowledge-rich data enabling it to be more interpretable and comprehensible (Dale, 2000). However, these two main pillars of computer science research possess their own set of advantages and disadvantages.

Machine learning approaches are stronger in learning patterns and relationships in data via optimization strategies. However, they often fall short in capturing and interpreting the factual knowledge required for most downstream NLP tasks. Instead, they attempt to mimic the facts and knowledge present in the training data (Pan et al., 2023). This hinders both traditional machine learning approaches and deep learning methods from achieving high performance in knowledge-intensive tasks. Further, recent studies (Petroni et al., 2019) have demonstrated that even robust large language models trained using vast amounts of training data and parameters suffer from hallucinations generating non-factual responses (Zhang et al., 2023b). This raises significant concerns regarding the trustworthiness of such expensive large language models. Similarly, symbolic methods are stronger in resembling human cognitive abilities by explicitly capturing the knowledge required for a task. On the other hand, symbolic methods have shown poor learning skills compared to machine learning approaches, especially in handling dynamic data and generalizing the knowledge beyond training data (Pan et al., 2023).

Recent focus has been turned into bridging the gap between machine learning approaches and symbolic methods to overcome the limitations of both when applied independently while retaining their strengths (Zhu et al., 2023). This hybrid approach enables the statistical methods to utilize knowledge-enriched input data to improve the inference with external knowledge and to produce interpretable results. At the same time, symbolic methods are empowered with statistical learning to incorporate semantic knowledge and generate generalized knowledge representations. In particular, the combination of deep learning and symbolic methods has paved the way for a new research era called Neuro-symbolic methods (Hamilton et al., 2022) aiming to develop trustworthy and interpretable NLP solutions. Furthermore, recent studies (Sarker et al., 2021) have shown, that neuro-symbolic solutions gain benefits in four key research aspects including interpretability, generalization to handle both small training data and out of distributions, and error recovery due to aggregated advantages of both deep learning and symbolic methods. This has drawn the attention of the artificial intelligence research community to develop effective hybrid solutions to solve various real-world problems.

The field of NLP has also begun to embrace hybrid techniques to develop effective real-world solutions. Especially, the existence of a larger number of textual knowledge bases has aided the rapid adoption of hybrid techniques over the past few years. Further, the representation of natural language is inherently a symbolic representation, hence the emergence of this new field, *Hybrid NLP*, is an unsurprising development within the field of NLP. Inspired by this rapidly progressing research paradigm, this survey presents an overview of the literature on hybrid approaches that combine machine learning techniques and symbolic methods for NLP. Specifically, we address the following research questions in this survey.

- What are the hybrid techniques used in different NLP tasks requiring Natural language understanding (NLU), Natural language generation (NLG), and Natural language reasoning (NLR)?
- What are the recent developments in the adoption of hybrid approaches for NLU, NLG, and NLR tasks?
- What kinds of symbolic representations and corresponding resources are used in the literature for the development of hybrid solutions in NLP?
- What are the challenges faced during the adoption of hybrid solutions to the field of NLP?
- What are the immediate future research directions involving hybrid NLP?

Figure 1 depicts the NLP applications discussed in this survey.

The closest study of our survey paper was presented by Zhu et al. (2023) during a recent tutorial. The authors briefly introduced the knowledge augmentation methods used for NLP. Apart from this study, many survey articles

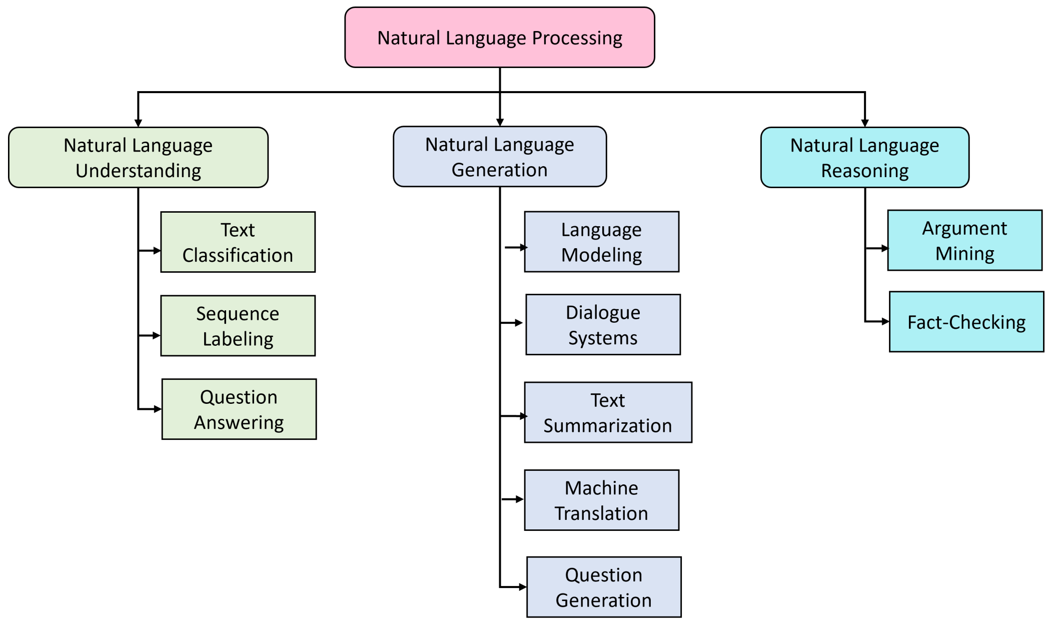



Figure 1: Taxonomy of Natural Language Processing Applications

have focused on a particular aspect of machine learning (e.g. large language models (Safavi and Koutra, 2021; Hu et al., 2023; Yin et al., 2022; Pan et al., 2023)) or symbolic methods (e.g. knowledge graphs (Schneider et al., 2022)) or a particular NLP task (e.g. text generation (Yu et al., 2022b)) employing hybrid solutions. Further details on the related surveys are discussed in Section 7. To the best of our knowledge, none of the existing studies present the hybrid approaches used for NLP tasks in a wider outlook. Specifically, we discuss state-of-the-art hybridization techniques combining machine learning and symbolic methods for a wide range of NLP tasks requiring, NLU, NLG, and NLR. For each NLP task, we analyzed the research articles retrieved using the keywords *Symbolic methods*, *Neuro-symbolic*, *Knowledge-augmented*, *Knowledge-enriched*, *Knowledge-aware*, and *Knowledge base* along with the task name in Google Scholar, as well as other articles related to or citing those. After an in-depth analysis of research articles published during the past 6 years (2018 - 2023), we chose the state-of-the-art hybrid approaches used for each NLP task, and present them in this survey paper. Articles were selected for inclusion in the survey based on relevance to the use of hybrid approaches.

Figure 2 shows the trend on the number of publications retrieved using the

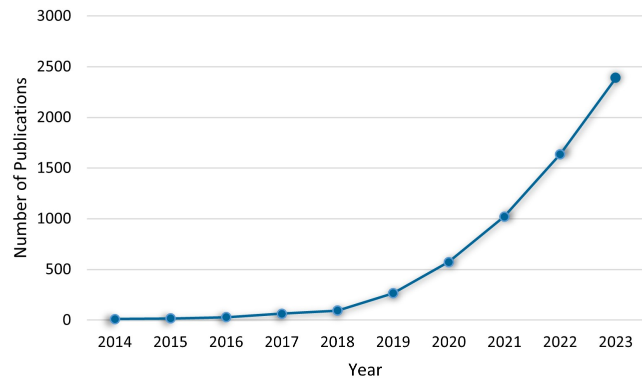



Figure 2: Number of Papers Published related to Hybrid NLP

keywords *Knowledge-augmented NLP, Knowledge-enriched NLP, Knowledge-aware NLP, Neuro-symbolic NLP*[1]. This shows the surge of interest in the topic of *Hybrid NLP*, which in turn motivates our survey detailing the current research trend in this topic and the potential future directions.

The remainder of the paper is organized as follows. Section 2 introduces the background knowledge required to understand the rest of the paper. State-of-the-art hybrid approaches proposed in the literature are presented in Section 3, Section 4, and Section 5 focusing on three fundamental objectives of Natural language processing, NLU, NLG, and NLR respectively. Section 6 discusses open challenges in the development of hybrid approaches. We discuss related survey articles as well as how they differ from ours in Section 7. Further, we detail future directions in Section 8. Finally, we conclude this review in Section 9 along with a discussion of possible future directions.

## 2. Background

This section introduces machine learning and symbolic methods, and the latest advancements in both fields. Further, we define the term *hybrid approaches* with respect to the context of this review.

[1] https://www.dimensions.ai/

### *2.1. Machine Learning*

Machine learning has a long history of development, especially in NLP with the transformation from statistical models such as Naive Bayes classification, Hidden Markov models, Support Vector Machines, and Logistic regression to powerful neural models. Specifically, the transformers (Vaswani et al., 2017) proposed in 2017 was a breakthrough in the evolution of neural architectures. The model comprised layers of encoders and decoders with multi-headed attention, enabling them to be very effective and powerful in sequence transduction tasks. Later, the architecture was explored by other researchers to introduce several influential models, resulting in a family of transformer models. Some of the notable models include BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), GPT (Generative Pre-trained Transformer) (Radford and Narasimhan, 2018), BART (Bidirectional Auto-Regressive Transformers) (Lewis et al., 2020), and their variations such as RoBERTa (Liu et al., 2019b), XLM-r (Kalyan et al., 2021), Sentence-BERT (Reimers and Gurevych, 2019). Often these pre-trained models are fine-tuned with limited training data to achieve state-of-the-art performance in various fields related to machine learning, including NLP.

### *2.2. Symbolic Methods*

Similar to how humans treat words and sequences of words for communication, symbolic methods simulate this cognitive behavior, by considering entities referred to as *symbols* to refer to the basic unit of information in a symbolic system. The symbols can be obtained from different data structures and processed to produce new symbols (Hoehndorf et al., 2017). Numerous symbolic representations have been used to explicitly capture the knowledge. The following are some of the notable symbolic representations apart from the general graph structure.

- Knowledge base (KB) - Knowledge base refers to a collection of items usually associated with an in-built index, mapping the index to items (Zouhar et al., 2022). The item can be either structured or unstructured conveying a piece of information in any mode. Wikipedia[2] is an

[2]https://www.wikipedia.org/

example of a knowledge base with a semi-structured collection of documents. Knowledge graphs and ontologies, discussed in what follows, are examples of structured knowledge bases.

- Knowledge graphs (KG) - KG stores inter-related factual knowledge of entities in a directed labeled graph. The nodes represent the entities and the edges denote the relationship between the two entities. This enables the retrieval of information about entities from a knowledge graph in the form of a triplet, *¡head entity, relationship, tail entity¿* (Pan et al., 2023). Figure 3a shows an example of entities and their relationship from Wikidata5M KG.[3] Table 1 summarizes the knowledge graph types used in the literature.

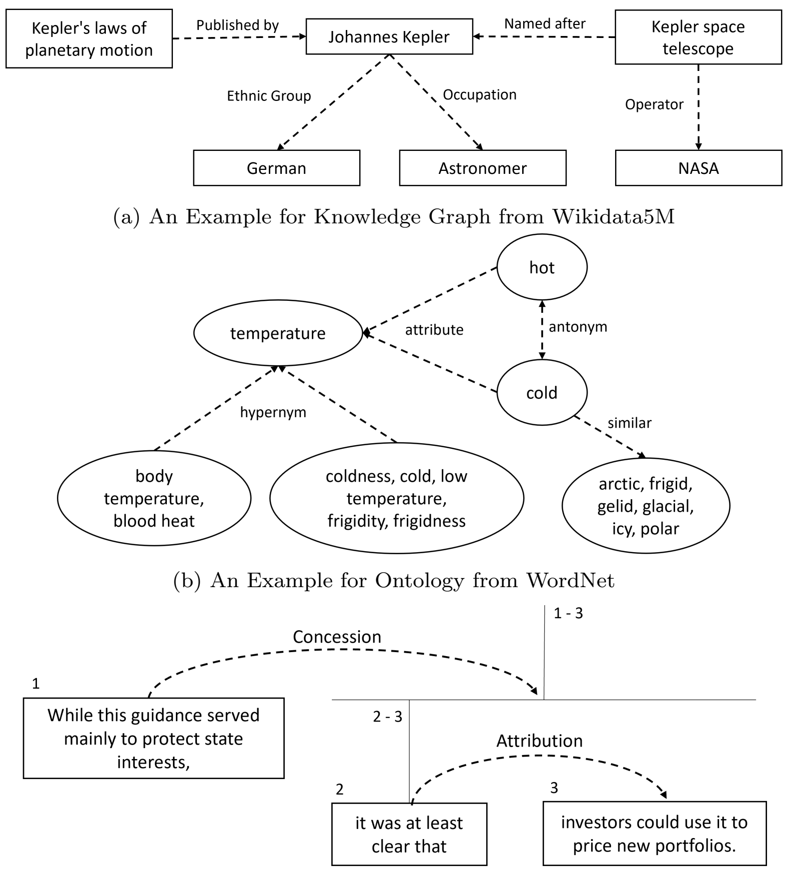
- Ontologies - Ontologies are referred to as a formal, explicit specification of a shared conceptualization. They store concepts and their relationship (e.g. hypernyms) in a graph structure, and serve as the basis for KGs. Some of the well-known ontologies include WordNet (Fellbaum, 2010), Probase (Wu et al., 2012), and FrameNet (Baker et al., 2003). Figure 3b shows an example from the WordNet ontology[4].

- Rhetorical Structure Theory (RST) graphs - RST describes the organization of natural text and the relationship between their parts. The text structure is identified as a tree, explaining the transition point of the relations and the extent of the relation (Hou et al., 2020). The construction of the RST graph is referred to as RST parsing, which involves the identification of roles for different granularity of text (phrases, sentences, paragraphs, and collection of paragraphs). Figure 3c shows an example of RST graph (Hou et al., 2020).

### *2.3. Hybrid Approaches*

The term *hybrid* is used often in the scientific domain referring to the synergization of two aspects for solving a problem. Especially the following scenarios are considered as *hybrid* in computer science research.

[3]https://deepgraphlearning.github.io/project/wikidata5m
[4]https://lexicala.com/review/2020/mccrae-rudnicka-bond-english-wordnet/




(a) An Example for Knowledge Graph from Wikidata5M

(b) An Example for Ontology from WordNet

(c) An Example for RST Graph

Figure 3: Examples for KG, Ontology, and RST Graph.

Table 1: Summary of Knowledge Graph Types.

| Type | Example | Notable Feature |
|---|---|---|
| Encyclopedic KG | WikiData (Vrandečić and Krötzsch, 2014) | Constructed from Wikipedia |
| | Freebase (Bollacker et al., 2008) | Stores more structured Wikipedia data |
| | Wikidata5M (Wang et al., 2021) | Combines WikiData and Wikipedia via entity linking |
| | Dbpedia (Auer et al., 2007) | Constructed from Wikipedia and links other KBs |
| | YAGO (Suchanek et al., 2007) | Constructed from Wikipedia, GeoNames, and WordNet |
| | NELL (Carlson et al., 2010) | Derived from web content |
| Commonsense KG | ConceptNet (Speer et al., 2017) | Covers a wide range of commonsense concepts |
| | ATOMIC (Sap et al., 2019) | Covers everyday commonsense inferential knowledge |
| | ATOMIC$^{20}_{20}$ (Hwang et al., 2021) | Covering social, physical, and eventive aspects of everyday inferential knowledge |
| | ASER (Zhang et al., 2020b) | An eventuality KG, with events as nodes and discourse relations as edges |
| | TransOMCS (Zhang et al., 2020a) | Derived from ConceptNet and ASER |
| | CausalBanK (Li et al., 2021) | Covers commensense related to causal |
| Domain-specific KG | UMLS (Bodenreider, 2004) | Specific to medical domain |
| | Finance KG (Liu et al., 2019c) | Specific to finance domain |
| | Geo KG (Zhu et al., 2017) | Specific to geology domain |
| | BKG Huo et al. (2022) | Bibliography KG |
| Multimodal KG | IMGpedia (Ferrada et al., 2017) | Incorporates both text and image of DBpedia resources |
| | MMKG (Liu et al., 2019a) | Incorporates both text and image of Freebase entities |
| | Richpedia (Wang et al., 2020b) | Incorporates both text and image of Wikepedia entities |

- Hybrid models - Combining different types of models or architectures to solve a problem, e.g. integrating statistical machine learning approaches with deep learning, combining two types of deep learning models such as Long short-term memory (LSTM) network and Convolutional Neural Network (CNN). Hybrid models are often used to overcome the limitations of individual models via integration.

- Hybrid systems - Combining multiple solutions to solve a single goal, e.g. classifying items and clustering the items classified as relevant for the primary task. Hybrid systems are often employed to achieve multiple objectives in pursuit of a primary goal.

- Hybrid data - Combining different types of data (e.g. different data structures, different modalities, different data sources) to achieve a common goal. Hybridizing data gives access to rich sources of information to solve the problem. However, the challenge in handling hybrid data escalates and requires more techniques to retrieve and process the hybrid information.

In this survey, we focus on the synergization of machine learning and symbolic methods for natural language processing as *hybrid* approach, and the hybridization of these two techniques may lead to hybrid models or systems or data. Figure 4 summarizes the machine learning approaches, symbolic methods, and hybrid approaches discussed in this paper.

Compared to the other two hybridization settings, hybrid data is a widely used technique for injecting symbolic representation into machine learning models. The knowledge bases can be directly queried using searching techniques such as Elastic search[5] or can be linked using techniques such as entity linking (Schneider et al., 2022) and keyword matching. The retrieved information can either be directly injected into machine learning models or transformed into vector representations. The pre-trained language models and topic models can be used to obtain vector representation for textual information retrieved from knowledge bases. Similarly, graph embedding techniques can be used to generate vector representations for nodes and edges retrieved. Table 2 presents the graph embedding techniques used in the literature.

[5] https://www.elastic.co/

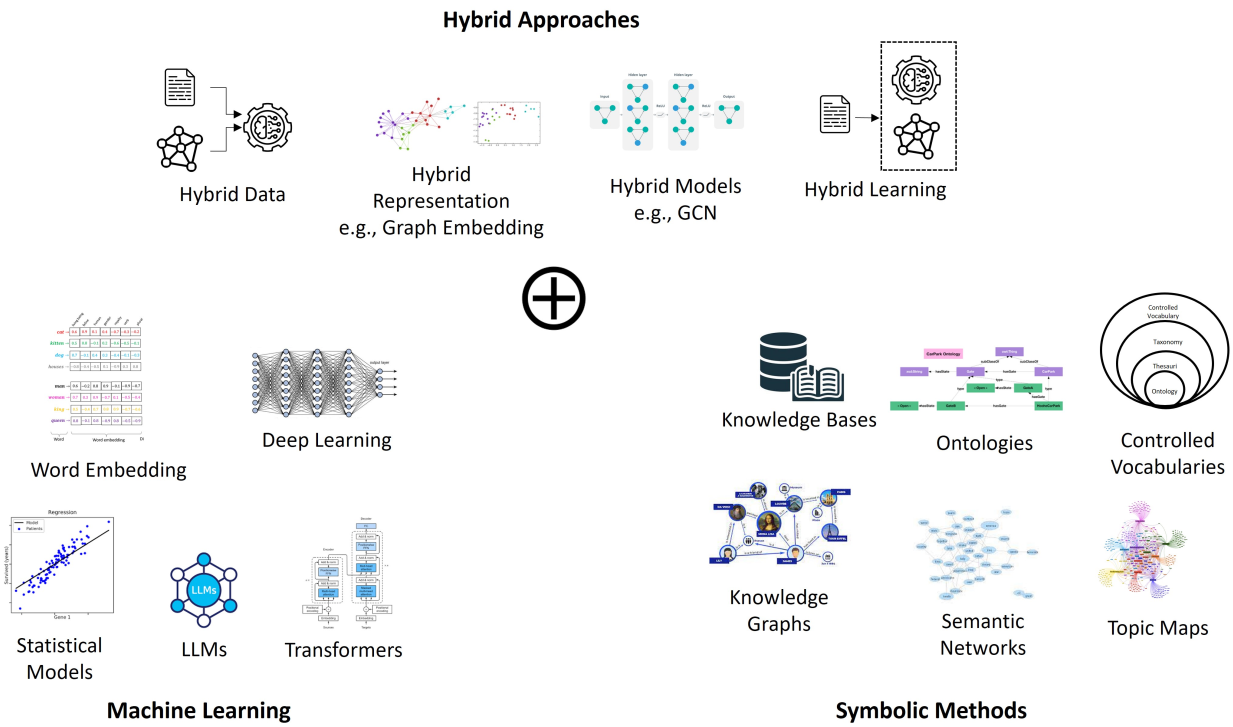



Figure 4: Summary of Machine Learning Approaches, Symbolic Methods, and Hybrid Approaches

Table 2: Graph Embedding Techniques.

| Type | Example | Applicable Knowledge Base |
|---|---|---|
| Tensor factorization methods | RotatE (Sun et al., 2019)<br>QuantE (Zhang et al., 2019a) | KG |
| Translation-based methods | TransE (Bordes et al., 2013)<br>TransA (Jia et al., 2016) | KG |
| Neural Network-based Methods | Graph Convolutional Networks (GCN) (Schlichtkrull et al., 2018)<br>Graph Attention Networks (GAN) (Velickovic et al., 2017) | Graphs |
| Graph Traversal Methods | Node2Vec (Grover and Leskovec, 2016)<br>Struc2Vec (Ribeiro et al., 2017) | Graphs |
| Language Modeling | WordNet Embedding (Saedi et al., 2018) | KB with associated text corpus |

## 3. Hybrid Approaches to Natural Language Understanding

Natural Language Understanding (NLU) is a branch of Natural Language Processing (NLP) that focuses on understanding the meaning and semantics of text. This includes various downstream NLU tasks such as text classification, sequence labeling, and question answering. This section briefly introduces the latest hybrid approaches used to produce state-of-the-art performance in popular NLU applications.

### *3.1. Text Classification*

Text classification is the task of automatically assigning a label from a predefined set of labels for an input text (Sebastiani, 2002). While there are enormous amounts of text classification tasks explored across various domains, sentiment analysis, stance detection, and language detection are some of the prominent tasks that attracted considerable focus of the NLU research community. With the introduction of large language models and their exceptional performance across various NLP tasks, text classification tasks generally exploit fine-tuning language models on limited domain-specific training data (Pittaras et al., 2023).

A straightforward approach for adopting hybrid approaches for text classification is by incorporating external knowledge obtained via symbolic methods as additional input to the classification models. Škrlj et al. (2021) used word taxonomies from WordNet (Fellbaum, 2010) to generate new semantic features for text classification. The authors transformed the input document into a semantic feature representation by extracting hypernyms of words from the documents and obtaining their double normalized TF-IDF (Term Frequency - Inverse Document Frequency) scores, followed by a wide range of feature selection techniques. The authors observed a significant improvement in the performance across six short text classification tasks when these external features were used with neural classifiers. Similarly, Liu et al. (2022) extracted concept words from the Probase knowledge base (Wang et al., 2010) and obtained a concept word embedding by aggregating word embeddings of all the concept words present in a text. This knowledge-enriched vector representation was provided as additional input to a neural model for text classification.

Instead of constructing a knowledge-rich vector representation as additional input, Liu et al. (2020a) generated a sentence tree as input to the BERT model by injecting knowledge graph triplets to corresponding places in the sentence. Injected knowledge was controlled via techniques such as soft-positioning and visible matrix. The authors experimented with three language-specific knowledge graphs and demonstrated promising results for the knowledge-injected BERT named K-BERT in 12 NLP tasks including text classification, sequence labeling, and question answering. Li et al. (2022) followed the same approach and showed that developing a domain-specific knowledge graph benefits more when the knowledge graph triplets are injected into the input text.

An alternative to injecting external knowledge as inputs to classifiers is using *hybrid architectures* which can combine symbolic representation learning models with traditional classifiers to improve the inference. One of the pioneering works in this direction for text classification was experimented by Yao et al. (2019). The authors proposed TextGCN, a text graph convolutional network neural network architecture that models documents and words as nodes in a graph to generate a heterogeneous graph. Embedding representation for words and documents or nodes in the graph is jointly learned as supervised learning while performing text classification. This enabled the authors to transform the text classification problem into a node classification problem in a heterogeneous graph. TextGCN was shown to outperform standalone neural models in several text classification tasks. Extending TextGCN, various graph convolution network architectures (Gu et al., 2023) were proposed in the literature for text classification.

### *3.2. Sequence Labeling*

Sequence labeling is the task of assigning a label at the word level instead of the sentence or phrase level from a predefined set of labels (El Mekki et al., 2022). Named entity recognition, part-of-speech tagging, and language detection are some of the well-known sequence labeling tasks. Information predicted at a word level in sequence labeling tasks generally serves as input features for various other downstream NLP tasks.

Similar to injecting external knowledge extracted from a knowledge graph into an input text, several attempts have been made to adopt similar hybrid techniques for sequence labeling tasks. In particular, the entities present in a text can be linked with external information for knowledge-augmented learning. Motivated by that direction, Enhanced Language RepresentatioN with Informative Entities (ERNIE) (Zhang et al., 2019b) was proposed by Zhang et al. to learn language representation with infused entity knowledge. The authors identified entities in the text first, aligned them with a knowledge graph, and obtained their entity embeddings using the graph embedding technique TransE (Bordes et al., 2013). Following that, the authors trained an auto-encoder architecture with random entities masked in the Wikipedia text corpus to enforce the representation learning to incorporate entity knowledge. This model was shown to outperform BERT-based architectures in various NLP tasks including sequence labeling with limited finetuning (Wang et al., 2020c). The underlying idea of ERNIE was later extended by replacing the auto-encoder with BERT-base architectures for sequence labeling (Hu et al., 2022).

Linking text with Wikipedia data for improving sequence labeling, especially for recognizing named entities, has been explored widely in recent research works (Tedeschi et al., 2021; Wang et al., 2022b; Boros et al., 2022). Tedeschi et al. (2021) exploited Wikipedia data to automatically create annotated training data. The authors utilized one-to-one linkage between Wikipedia articles to generate named entity candidates and automatically annotated them as abstract concepts or named entities using the knowledge bases WordNet (Fellbaum, 2010) and BabelNet.[6] Different from this approach, Wang et al. (2022b) used Wikipedia as an external knowledge source to extract information related to the input sentence. The authors used Elastic search to extract relevant content from the Wikipedia articles by considering the input sentence as a query. The retrieved contexts were injected into the XLM-R (Kalyan et al., 2021) model along with the input text for sequence classification.

Boros et al. (2022) experimented with the injection of both the relevant context retrieved using Elastic search on Wikipedia and knowledge graph embedding for knowledge-enriched training. The authors used Wikidata5M (Wang et al., 2021), a large-scale knowledge graph, and the RotatE embedding model (Sun et al., 2019) to generate the embedding of entities in the knowledge graph. Similar to the search in Wikipedia articles, an embedding-based search was performed in the knowledge graph to retrieve entity embeddings similar to the input document's embedding representation. Finally, the authors convert the relevant context into embedding representation using Sentence-BERT (Reimers and Gurevych, 2019) and inject both context embedding and entity embedding into the classification model. Apart from these techniques, graph neural architectures were also explored for sequence

[6] https://babelnet.org/

labeling tasks by converting the word-level classification problem into a node classification problem in a graph (Gui et al., 2019), enabling the authors to capture non-sequential dependencies via a graph structure.

### *3.3. Question Answering*

Question Answering (QA) is the task of generating or finding relevant answers for a question in natural language (Zaib et al., 2022). It is one of the notable NLU tasks that often requires commonsense and external knowledge, hence demands the adoption of hybrid solutions. While training large language models in Wikipedia data and books enabled them to show exceptional performance in QA tasks (Mitra et al., 2019), numerous other hybrid solutions have been explored in the literature recently.

Similar to other NLU applications, an evident hybrid technique is to provide external knowledge required to perform the QA task as input to the inference model. Mitra et al. (2019) performed an Elastic search using question and answers as a query in external sources and retrieved the top 50 related sentences. Authors reranked the retrieved sentences using sentence similarity and provided the top 10 sentences as additional information to a BERT model. For effective context retrieval, Karpukhin et al. (2020) proposed obtaining dense vector representation of passages using both the TF-IDF approach and a BERT model. During the inference, the dot-product between the dense representation of the question and the passage was used to determine the relevant context to be used as the external input to the QA model. Similarly, Noraset et al. (2021) experimented with different document retrieval mechanisms to extract information from Wikipedia. The authors used the BM25F retrieval algorithm (Pérez-Agüera et al., 2010), Google search API, and TF-IDF for retrieving the relevant content, and injected it as additional information to a Bi-LSTM model. The experiment results showed the content retrieved using the BM25F algorithm was very effective for QA.

The integration of knowledge graphs into QA tasks is an expected development in the progression of hybrid approaches. Lv et al. (2020) extracted evidence from the structured knowledge base ConceptNet (Speer et al., 2017) and plain texts of Wikipedia articles and generated knowledge graphs for both sources for further inference. A graph convolution network (GNN) was used to extract node representation from the two knowledge graphs and the obtained contextual representation of questions and answers was fed into another graph-based attention model to choose the right answer. In addition to

Table 3: Summary of Hybrid Techniques used for NLU Tasks.

| Task | Research | Hybrid Approach | | | | | |
|---|---|---|---|---|---|---|---|
| | | Input | | | Model | Transfer Learning | Training Data Construction |
| | | KB Text | KG Triplet | KG Embedding | | | |
| Text Classification | Škrlj et al. (2021); Liu et al. (2022) | ✓ | - | - | - | - | - |
| | Liu et al. (2020a); Li et al. (2022) | - | ✓ | - | - | - | - |
| | Yao et al. (2019); Gu et al. (2023) | - | - | - | ✓ | - | - |
| Sequence Labeling | Wang et al. (2020c) | - | - | ✓ | - | - | - |
| | Hu et al. (2022) | - | ✓ | - | - | - | - |
| | Tedeschi et al. (2021) | - | - | - | - | - | ✓ |
| | Wang et al. (2022b) | ✓ | - | - | - | - | - |
| | Boros et al. (2022) | ✓ | - | ✓ | - | - | - |
| | Gui et al. (2019) | - | - | - | ✓ | - | - |
| Question Answering | Mitra et al. (2019); Karpukhin et al. (2020); Noraset et al. (2021) | ✓ | - | - | - | - | - |
| | Lv et al. (2020) | ✓ | - | ✓ | ✓ | - | - |
| | Feng et al. (2020) | - | - | ✓ | ✓ | - | - |
| | Yasunaga et al. (2021); Zhang et al. (2022, 2023a) | - | - | ✓ | ✓ | ✓ | - |

the enriched information available in each node in the knowledge graph, Feng et al. (2020) proposed techniques to utilize relational paths using a Multi-hop graph relation network, leading to more interpretable models. The proposed approach chooses a sub-graph related to the input query and both node embedding and path embeddings are obtained using a GNN architecture. Finally, the correct answer was chosen using a text encoder model given the embedding representations of questions and answers. Yasunaga et al. (2021) extended this work to obtain a subgraph related to the QA context and then score each node in the subgraph using a language model. Later the authors jointly learned the representation of QA context and nodes in the graph using a graph neural network. The proposed model was shown to outperform other knowledge graph augmented language models in QA tasks in commonsense and biomedical domains. Following this approach, similar ideas of mutually exchanging information between language models and knowledge graphs were also explored in the literature for the QA task (Zhang et al., 2022, 2023a).

Table 3 summarizes the hybrid techniques used for natural language understanding tasks. Here, we consider hybrid architectures such as Graph

Neural Network modeling both distributed representation and symbolic representation as hybrid *Models*. Hybrid solutions, which jointly learn the distributed representation and symbolic representation by mutually exchanging the knowledge are marked as *Transfer Learning* approaches. It can be observed that the injection of external knowledge as text data and the injection of knowledge graph triplets and knowledge graph embedding are widely used as hybrid solutions across all three tasks. Further, the learning of knowledge graph embedding using graph neural networks is commonly used for question-answering tasks, possibly due to the demand of the task for retrieving supporting knowledge-enriched evidence to find the right answer.

## 4. Hybrid Approaches to Natural Language Generation

Natural Language Generation (NLG) is a fundamental task in Natural Language Processing, whose aim is to produce meaningful text in natural language, generally by giving it a prompt as input. This requires semantic and syntactic understanding of the language to generate text which makes it challenging while also ensuring its applicability across a broad spectrum of NLG tasks such as dialogue systems, summarization, machine translation, and question answering. NLG solutions generally follow the encoder-decoder architecture (Vaswani et al., 2017), where the encoder understands the input text or prompt, and generates hidden states interpreted by the decoder to generate meaningful text (Yu et al., 2022b). However, the text generated by those powerful models often fails to match human responses due to the limited knowledge available in the training data and the lack of generalization capabilities. This demands rapid embracement of hybrid techniques to generate knowledge-enhanced text. This section presents state-of-the-art hybrid approaches used across a range of prominent NLG tasks.

### *4.1. Language Modeling*

Language modeling is the task of learning a universal representation of the language from an unlabelled text corpus (Rosenfeld, 2000). The task is often modeled as a next word or token prediction, where the language model is trained to predict the next word given its previous or surrounding words in a piece of text. Substantial effort has been made in the direction of integrating external knowledge into language models.

Integration of knowledge sources into the input representation of the language models is a prominent way of infusing external knowledge prior to the inference. Analyzing in this direction, Peters et al. (2019) performed entity linking and language modeling jointly as a multitask learning. The authors used an existing deep learning solution (Ganea and Hofmann, 2017) for linking entities in the text to Wikipedia pages (i.e. for wikification) and obtained their embedding representation from entity descriptions. The resulting entity embedding was injected into the language model to generate the knowledge-enhanced representation of the text, and both entity linking and language models were jointly optimized. The authors observed an increase in the ability to recall facts in the resulting model called KnowBERT. Instead of linking entities, Ji et al. (2020) extracted the relevant subgraph from a knowledge graph using the Multi-hop technique and input their embedding representation by aggregating the node embedding obtained via a graph neural network. This concept representation was combined with the output of the encoder for predicting the next word. Liu et al. (2021) extended this idea by modifying the encoder-decoder architecture with a dedicated encoder and decoder augmented with an embedding representation obtained from the knowledge graph. Further, a dedicated convolutional neural network was used to generate the vector representation for concept words from the subgraph. Following the utilization of node embeddings as input for language representation learning, jointly optimizing the node embedding representation as well as the language representation is observed as a promising direction of improvement (Wang et al., 2021; Yu et al., 2022a).

Different from these approaches, Xiong et al. (2019) proposed to modify the training objective to force the language model to learn about real-world entities. The authors identified and linked entities mentioned in the text to Wikipedia and generated negative statements of the corresponding text by randomly replacing the entity occurrences with the names of the same entity types. During the training, the model learns to identify the correct entity. Similarly, Zhou et al. (2020) modified the training objective to enforce the model to generate concept-aware text. The first objective imposed the model to predict the original sentence given some unordered keywords of the sentence, whereas the next objective was aimed at recovering the order of concepts in a sentence given a shuffled list of concepts. Here, the authors define verbs, nouns, and proper nouns present in a sentence as concepts. Instead of modifying the training objective, Guan et al. (2020) post-trained the language models on sentences reconstructed from a knowledge graph. The authors converted commonsense triplets from ConceptNet (Speer et al., 2017) and ATOMIC (Sap et al., 2019) into readable sentences using a templatebased method (Levy et al., 2017).

Generating text with complex ideas may require capturing the knowledge from structured or unstructured knowledge from external sources. Recently, this objective has been studied as a Knowledge graph to text generation problem, where the information available from external sources is converted into a knowledge graph first and the output text is generated based on the knowledge graph. In this direction, Koncel-Kedziorski et al. (2019) propose GraphWriter, an extension of the transformer model for knowledge-graph-to-text generation. Following this study, various extensions of it were proposed (Cai and Lam, 2020) for encoding structural information in the knowledge graph.

### *4.2. Dialogue Systems*

Dialogue systems are designed to coherently converse with humans in natural language (Ni et al., 2023). This requires understanding the language, recalling the conversation history, and producing accurate responses. Undoubtedly, the ability to generate precise responses hinges on the understanding of the external world and utilizing commonsense.

One of the pioneering works in the hybrid application for dialogue systems was experimented by Ghazvininejad et al. (2018). The authors introduced two encoders dedicated to encoding the conversation history as well as the external facts, enabling the responses to be conditioned on both factors. The authors extracted focus phrases containing entities from the input query and collected raw text related to the focus phrases from external sources such as Wikipedia using entity linking techniques. A Recurrent Neural Network (RNN) encoder is used as a fact-encoder to convert the raw text with related facts into a hidden state in the proposed encoder-decoder model. Meng et al. (2020) experimented with a similar solution by backing up the conversion using a domain-specific knowledge base and applying a dedicated RNN to encode the background knowledge. Instead of encoding all the related facts retrieved, Dinan et al. (2018) used an attention-based component to carefully choose the relevant information gathered from external sources, and used a shared encoder to encode the knowledge and the dialogue context. Here, the TF-IDF vector representation of the articles and input query was used to retrieve relevant context from Wikipedia. The authors proposed a generative transformer memory network capable of retrieving relevant information from large memory and generating responses conditioned on both relevant information and dialogue history.

Integration of knowledge graphs into encoder-decoder models is an anticipated research trajectory in dialogue systems study. As evidence of this, Zhou et al. (2018) proposed an encoder-decoder model coupled with graph attention mechanisms. The authors retrieved one knowledge graph per word present in the input query and converted it into a vector representation using the graph attention mechanism. This vector representation was concatenated with the vector presentation of the corresponding word and provided as input to the encoder-decoder model. The decoder model was also combined with a dynamic graph attention mechanism to attend to all the relevant knowledge graphs retrieved to generate the output. Instead of utilizing the existing knowledge graphs, Zhang et al. (2020c) generated a concept graph by starting with grounded concepts present in the input and expanding it to more meaningful conversations by traversing through the related concepts. Following the other knowledge graph integration approaches, the author encoded the concept graph into a vector representation using a graph neural network and inputted to the encoder-decoder model along with the input query.

### *4.3. Text Summarization*

Text summarization is one of the core challenges in NLG which aims to generate summaries based on sources ranging from a single document to a collection of documents (El-Kassas et al., 2021). There are two types of underlying approaches for text summarization namely, extractive summarization and abstractive summarization. The first approach strives to choose key sentences or phrases from the source, and it is often solved as a ranking or scoring task of existing sentences in the source. On the other hand, abstractive summarization aims at producing the summary by constructing sentences or phrases using words available in the source which is commonly modeled as an NLG problem. The latest studies have shown that the summary generated by NLG models suffers from factual inconsistency issues, demanding more robust solutions.

Aiming at resolving the factual inconsistency issue, Cao et al. (2018) extracted fact descriptions from source sentences in the form of *(subject, predicate, object)* using Open Information Extraction (OpenIE) tool (Etzioni et al., 2008) and a dependency parser. The fact descriptions were provided as additional input to a neural model composed of two encoders and a dual attention decoder. Li et al. (2018) experimented with a similar strategy by introducing a shared encoder for external knowledge and input source, followed by a single decoder to generate the summary. A similar approach was carried out by Wang et al. (2022a), where the authors extracted knowledge graph triplets related to the input text, mapped them into a low dimensional vector space and trained a graph embedding classifier to determine whether the triplet should be included in the summary or not. The embedding of triplets classified as key information was fed into a decoder along with the output of the input encoder for the summary generation.

Topic models are also integrated with summarization models to genre topic-aware summaries in the literature. Researching in this direction, Narayan et al. (2018) attempted to enforce the generation of topic-aware summaries by integrating the topic models with neural approaches. The authors first applied the topic model to the source document and input the topic distribution as an additional input of an attention-based convolutional encoder-decoder model. This enabled the model to associate each word in the document with key topics and condition the output words on the topic distribution of the document. Here, the Latent Dirichlet Allocation (LDA) model (Blei et al., 2003) was used to extract the topics from input documents.

Similar to abstractive summarization, extractive summarization techniques have also adopted hybrid approaches to effectively model cross-sentence relations prior to the selection of summary-worthy sentences from the source. Wang et al. (2020a) proposed a heterogeneous graph-based neural network to model the inter-sentence relationships. The authors constructed a heterogeneous graph by modeling words and sentences as nodes in the graph. Semantic features of the nodes and edges were modeled using various techniques, including Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) based sentence representation and TF-IDF-based edge weights. Graph attention networks (GAN) combined with transformers were used to obtain the final representation of nodes. Finally, the authors chose sentence nodes in the heterogeneous graph for summary generation via node classification. Different from this approach, Xu et al. (2020) modeled the source document as a Rhetorical Structure Theory (RST) graph and a coreference graph, to capture long-term dependencies among the discourse units in the input document. Here, the coreference graph was constructed using the entities and their coreferences. Both the document and graph were encoded using a BERT model and GCN respectively, and the encoded information was used to predict whether the input sentence should appear in the summary or not.

### 4.4. Machine Translation

Machine translation involves the automated conversion of text from one language to another (Lopez, 2008). Initially, rule-based approaches and statistical approaches were prevalent in this field and later neural machine translation (NMT) turned out to be a key milestone in the current era. Compared to other NLG tasks, machine translation requires less information from external sources as it is enforced to preserve the content during the conversion from the source language to the target language. However, enhancing the input to NMT with linguistic features such as morphological analysis, part-of-speech tags, and dependency labels is shown to improve the quality of the task (Sennrich and Haddow, 2016; Chen et al., 2018). Bastings et al. (2017) extended this idea by applying a graph convolution network on the dependency trees to obtain a dense vector representation for the sentence structure. Apart from utilizing the linguistic features, Chen et al. (2018) aided the translation using search engines by extracting similar source sentences and their corresponding translation. Among the retrieved sentence pairs, top K sentences were chosen using edit distance and provided as additional input to an attention-based NMT model, enabling it to carefully attend to relevant sentence pair examples.

### 4.5. Question Generation

The question generation task in NLG involves the automatic generation of questions from a given passage or document about a topic or context (Kurdi et al., 2020). This task serves as an underlying objective of various other NLG tasks such as conversation systems and plays a key role in various domains including education. While the nature of the task may vary depending on the type of question or answer (Mulla and Gharpure, 2023), the primary objective of question answering remains consistent: producing meaningful questions that are both syntactically and semantically accurate.

Generating questions will be centered on a certain topic or context, and the knowledge bases can serve as a rich resource of the topic or context for the question generation model. Therefore, a straightforward solution to develop a hybrid approach for question generation is either training a text generation model by extracting the text from the knowledge base as input source (Du and Cardie, 2018) or applying rule-based techniques on structured knowledge bases such as ontologies (Stasaski and Hearst, 2017) and knowledge graphs (Reddy et al., 2017) to produce questions. Elsahar et al. (2018) utilized both textual and structured context of the knowledge base Freebase (Bollacker et al., 2008). The authors extracted the triplets, text descriptions containing the subject and object of the triplet, and the phrase containing the lexicalization of the predicate of the triplet from the knowledge base. Both the textual and structured information extracted were fed as input to an attention-based encoder-decoder model for question generation. Instead of directly injecting the knowledge graph triplets as input to the question generation model, Kumar et al. (2019) observed a significant improvement in the performance, when the embedding representation of the triplet was utilized as the input. The authors used pre-trained TransE embeddings (Bordes et al., 2013) of Freebase from OpenKE tool (Han et al., 2018) for the experiment. Following these studies, further advancements were observed in this direction by integrating graph neural networks for embedding knowledge graphs into the question generation model (Chen et al., 2023, 2019).

Table 4 summarizes the existing hybrid techniques for NLG tasks. It can be observed that a wide range of techniques are adopted for NLG compared to NLU tasks, especially as hybrid inputs and learning techniques. A notable hybrid input will be the injection of the whole knowledge base as input to the machine learning model, where the requirement of the task is to convert or derive text from the knowledge base (e.g. knowledge graph to text generation, knowledge base to question generation). Hybrid learning techniques are mainly used with language modeling tasks for developing knowledge-aware and generalized models. While the injection of KB textual content, structured content, or embedding representations are widely used across many NLG tasks, text summarization tasks exploit various other hybrid approaches such as injection of RST and coreference graph embedding and topic vectors. The machine translation task demands less amount of external knowledge compared to the other NLG tasks, hence, very little attention has been given to the development of hybrid solutions for machine translation. Apart from these hybrid solutions for various NLG tasks, another promising research direction related to NLG will be the generation of knowledge-aware explanations of inferences using hybrid solutions (Yang et al., 2023).

## 5. Hybrid Approaches to Natural Language Reasoning

Natural language reasoning (NLR) aims to integrate diverse knowledge sources such as encyclopedic and commonsense knowledge, to draw new logical conclusions about the actual or hypothetical world (Yu et al., 2023b).

Table 4: Summary of Hybrid Techniques used for NLG Tasks.

| Task | Research | Hybrid Approach | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Input | | | | | | Model | Learning | | Training Data Construction |
| | | KB Text | KG Triplet | KG Embedding | Graph Emebdding | KB | Topic | | Transfer Learning | Objective | |
| Language Modeling | Peters et al. (2019) | ✓ | - | - | - | - | - | - | - | - | - |
| | Ji et al. (2020) | - | - | ✓ | - | - | - | ✓ | - | - | - |
| | Liu et al. (2021) | - | - | ✓ | - | - | - | - | - | - | - |
| | Wang et al. (2021) | - | - | ✓ | - | - | - | - | ✓ | - | - |
| | Yu et al. (2022a) | - | - | ✓ | - | - | - | ✓ | ✓ | - | - |
| | Xiong et al. (2019); Zhou et al. (2020) | - | - | - | - | - | - | - | - | ✓ | - |
| | Guan et al. (2020) | - | - | - | - | - | - | - | - | - | ✓ |
| | Koncel-Kedziorski et al. (2019); Cai and Lam (2020) | - | - | - | - | ✓ | - | - | - | - | |
| Dialogue Systems | Ghazvininejad et al. (2018); Dinan et al. (2018); Liu et al. (2019d); Meng et al. (2020) | ✓ | - | - | - | - | - | - | - | - | - |
| | Zhou et al. (2018); Zhang et al. (2020c) | - | - | ✓ | - | - | - | ✓ | - | - | - |
| Text Summarization | Cao et al. (2018); Li et al. (2018) | ✓ | - | - | - | - | - | - | - | - | - |
| | Wang et al. (2022a) | - | - | ✓ | - | - | - | ✓ | - | - | - |
| | Narayan et al. (2018) | - | - | - | - | - | ✓ | - | - | - | - |
| | Wang et al. (2020a) | - | - | - | - | - | - | ✓ | - | - | - |
| | Xu et al. (2020) | - | - | - | ✓ | - | - | - | - | - | - |
| Machine Translation | Bastings et al. (2017) | - | - | - | - | - | - | ✓ | - | - | - |
| | Chen et al. (2018) | ✓ | - | - | - | - | - | - | - | - | - |
| Question Generation | Reddy et al. (2017); Du and Cardie (2018) | - | - | - | - | - | - | - | - | - | ✓ |
| | Stasaski and Hearst (2017) | - | - | - | - | ✓ | - | - | - | - | - |
| | Elsahar et al. (2018) | ✓ | ✓ | - | - | - | - | - | - | - | - |
| | Kumar et al. (2019) | - | - | ✓ | - | - | - | - | - | - | - |
| | Chen et al. (2023, 2019) | - | ✓ | - | - | - | ✓ | - | - | - | - |

The knowledge can be integrated from both implicit (e.g. pre-trained language models) and explicit sources (e.g. knowledge bases). Reasoning plays a vital role in various NLU and NLG tasks, where neither memorizing the knowledge in the training data nor understanding the context is sufficient for deriving conclusions and requiring the integration of knowledge. This section presents hybrid approaches used for notable NLR tasks including argument mining, and automated fact-checking.

### *5.1. Argument Mining*

Argument mining involves the identification of structured argument data from unstructured text, including the identification of premise and conclusion (Lawrence and Reed, 2019). This enables understanding of the individual components of the arguments and their relationships used to convey the overall message. Argument mining is widely used for the development of qualitative assessment tools for online content, grabbing the attention of both policymakers and social media researchers.

Integration of knowledge graphs, Wikipedia, search engines, and pre-trained language models are often used as a solution for argument mining to infuse external knowledge and commonsense required for the inference (Fromm et al., 2019; Abels et al., 2021; Saadat-Yazdi et al., 2022, 2023). Fromm et al. (2019) integrated three knowledge sources (i.e. Word2Vec (Mikolov et al., 2013), DBpedia knowledge graph (Auer et al., 2007), and a pre-trained BERT model) for classifying a sentence as an argument or not for the given topic. The authors obtained vector representations of the sentence and the topic using the Word2Vec model and input them into a BiLSTM to encode them. Triplets of the entities present in the argument were extracted from the knowledge graph and converted into embedding vectors using a graph embedding technique TransE (Bordes et al., 2013). Encoded topic and argument vectors and entity embeddings were used to fine-tune a BERT classifier. Similarly, Abels et al. (2021) integrated topic modeling, Wikipedia, Knowledge graph, and search engine for argument mining. The authors learned topics using Latent Dirichlet Allocation (LDA) (Blei et al., 2003) from entity-linked Wikipedia pages from the input sentence. Further, they obtained the subgraph related to the topic words from the existing knowledge graph, Wikidata (Vrandečić and Krötzsch, 2014). To resolve the incompleteness issue in Wikipedia, the authors constructed another knowledge graph using content that resulted in a Google search for the topic words.

Finally, the authors extracted evident paths from both knowledge graphs using a breadth-first search, converted each path into a vector representation using a Bi-LSTM, and used it as additional input for argument mining.

### 5.2. Automated Fact-Checking

Automated fact-checking is an essential task for detecting and mitigating the impact of misinformation (Guo et al., 2022; Zeng et al., 2021). This is generally composed of three stages: 1. claim detection to identify sentences with check-worthy or verifiable claims; 2. evidence retrieval to extract supporting statements of the claim; and 3. claim verification to validate whether the retrieved claim is true or not based on the evidence. An intermediate stage of 'claim matching' is sometimes added before the 'evidence retrieval', where the claim matching task consists in grouping together claims that can be resolved with the same fact-check and therefore need not be treated separately in the subsequent stages (Kazemi et al., 2021). The evidence retrieval and claim verification tasks are often combined and handled as fact verification (Guo et al., 2022). It is evident that leveraging hybrid knowledge sources and techniques can significantly enhance fact-verification tasks for precise inference.

Fact-checking using knowledge sources was often resolved by constructing knowledge graphs using the evidence gathered and executing path detection algorithms in knowledge graphs for claim verification (Ciampaglia et al., 2015; Shi and Weninger, 2016; Shiralkar et al., 2017) until the integration of large language models. Addressing this aspect, Zhou et al. (2019) integrated Wikipedia data, the BERT model, and a graph neural network for claim verification. The authors retrieved related sentences to the claim from Wikipedia using MediaWiki API[7] and chose the top 5 relevant sentences using the hinge loss function. Both the claim and relevant sentences were encoded using the BERT model and input to a graph neural network for veracity detection of the claim. Zhong et al. (2020) extended this approach by explicitly modeling the relationship between the evidence sentences by constructing an evidence graph. The authors applied the AllenNLP tool (Gardner et al., 2017) for semantic role labeling and modeled arguments and links between arguments as nodes and edges in the evidence graph. A Graph-enhanced contextual representation of the words in the evidence graph was

[7] https://www.mediawiki.org/wiki/API

extracted by the pre-trained model XL-net (Yang et al., 2019) and inputted to graph neural network for veracity classification of claims. Adopting this methodology, numerous inference techniques using graph neural networks and evidence graphs for fact-checking have been employed in the literature (Liu et al., 2020b; Xu et al., 2022).

Apart from the graph-based techniques, Si et al. (2021) integrated topic models and neural networks for retrieving topic-constrained evidence information. Given a claim and set of evidence sentences, the authors applied Latent Dirichlet Allocation to extract the topics from the evidence sentences, and the topic distributions learned are used to obtain a topic representation of the evidence via a co-attention mechanism. This enabled the authors to incorporate topic consistency between the evidence and the topic consistency between the claim and evidence into dense representations. This topic-aware evidence representation and claim were input to a capsule network for determining the stance of the evidence towards the claim.

It can be observed that, compared to other NLU and NLG tasks, natural language reasoning demands multiple explicit and implicit knowledge sources integrated together for deriving new conclusions. This includes knowledge bases, graph embedding solutions, pre-trained language models, and topic modeling.

## 6. Challenges

While hybrid approaches eliminate weaknesses of symbolic methods and machine learning approaches, they pose certain challenges that would significantly impact practical usage. Following are some of the key challenges in the implementation of hybrid approaches.

- **Generalization of knowledge:** Although existing powerful models are infused with external knowledge sources for accurate inference, machine learning models tend to remember the knowledge provided during the learning and fail to update their internal memory according to the changes in the real world. This requires more generalized solutions without the need to retrain the model with changes in the real world and adapt to temporal changes for reduced model deterioration and improved persistence (Alkhalifa et al., 2023).

- **Generalization across tasks:** Hybrid models are often trained for a specific task with the integration of symbolic representation required to accomplish inference for the given task, making them incompatible across other NLP tasks. Research in model generalizability across tasks is still in its infancy, including in the development of models for zero-shot adaptation to new tasks.

- **Human-level reasoning:** Hybrid approaches are relatively powerful in natural language reasoning. However, human-level reasoning remains an open research problem (Yin et al., 2022), requiring more robust reasoning models simulating human thoughts. Especially, recent studies (Branco et al., 2021) have reported the inconsistent performance of hybrid models in reasoning tasks and observed that the models tend to remember the shortcuts present in the training data. Further, these models tend to rely more on reasoning on entities or the syntactic structure of the data. Hence, it is unclear whether the implicit commonsense knowledge of the hybrid models or the correlation in data results in superior performance in handling complex situations. This demands the development of more robust reasoning models encouraging the acquisition of commonsense knowledge rather than remembering training data.

- **Reliability of knowledge sources:** Another key concern of hybrid models is the reliability of external sources infused during the training which are often curated using automated tools and search engines. Curated data may contain biased and/or false information. Further, the knowledge related to a topic or context may vary over time, and utilizing outdated information may result in inaccurate inferences. This questions the reliability of the knowledge bases and the factual inferences obtained by hybrid models (Yin et al., 2022).

- **Increased computational requirements:** Integration of KBs with machine learning models increases the requirement of both computational resources and inference power to deal with a large number of logical rules and constraints. While approximate inference can be employed as a solution at a cost of reduced performance (Yu et al., 2023a), this does not serve the purpose of adopting hybrid solutions. Therefore, it is required to explore effective solutions utilizing the computational strength of neural models to tackle the increased computational requirements.

- **Requirement of customized KBs:** Even though most of the hybrid solutions employ existing general-purpose KBs, various studies Li et al. (2022); Meng et al. (2020); Huo et al. (2022) have shown the requirement of developing task-specific or domain-specific KBs and reported promising results when KBs are developed specifically for an NLP task or domain. This hinders the prompt adoption of hybrid solutions across other NLP tasks and domains.

- **Uniform knowledge acquisition and representation:** KBs support different retrieval systems and knowledge representation mechanisms depending on the information stored. Therefore, the existing approaches often rely on a single KB and utilize a KB-based technique to retrieve and represent the information. This restricts the utilization of multiple KBs for accomplishing an NLP task and demands more uniform approaches to retrieve and represent knowledge across various KBs (Zouhar et al., 2022).

## 7. Related Surveys

Zhu et al. (2023) briefly introduced the knowledge augmentation methods used for NLP during a recent tutorial. The study was centered around the integration of knowledge into language models for NLP. Apart from this study, various reviews have discussed the integration of knowledge graphs for NLP (Schneider et al., 2022), and integration of knowledge into large language models (LLMS) (Safavi and Koutra, 2021; Hu et al., 2023; Yin et al., 2022; Pan et al., 2023). Pan et al. (2023) presented the latest survey on this direction focusing on KG-enhanced LLMs, LLM-enhance KGs, and scenarios where LLM and KG can play equal roles. Focusing on specific NLP tasks, Yu et al. (2022b) detailed a survey on knowledge-enhanced text generation. From a more theoretical perspective, Ferrone and Zanzotto (2020) presented a survey summarizing the link between distributed and symbolic representations, and explained how symbols are represented in deep learning for NLP. Similarly, Zouhar et al. (2022) systematically described artifacts (items retrieved from the knowledge base) and techniques used in the literature to retrieve and inject the artifacts into NLP models.

Hamilton et al. (2022) examined the impact of neuro-symbolic approaches in NLP by analyzing the performance of the models in terms of five key criteria, out-of-distribution generalization, interpretability, reduced training data, transferability across domains, and reasoning. The authors concluded in their study, that there is no clear correlation is met between the integration of knowledge and the performance criteria analyzed.

Different from all these studies, we focus on the broader aspect of synergizing machine learning techniques and symbolic methods for NLP and present the state-of-the-art hybrid approaches proposed in the literature for a wide range of NLP tasks, providing a unique overview of this increasingly popular trend in NLP.

## 8. Future Directions

In the previous sections, we introduced the latest advancements in Hybrid NLP and the challenges linked to embracing this emerging line of research. Despite the challenges to be addressed, numerous open research problems persist in this promising research topic. We next put forth a set of immediate future research directions in hybrid NLP.

- **Introduction of pre-trained hybrid models:** Training hybrid models is computationally complex, and this hinders the accessibility of hybrid solutions for researchers with diverse levels of (typically limited) training resources. Similar to the effectiveness of pre-trained language models, a comparable approach can be embraced for hybrid models. Therefore, the introduction of pre-trained hybrid models has the potential to streamline computational efficiency, making them more accessible and appealing to a broader community of researchers.

- **Zero-short or few-shot reasoning using hybrid models:** Training hybrid models requires a massive amount of quality labeled data. However, certain domains face challenges due to limited labeled data. Therefore, zero-shot or few-short learning using hybrid models can facilitate learning with limited training data. While Large Language Models (LLMs) have shown impressive performance in zero-shot/few-shot learning, this learning trend is yet to be further studied for hybrid solutions.

- **Dynamic Reasoning:** Hybrid solutions play a key role in Natural Language Reasoning (NLR) by facilitating the generation of more meaningful and interpretable conclusions. Nevertheless, the existing hybrid solutions for NLR primarily focus on static symbolic representations. However, the information in symbolic representations may get outdated, e.g. a tuple (Donald Trump, President of, America) in a knowledge graph is no longer valid. Addressing this issue requires both the symbolic representation and reasoning to dynamically get updated. This aspect of the research remains unexplored and needs further investigation.

- **Automatic construction of symbolic knowledge:** While the present hybrid solutions highly benefit from the existing symbolic knowledge sources, there is a growing need for knowledge sources with various characteristics, e.g. domain-specific knowledge sources. Therefore, it will be an interesting research direction to explore the possibility of automatically generating symbolic knowledge through hybrid solutions (Yu et al., 2023a).

- **Handling multimodal data**: The present focus on hybrid techniques involves the utilization of textual knowledge in knowledge bases to develop effective NLP solutions. However, much of today's real-world data is predominantly multimodal, and this necessitates the introduction of multimodal hybrid techniques to effectively harness the knowledge across different modalities. The existence of multimodal knowledge bases and the latest developments in multimodal LLMs may expedite the adoption of multimodal hybrid techniques. However, this research direction requires further investigation and advancements.

## 9. Conclusion

The use of a hybrid approach is a promising direction of combining rich knowledge in symbolic methods with machine learning to enhance their inference capabilities. Moreover, it facilitates the generation of more credible and factually grounded inferences by incorporating external knowledge and commonsense, ultimately improving the reliability and adaptability of hybrid solutions across a broad spectrum of NLP tasks. Our survey paper is the first to provide a broad overview of hybrid approaches presented in three subareas of NLP including understanding (NLU), generation (NLG), and reasoning (NLR). We further delve into a set of tasks in each of these subareas, discussing state-of-the-art methods and progress made in the scientific community. We conclude the overview by discussing open research challenges for further development of hybrid approaches to NLP.

Hybrid approaches for NLP generally exploit two types of techniques, integration of symbolic knowledge into the input of statistical or deep learning models and modifying the deep learning models with symbolic structures resulting in architectures such as graph neural networks and hybrid data representation such as graph embedding. The first approach enables a wide range of adoption of symbolic knowledge sources such as external databases, knowledge graphs, and topic models and encourages the model to utilize this knowledge as additional information during inference to make accurate decisions. Similarly, hybrid architectures empower powerful representations of data and their relationships via hybrid structures, improving the scalability and explainability of the task. There is no doubt, that a clear understanding of both machine learning approaches and symbolic methods will lead to innovative architectures with substantial advancements in the future.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This project is funded by the European Union and UK Research and Innovation under Grant No. 101073351 as part of Marie Skłodowska-Curie Actions (MSCA Hybrid Intelligence to monitor, promote, and analyze transformations in good democracy practices).

## References

Abels, P., Ahmadi, Z., Burkhardt, S., Schiller, B., Gurevych, I., Kramer, S., 2021. Focusing knowledge-based graph argument mining via topic modeling. arXiv preprint arXiv:2102.02086 .

Alkhalifa, R., Kochkina, E., Zubiaga, A., 2023. Building for tomorrow: Assessing the temporal persistence of text classifiers. Information Processing & Management 60, 103200.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z., 2007. Dbpedia: A nucleus for a web of open data, in: international semantic web conference, Springer. pp. 722–735.

Baker, C.F., Fillmore, C.J., Cronin, B., 2003. The structure of the framenet database. International Journal of Lexicography 16, 281–296.

Bastings, J., Titov, I., Aziz, W., Marcheggiani, D., Sima'an, K., 2017. Graph convolutional encoders for syntax-aware neural machine translation, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 1957–1967. URL: https://aclanthology.org/D17-1209, doi:10.18653/v1/D17-1209.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research 3, 993–1022.

Bodenreider, O., 2004. The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research 32, D267–D270.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J., 2008. Freebase: a collaboratively created graph database for structuring human knowledge, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 1247–1250.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O., 2013. Translating embeddings for modeling multi-relational data. Advances in neural information processing systems 26.

Boros, E., González-Gallardo, C.E., Giamphy, E., Hamdi, A., Moreno, J.G., Doucet, A., 2022. Knowledge-based contexts for historical named entity recognition & linking. Experimental IR Meets Multilinguality, Multimodality, and Interaction .

Branco, R., Branco, A., Silva, J.M., Rodrigues, J., 2021. Commonsense reasoning: how do neuro-symbolic and neuro-only approaches compare?, in: CIKM Workshops.

Cai, D., Lam, W., 2020. Graph transformer for graph-to-sequence learning, in: Proceedings of the AAAI conference on artificial intelligence, pp. 7464–7471.

Cao, Z., Wei, F., Li, W., Li, S., 2018. Faithful to the original: Fact aware neural abstractive summarization, in: Proceedings of the AAAI Conference on Artificial Intelligence.

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E., Mitchell, T., 2010. Toward an architecture for never-ending language learning, in: Proceedings of the AAAI conference on artificial intelligence, pp. 1306–1313.

Chen, K., Wang, R., Utiyama, M., Sumita, E., Zhao, T., 2018. Syntax-directed attention for neural machine translation, in: Proceedings of the AAAI conference on artificial intelligence.

Chen, Y., Wu, L., Zaki, M.J., 2019. Reinforcement learning based graph-to-sequence model for natural question generation. arXiv preprint arXiv:1908.04942 .

Chen, Y., Wu, L., Zaki, M.J., 2023. Toward subgraph-guided knowledge graph question generation with graph neural networks. IEEE Transactions on Neural Networks and Learning Systems .

Ciampaglia, G.L., Shiralkar, P., Rocha, L.M., Bollen, J., Menczer, F., Flammini, A., 2015. Computational fact checking from knowledge networks. PloS one 10, e0128193.

Dale, R., 2000. Symbolic approaches to natural language processing. Handbook of Natural Language Processing , 1–9.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding, in: Burstein, J., Doran, C., Solorio, T. (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 4171–4186. URL: https://aclanthology.org/N19-1423, doi:10.18653/v1/N19-1423.

Dinan, E., Roller, S., Shuster, K., Fan, A., Auli, M., Weston, J., 2018. Wizard of wikipedia: Knowledge-powered conversational agents. arXiv preprint arXiv:1811.01241 .

Du, X., Cardie, C., 2018. Harvesting paragraph-level question-answer pairs from Wikipedia, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia. pp. 1907–1917. URL: https://aclanthology.org/P18-1177, doi:10.18653/v1/P18-1177.

El-Kassas, W.S., Salama, C.R., Rafea, A.A., Mohamed, H.K., 2021. Automatic text summarization: A comprehensive survey. Expert systems with applications 165, 113679.

El Mekki, A., El Mahdaouy, A., Berrada, I., Khoumsi, A., 2022. Adasl: an unsupervised domain adaptation framework for arabic multi-dialectal sequence labeling. Information Processing & Management 59, 102964.

Elsahar, H., Gravier, C., Laforest, F., 2018. Zero-shot question generation from knowledge graphs for unseen predicates and entity types, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana. pp. 218–228. URL: https://aclanthology.org/N18-1020, doi:10.18653/v1/N18-1020.

Etzioni, O., Banko, M., Soderland, S., Weld, D.S., 2008. Open information extraction from the web. Communications of the ACM 51, 68–74.

Fellbaum, C., 2010. Wordnet, in: Theory and applications of ontology: computer applications. Springer, pp. 231–243.

Feng, Y., Chen, X., Lin, B.Y., Wang, P., Yan, J., Ren, X., 2020. Scalable multi-hop relational reasoning for knowledge-aware question answering, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online. pp. 1295–1309. URL: https://aclanthology.org/2020.emnlp-main.99, doi:10.18653/v1/2020.emnlp-main.99.

Ferrada, S., Bustos, B., Hogan, A., 2017. Imgpedia: a linked dataset with content-based analysis of wikimedia images, in: The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II 16, Springer. pp. 84–93.

Ferrone, L., Zanzotto, F.M., 2020. Symbolic, distributed, and distributional representations for natural language processing in the era of deep learning: A survey. Frontiers in Robotics and AI 6, 153.

Fromm, M., Faerman, E., Seidl, T., 2019. Tacam: topic and context aware argument mining, in: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 99–106.

Ganea, O.E., Hofmann, T., 2017. Deep joint entity disambiguation with local neural attention, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 2619–2629. URL: https://aclanthology.org/D17-1277, doi:10.18653/v1/D17-1277.

Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N.F., Peters, M., Schmitz, M., Zettlemoyer, L.S., 2017. Allennlp: A deep semantic natural language processing platform. arXiv:arXiv:1803.07640.

Ghazvininejad, M., Brockett, C., Chang, M.W., Dolan, B., Gao, J., Yih, W.t., Galley, M., 2018. A knowledge-grounded neural conversation model, in: Proceedings of the AAAI Conference on Artificial Intelligence.

Grover, A., Leskovec, J., 2016. node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 855–864.

Gu, Y., Wang, Y., Zhang, H.R., Wu, J., Gu, X., 2023. Enhancing text classification by graph neural networks with multi-granular topic-aware graph. IEEE Access 11, 20169–20183.

Guan, J., Huang, F., Zhao, Z., Zhu, X., Huang, M., 2020. A knowledge-enhanced pretraining model for commonsense story generation. Transactions of the Association for Computational Linguistics 8, 93–108.

Gui, T., Zou, Y., Zhang, Q., Peng, M., Fu, J., Wei, Z., Huang, X.J., 2019. A lexicon-based graph neural network for chinese ner, in: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp. 1040–1050.

Guo, Z., Schlichtkrull, M., Vlachos, A., 2022. A survey on automated fact-checking. Transactions of the Association for Computational Linguistics 10, 178–206.

Hamilton, K., Nayak, A., Božić, B., Longo, L., 2022. Is neuro-symbolic ai meeting its promises in natural language processing? a structured review. Semantic Web , 1–42.

Han, X., Cao, S., Lv, X., Lin, Y., Liu, Z., Sun, M., Li, J., 2018. OpenKE: An open toolkit for knowledge embedding, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Brussels, Belgium. pp. 139–144. URL: https://aclanthology.org/D18-2024, doi:10.18653/v1/D18-2024.

Hoehndorf, R., Queralt-Rosinach, N., et al., 2017. Data science and symbolic ai: Synergies, challenges and opportunities. Data Science 1, 27–38.

Hou, S., Zhang, S., Fei, C., 2020. Rhetorical structure theory: A comprehensive review of theory, parsing methods and applications. Expert Systems with Applications 157, 113421.

Hu, L., Liu, Z., Zhao, Z., Hou, L., Nie, L., Li, J., 2023. A survey of knowledge enhanced pre-trained language models. IEEE Transactions on Knowledge and Data Engineering .

Hu, W., He, L., Ma, H., Wang, K., Xiao, J., 2022. Kgner: Improving chinese named entity recognition by bert infused with the knowledge graph. Applied Sciences 12, 7702.

Huo, C., Ma, S., Liu, X., 2022. Hotness prediction of scientific topics based on a bibliographic knowledge graph. Information Processing & Management 59, 102980. URL: https://www.sciencedirect.com/science/article/pii/S0306457322000966, doi:https://doi.org/10.1016/j.ipm.2022.102980.

Hwang, J.D., Bhagavatula, C., Le Bras, R., Da, J., Sakaguchi, K., Bosselut, A., Choi, Y., 2021. (comet-) atomic 2020: on symbolic and neural commonsense knowledge graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 6384–6392.

Ji, H., Ke, P., Huang, S., Wei, F., Zhu, X., Huang, M., 2020. Language generation with multi-hop reasoning on commonsense knowledge graph, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online. pp. 725–736. URL: https://aclanthology.org/2020.emnlp-main.54, doi:10.18653/v1/2020.emnlp-main.54.

Jia, Y., Wang, Y., Lin, H., Jin, X., Cheng, X., 2016. Locally adaptive translation for knowledge graph embedding, in: Proceedings of the AAAI Conference on Artificial Intelligence.

Kalyan, K.S., Rajasekharan, A., Sangeetha, S., 2021. Ammus: A survey of transformer-based pretrained models in natural language processing. arXiv preprint arXiv:2108.05542 .

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., Yih, W.t., 2020. Dense passage retrieval for open-domain question answering, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online. pp. 6769–6781. URL: https://aclanthology.org/2020.emnlp-main.550, doi:10.18653/v1/2020.emnlp-main.550.

Kazemi, A., Garimella, K., Gaffney, D., Hale, S., 2021. Claim matching beyond english to scale global fact-checking, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 4504–4517.

Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., Hajishirzi, H., 2019. Text Generation from Knowledge Graphs with Graph Transformers, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 2284–2293. URL: https://aclanthology.org/N19-1238, doi:10.18653/v1/N19-1238.

Kumar, V., Hua, Y., Ramakrishnan, G., Qi, G., Gao, L., Li, Y.F., 2019. Difficulty-controllable multi-hop question generation from knowledge graphs, in: The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18, Springer. pp. 382–398.

Kurdi, G., Leo, J., Parsia, B., Sattler, U., Al-Emari, S., 2020. A systematic review of automatic question generation for educational purposes. International Journal of Artificial Intelligence in Education 30, 121–204.

Lawrence, J., Reed, C., 2019. Argument mining: A survey. Computational Linguistics 45, 765–818. URL: https://aclanthology.org/J19-4006, doi:10.1162/coli_a_00364.

Levy, O., Seo, M., Choi, E., Zettlemoyer, L., 2017. Zero-shot relation extraction via reading comprehension, in: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), pp. 333–342.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 7871–7880. URL: https://aclanthology.org/2020.acl-main.703, doi:10.18653/v1/2020.acl-main.703.

Li, H., Zhu, J., Zhang, J., Zong, C., 2018. Ensure the correctness of the summary: Incorporate entailment knowledge into abstractive sentence summarization, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA. pp. 1430–1441. URL: https://aclanthology.org/C18-1121.

Li, T., Chen, X., Dong, Z., Keutzer, K., Zhang, S., 2022. Domain-adaptive text classification with structured knowledge from unlabeled data, IJCAI International Joint Conference on Artificial Intelligence.

Li, Z., Ding, X., Liu, T., Hu, J.E., Van Durme, B., 2021. Guided generation of cause and effect. arXiv preprint arXiv:2107.09846 .

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P., 2020a. K-bert: Enabling language representation with knowledge graph, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2901–2908.

Liu, Y., Li, H., Garcia-Duran, A., Niepert, M., Onoro-Rubio, D., Rosenblum, D.S., 2019a. Mmkg: multi-modal knowledge graphs, in: The Semantic Web: 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2–6, 2019, Proceedings 16, Springer. pp. 459–474.

Liu, Y., Li, P., Hu, X., 2022. Combining context-relevant features with multi-stage attention network for short text classification. Computer Speech & Language 71, 101268.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019b. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 .

Liu, Y., Wan, Y., He, L., Peng, H., Philip, S.Y., 2021. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 6418–6425.

Liu, Y., Zeng, Q., Ordieres Meré, J., Yang, H., 2019c. Anticipating stock market of the renowned companies: A knowledge graph approach. Complexity 2019.

Liu, Z., Niu, Z.Y., Wu, H., Wang, H., 2019d. Knowledge aware conversation generation with explainable reasoning over augmented graphs, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 1782–1792.

Liu, Z., Xiong, C., Sun, M., Liu, Z., 2020b. Fine-grained fact verification with kernel graph attention network, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 7342–7351. URL: https://aclanthology.org/2020.acl-main.655, doi:10.18653/v1/2020.acl-main.655.

Lopez, A., 2008. Statistical machine translation. ACM Computing Surveys (CSUR) 40, 1–49.

Lv, S., Guo, D., Xu, J., Tang, D., Duan, N., Gong, M., Shou, L., Jiang, D., Cao, G., Hu, S., 2020. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering, in: Proceedings of the AAAI conference on artificial intelligence, pp. 8449–8456.

Meng, C., Ren, P., Chen, Z., Monz, C., Ma, J., de Rijke, M., 2020. Refnet: A reference-aware network for background based conversation, in: Proceedings of the AAAI conference on artificial intelligence, pp. 8496–8503.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 .

Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I., Roth, D., 2023. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Computing Surveys 56, 1–40.

Mitra, A., Banerjee, P., Pal, K.K., Mishra, S., Baral, C., 2019. How additional knowledge can improve natural language commonsense question answering? arXiv preprint arXiv:1909.08855 .

Mulla, N., Gharpure, P., 2023. Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. Progress in Artificial Intelligence 12, 1–32.

Narayan, S., Cohen, S.B., Lapata, M., 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium. pp. 1797–1807. URL: https://aclanthology.org/D18-1206, doi:10.18653/v1/D18-1206.

Ni, J., Young, T., Pandelea, V., Xue, F., Cambria, E., 2023. Recent advances in deep learning based dialogue systems: A systematic survey. Artificial intelligence review 56, 3055–3155.

Noraset, T., Lowphansirikul, L., Tuarob, S., 2021. Wabiqa: A wikipedia-based thai question-answering system. Information Processing & Management 58, 102431. URL: https://www.sciencedirect.com/science/article/pii/S0306457320309249, doi:https://doi.org/10.1016/j.ipm.2020.102431.

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., Wu, X., 2023. Unifying large language models and knowledge graphs: A roadmap. arXiv preprint arXiv:2306.08302 .

Pérez-Agüera, J.R., Arroyo, J., Greenberg, J., Iglesias, J.P., Fresno, V., 2010. Using bm25f for semantic search, in: Proceedings of the 3rd international semantic search workshop, pp. 1–8.

Peters, M.E., Neumann, M., Logan, R., Schwartz, R., Joshi, V., Singh, S., Smith, N.A., 2019. Knowledge enhanced contextual word representations, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China. pp. 43–54. URL: https://aclanthology.org/D19-1005, doi:10.18653/v1/D19-1005.

Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., Miller, A., 2019. Language models as knowledge bases?, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China. pp. 2463–2473. URL: https://aclanthology.org/D19-1250, doi:10.18653/v1/D19-1250.

Pittaras, N., Giannakopoulos, G., Stamatopoulos, P., Karkaletsis, V., 2023. Content-based and knowledge-enriched representations for classification across modalities: a survey. ACM Computing Surveys .

Radford, A., Narasimhan, K., 2018. Improving language understanding by generative pre-training. URL: https://api.semanticscholar.org/CorpusID:49313245.

Reddy, S., Raghu, D., Khapra, M.M., Joshi, S., 2017. Generating natural language question-answer pairs from a knowledge graph using a RNN based question generation model, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Association for Computational Linguistics, Valencia, Spain. pp. 376–385. URL: https://aclanthology.org/E17-1036.

Reimers, N., Gurevych, I., 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 .

Ribeiro, L.F., Saverese, P.H., Figueiredo, D.R., 2017. struc2vec: Learning node representations from structural identity, in: Proceedings of the 23rd

ACM SIGKDD international conference on knowledge discovery and data mining, pp. 385–394.

Rosenfeld, R., 2000. Two decades of statistical language modeling: Where do we go from here? Proceedings of the IEEE 88, 1270–1278.

Saadat-Yazdi, A., Li, X., Chausson, S., Belle, V., Ross, B., Pan, J.Z., Kökciyan, N., 2022. KEViN: A knowledge enhanced validity and novelty classifier for arguments, in: Proceedings of the 9th Workshop on Argument Mining, International Conference on Computational Linguistics, Online and in Gyeongju, Republic of Korea. pp. 104–110. URL: https://aclanthology.org/2022.argmining-1.9.

Saadat-Yazdi, A., Pan, J.Z., Kokciyan, N., 2023. Uncovering implicit inferences for improved relational argument mining, in: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Dubrovnik, Croatia. pp. 2484–2495. URL: https://aclanthology.org/2023.eacl-main.182.

Saedi, C., Branco, A., António Rodrigues, J., Silva, J., 2018. WordNet embeddings, in: Proceedings of the Third Workshop on Representation Learning for NLP, Association for Computational Linguistics, Melbourne, Australia. pp. 122–131. URL: https://aclanthology.org/W18-3016, doi:10.18653/v1/W18-3016.

Safavi, T., Koutra, D., 2021. Relational world knowledge representation in contextual language models: A review, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 1053–1067.

Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N.A., Choi, Y., 2019. Atomic: An atlas of machine commonsense for if-then reasoning, in: Proceedings of the AAAI conference on artificial intelligence, pp. 3027–3035.

Sarker, M.K., Zhou, L., Eberhart, A., Hitzler, P., 2021. Neuro-symbolic artificial intelligence. AI Communications 34, 197–209.

Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M., 2018. Modeling relational data with graph convolutional networks, in:

The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15, Springer. pp. 593–607.

Schneider, P., Schopf, T., Vladika, J., Galkin, M., Simperl, E., Matthes, F., 2022. A decade of knowledge graphs in natural language processing: A survey, in: Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online only. pp. 601–614. URL: https://aclanthology.org/2022.aacl-main.46.

Sebastiani, F., 2002. Machine learning in automated text categorization. ACM computing surveys (CSUR) 34, 1–47.

Sennrich, R., Haddow, B., 2016. Linguistic input features improve neural machine translation, in: Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers, Association for Computational Linguistics, Berlin, Germany. pp. 83–91. URL: https://aclanthology.org/W16-2209, doi:10.18653/v1/W16-2209.

Shi, B., Weninger, T., 2016. Fact checking in heterogeneous information networks, in: Proceedings of the 25th International Conference Companion on World Wide Web, pp. 101–102.

Shiralkar, P., Flammini, A., Menczer, F., Ciampaglia, G.L., 2017. Finding streams in knowledge graphs to support fact checking, in: 2017 IEEE International Conference on Data Mining (ICDM), IEEE. pp. 859–864.

Si, J., Zhou, D., Li, T., Shi, X., He, Y., 2021. Topic-aware evidence reasoning and stance-aware aggregation for fact verification, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online. pp. 1612–1622. URL: https://aclanthology.org/2021.acl-long.128, doi:10.18653/v1/2021.acl-long.128.

Škrlj, B., Martinc, M., Kralj, J., Lavrač, N., Pollak, S., 2021. tax2vec: Constructing interpretable features from taxonomies for short text classification. Computer Speech & Language 65, 101104.

Speer, R., Chin, J., Havasi, C., 2017. Conceptnet 5.5: An open multilingual graph of general knowledge, in: Proceedings of the AAAI conference on artificial intelligence.

Stasaski, K., Hearst, M.A., 2017. Multiple choice question generation utilizing an ontology, in: Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, Copenhagen, Denmark. pp. 303–312. URL: https://aclanthology.org/W17-5034, doi:10.18653/v1/W17-5034.

Suchanek, F.M., Kasneci, G., Weikum, G., 2007. Yago: a core of semantic knowledge, in: Proceedings of the 16th international conference on World Wide Web, pp. 697–706.

Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J., 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv preprint arXiv:1902.10197 .

Tedeschi, S., Maiorca, V., Campolungo, N., Cecconi, F., Navigli, R., 2021. Wikineural: Combined neural and knowledge-based silver data creation for multilingual ner, in: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 2521–2533.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al., 2017. Graph attention networks. stat 1050, 10–48550.

Vrandečić, D., Krötzsch, M., 2014. Wikidata: a free collaborative knowledgebase. Communications of the ACM 57, 78–85.

Wang, D., Liu, P., Zheng, Y., Qiu, X., Huang, X., 2020a. Heterogeneous graph neural networks for extractive document summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 6209–6219. URL: https://aclanthology.org/2020.acl-main.553, doi:10.18653/v1/2020.acl-main.553.

Wang, G.H., Li, W., Lai, E.M.K., Jiang, J., 2022a. Katsum: Knowledge-aware abstractive text summarization. ArXiv abs/2212.03371. URL: https://api.semanticscholar.org/CorpusID:254366408.

Wang, M., Wang, H., Qi, G., Zheng, Q., 2020b. Richpedia: a large-scale, comprehensive multi-modal knowledge graph. Big Data Research 22, 100159.

Wang, X., Gao, T., Zhu, Z., Zhang, Z., Liu, Z., Li, J., Tang, J., 2021. Kepler: A unified model for knowledge embedding and pre-trained language representation. Transactions of the Association for Computational Linguistics 9, 176–194.

Wang, X., Shen, Y., Cai, J., Wang, T., Wang, X., Xie, P., Huang, F., Lu, W., Zhuang, Y., Tu, K., Lu, W., Jiang, Y., 2022b. DAMO-NLP at SemEval-2022 task 11: A knowledge-based system for multilingual named entity recognition, in: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), Association for Computational Linguistics, Seattle, United States. pp. 1457–1468. URL: https://aclanthology.org/2022.semeval-1.200, doi:10.18653/v1/2022.semeval-1.200.

Wang, Y., Sun, Y., Ma, Z., Gao, L., Xu, Y., 2020c. An ernie-based joint model for chinese named entity recognition. Applied Sciences 10, 5711.

Wang, Z., Huang, J., Li, H., Liu, B., Shao, B., Wang, H., Wang, J., Wang, Y., Wu, W., Xiao, J., et al., 2010. Probase: a universal knowledge base for semantic search. Microsoft Research Asia .

Wu, W., Li, H., Wang, H., Zhu, K.Q., 2012. Probase: A probabilistic taxonomy for text understanding, in: Proceedings of the 2012 ACM SIGMOD international conference on management of data, pp. 481–492.

Xiong, W., Du, J., Wang, W.Y., Stoyanov, V., 2019. Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model. arXiv preprint arXiv:1912.09637 .

Xu, J., Gan, Z., Cheng, Y., Liu, J., 2020. Discourse-aware neural extractive text summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 5021–5031. URL: https://aclanthology.org/2020.acl-main.451, doi:10.18653/v1/2020.acl-main.451.

Xu, W., Wu, J., Liu, Q., Wu, S., Wang, L., 2022. Evidence-aware fake news detection with graph neural networks, in: Proceedings of the ACM Web Conference 2022, pp. 2501–2510.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., Le, Q.V., 2019. Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems 32.

Yang, Z., Xu, Y., Hu, J., Dong, S., 2023. Generating knowledge aware explanation for natural language inference. Information Processing & Management 60, 103245. URL: https://www.sciencedirect.com/science/article/pii/S0306457322003466, doi:https://doi.org/10.1016/j.ipm.2022.103245.

Yao, L., Mao, C., Luo, Y., 2019. Graph convolutional networks for text classification, in: Proceedings of the AAAI conference on artificial intelligence, pp. 7370–7377.

Yasunaga, M., Ren, H., Bosselut, A., Liang, P., Leskovec, J., 2021. QA-GNN: Reasoning with language models and knowledge graphs for question answering, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online. pp. 535–546. URL: https://aclanthology.org/2021.naacl-main.45, doi:10.18653/v1/2021.naacl-main.45.

Yin, D., Dong, L., Cheng, H., Liu, X., Chang, K.W., Wei, F., Gao, J., 2022. A survey of knowledge-intensive nlp with pre-trained language models. arXiv preprint arXiv:2202.08772 .

Yu, D., Yang, B., Liu, D., Wang, H., Pan, S., 2023a. A survey on neural-symbolic learning systems. Neural Networks 166, 105–126. URL: https://www.sciencedirect.com/science/article/pii/S0893608023003398, doi:https://doi.org/10.1016/j.neunet.2023.06.028.

Yu, D., Zhu, C., Yang, Y., Zeng, M., 2022a. Jaket: Joint pre-training of knowledge graph and language understanding, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 11630–11638.

Yu, F., Zhang, H., Wang, B., 2023b. Nature language reasoning, a survey. arXiv preprint arXiv:2303.14725 .

Yu, W., Zhu, C., Li, Z., Hu, Z., Wang, Q., Ji, H., Jiang, M., 2022b. A survey of knowledge-enhanced text generation. ACM Computing Surveys 54, 1–38.

Zaib, M., Zhang, W.E., Sheng, Q.Z., Mahmood, A., Zhang, Y., 2022. Conversational question answering: A survey. Knowledge and Information Systems 64, 3151–3195.

Zeng, X., Abumansour, A.S., Zubiaga, A., 2021. Automated fact-checking: A survey. Language and Linguistics Compass 15, e12438.

Zhang, H., Khashabi, D., Song, Y., Roth, D., 2020a. Transomcs: From linguistic graphs to commonsense knowledge. arXiv preprint arXiv:2005.00206 .

Zhang, H., Liu, X., Pan, H., Song, Y., Leung, C.W.K., 2020b. Aser: A large-scale eventuality knowledge graph, in: Proceedings of the web conference 2020, pp. 201–211.

Zhang, H., Liu, Z., Xiong, C., Liu, Z., 2020c. Grounded conversation generation as guided traverses in commonsense knowledge graphs, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 2031–2043. URL: https://aclanthology.org/2020.acl-main.184, doi:10.18653/v1/2020.acl-main.184.

Zhang, Q., Chen, S., Fang, M., Chen, X., 2023a. Joint reasoning with knowledge subgraphs for multiple choice question answering. Information Processing & Management 60, 103297. URL: https://www.sciencedirect.com/science/article/pii/S0306457323000341, doi:https://doi.org/10.1016/j.ipm.2023.103297.

Zhang, S., Tay, Y., Yao, L., Liu, Q., 2019a. Quaternion knowledge graph embeddings. Advances in neural information processing systems 32.

Zhang, X., Bosselut, A., Yasunaga, M., Ren, H., Liang, P., Manning, C.D., Leskovec, J., 2022. Greaselm: Graph reasoning enhanced language models for question answering. arXiv preprint arXiv:2201.08860 .

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., et al., 2023b. Siren's song in the ai ocean: A survey on hallucination in large language models. arXiv preprint arXiv:2309.01219 .

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q., 2019b. ERNIE: Enhanced language representation with informative entities, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy. pp. 1441–1451. URL: https://aclanthology.org/P19-1139, doi:10.18653/v1/P19-1139.

Zhong, W., Xu, J., Tang, D., Xu, Z., Duan, N., Zhou, M., Wang, J., Yin, J., 2020. Reasoning over semantic-level graph for fact checking, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online. pp. 6170–6180. URL: https://aclanthology.org/2020.acl-main.549, doi:10.18653/v1/2020.acl-main.549.

Zhou, H., Young, T., Huang, M., Zhao, H., Xu, J., Zhu, X., 2018. Commonsense knowledge aware conversation generation with graph attention., in: IJCAI, pp. 4623–4629.

Zhou, J., Han, X., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2019. GEAR: Graph-based evidence aggregating and reasoning for fact verification, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy. pp. 892–901. URL: https://aclanthology.org/P19-1085, doi:10.18653/v1/P19-1085.

Zhou, W., Lee, D.H., Selvam, R.K., Lee, S., Lin, B.Y., Ren, X., 2020. Pre-training text-to-text transformers for concept-centric common sense. arXiv preprint arXiv:2011.07956 .

Zhu, C., Xu, Y., Ren, X., Lin, B.Y., Jiang, M., Yu, W., 2023. Knowledge-augmented methods for natural language processing, in: Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, pp. 1228–1231.

Zhu, Y., Zhou, W., Xu, Y., Liu, J., Tan, Y., et al., 2017. Intelligent learning for knowledge graph towards geological data. Scientific Programming 2017.

Zouhar, V., Mosbach, M., Biswas, D., Klakow, D., 2022. Artefact retrieval: Overview of nlp models with knowledge base access. arXiv preprint arXiv:2201.09651 .