Entity-Augmented Neuroscience Knowledge Retrieval Using Ontology and Semantic **Understanding Capability of LLM**

Pralaypati Ta, Sriram Venkatesaperumal, Keerthi Ram, and Mohanasankar Sivaprakasam

Abstract— Neuroscience research publications encompass a vast wealth of knowledge. Accurately retrieving existing information and discovering new insights from this extensive literature is essential for advancing the field. However, when knowledge is dispersed across multiple sources, current state-of-the-art retrieval methods often struggle to extract the necessary information. A knowledge graph (KG) can integrate and link knowledge from multiple sources, but existing methods for constructing KGs in neuroscience often rely on labeled data and require domain expertise. Acquiring large-scale, labeled data for a specialized area like neuroscience presents significant challenges. This work proposes novel methods for constructing KG from unlabeled large-scale neuroscience research corpus utilizing large language models (LLM), neuroscience ontology, and text embeddings. We analyze the semantic relevance of neuroscience text segments identified by LLM for building the knowledge graph. We also introduce an entity-augmented information retrieval algorithm to extract knowledge from the KG. Several experiments were conducted to evaluate the proposed approaches, and the results demonstrate that our methods significantly enhance knowledge discovery from the unlabeled neuroscience research corpus. It achieves an F1 score of 0.84 for entity extraction, and the knowledge obtained from the KG improves answers to over 54% of the questions.

Index Terms— Entity extraction, Knowledge discovery, Knowledge graph, Large language model, Ontology

I. INTRODUCTION

THE volume of published research articles in neuroscience has consistently increased over time [1] [2]. This extensive literature contains a wealth of knowledge in the form of new findings, claims, counterclaims, and various scholarly discussions. However, the unstructured nature of this data makes extracting valuable insights from these articles complex and time-consuming [3]. More importantly, the mutual implications of different published research works remain largely unexplored, as they have not been effectively linked [4]. Establishing such links is especially crucial for neuroscience because fully depicting the findings on a given neuroscience topic would include uncovering the relationships between the various research streams that examine multiple facets of that topic [4] [5]. Hence, there is a need for an automated and practical solution to extract information from research articles and facilitate navigation between interrelated findings across related studies. Such a system would significantly enhance the knowledge discovery process [3] [5].

Dense passage retrieval has emerged as a leading approach for extracting semantically similar knowledge from vast collections of textual data [6]. These methods utilize neural networks, such as BERT [7] or other transformer architectures, to represent questions and documents as dense vector embeddings within a common vector space. Relevance is then determined by calculating similarity scores between the question and document vectors. However, dense retrieval systems often encounter challenges in accurately retrieving all necessary passages to answer complex questions, particularly when the relevant information is dispersed across multiple documents that exhibit limited lexical overlap with the original question [8] [9].

The knowledge graph (KG) is an effective data structure for integrating and linking knowledge from multiple sources, enhancing accurate information retrieval and enabling automated knowledge discovery. In knowledge graphs, relationships are stored explicitly, while in vector embeddings, they are represented implicitly through proximity in the embedding space. KGs address the challenges associated with dense passage retrieval in accurately retrieving knowledge from multiple sources by connecting the research articles through shared entities [9].

The primary steps in converting unstructured content from neuroscience literature into a knowledge graph (KG) involve extracting establishing relationships neuroscience entities. Various approaches for extracting entities from unstructured text have been proposed, including rule-based, dictionary-based, and deep learning methods [15]. Recent studies have demonstrated that deep learning-based methods consistently outperform other techniques in neuroscience entity recognition tasks [12] [13] [14]. However, the deep-learning approaches require substantial amounts of

Pralaypati Ta, Sriram Venkatesaperumal, Keerthi Ram and Mohanasankar Siyaprakasam are with the Sudha Gopalakrishnan Brain Centre, Indian Institute of Technology, Madras, Chennai, India (e-mail: pralay@gmail.com).

Pralaypati Ta and Mohanasankar Sivaprakasam are also with the Department of Electrical Engineering, Indian Institute of Technology, Madras, Chennai, India.

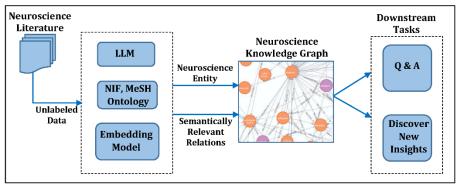


Fig. 1. Construction of neuroscience KG from unlabeled data and knowledge discovery using KG.

annotated text to train their models [15], which can be expensive and time-consuming, especially in specialized fields like neuroscience. Hence, there has been growing interest in exploring few-shot methods for entity and relation extraction tasks. Large Language Models (LLMs) have shown promising results in these tasks for the biomedical domain through zero-shot learning, few-shot learning, and minimal instruction tuning [16] [17] [21].

In this work, we extract entities and infer relationships between entities from unlabeled neuroscience research corpus using zero-shot and few-shot methods, as depicted in Fig. 1. We propose a novel approach for neuroscience entity extraction by combining entity linking and LLM-based methods. First, we formulate entity extraction as an entity linking (EL) problem [19] that jointly performs entity mention detection and disambiguation. It uses Neuroscience Information Framework Standardized Ontologies (NIFSTD) [18] and semantic similarity to perform entity linking. The entities generated by this method are combined with those extracted by LLM, and the combined set is finally filtered to remove false positives. We evaluate the performance of our approach against the performance of open-source Llama 3 [20] language models and other similarity-based unsupervised approaches.

Recent studies have harnessed the capabilities of large language models (LLMs) to extract relations among entities from unstructured scientific texts in biomedical and other domains [22] [23]. In the relation extraction (RE) task, given two specific entities and an input text, the LLM must determine whether a relationship exists between them in the text and extract that information. This task depends on the model's ability to recognize text segments relevant to the entities involved. We hypothesize that relation extraction is an entity-centric text span retrieval task. We propose an entityaugmented text extraction approach [24] to explore the language model's proficiency in identifying these text spans within unstructured neuroscience texts. This method extracts the appropriate text from a paragraph pertinent to a given entity. We formulate a dense retrieval task using neuroscience-specific questions from the PubMedQA [25] dataset to evaluate the relevance of the extracted data. Given a question, the task is to identify the corresponding context. Our evaluation shows that utilizing the entity-centric texts extracted from the PubMedQA contexts further improves the semantic retrieval performance of the state-of-the-art retrieval model NV-Embed-v2 [26].

Any effective question-answering system designed for a specialized domain like neuroscience must provide answers supported by research evidence. LLMs can answer domainspecific questions based on their parametric knowledge. However, when asked to provide proof for an answer, LLMs suffer from reference fabrication problems [27]. Our experiment indicates that these reference fabrication problems also exist in LLMs when applied to question-answering in neuroscience. The retrieval augmented generation (RAG) [28] helps mitigate LLM's hallucination and reference fabrication problem. However, the capability of RAG is limited when answers must be synthesized from multiple documents [29] [38]. Recent studies have proposed leveraging knowledge graphs to overcome this challenge, utilizing a relevant subgraph to provide context for the LLM in generating answers [29]. In this work, we construct a prototype knowledge graph from the 3248 neuroscience papers, 1764 full text & 1484 abstracts, employing the entity extraction and LLM-based relation extraction steps mentioned above (Fig. 1). We develop an entity-augmented subgraph retrieval module to extract pertinent subgraphs for specific questions. The efficacy of the extracted subgraph was assessed using an LLM as a judge [30] and 493 synthetic neuroscience questions. Our experiment shows that the knowledge graphbased retrieval system performs better in answering a significant percentage of questions than the baseline RAG method.

To summarize, our contributions are as follows:

- We present a novel approach for extracting neuroscience entities from unlabeled texts, combining entity linking and LLM-based extraction methods. Our approach achieves an F1 score of 0.84 and outperforms few-shot LLM-based and other techniques.
- We propose an entity-augmented text extraction approach to investigate LLMs' entity-specific text extraction capability from neuroscience literature. Our experiment demonstrates that the text extracted by LLM using the proposed approach has semantic significance. This method was subsequently employed

- to extract relationships between neuroscience entities.
- We show the reference fabrication issue prevalent in LLM's answers when addressing queries within the neuroscience domain.
- We construct a neuroscience knowledge graph based on the entity and relation extraction techniques mentioned above. We develop an entity-augmented information retrieval module to extract the relevant sub-graph from this knowledge graph. Experiments conducted on a synthetically generated neuroscience question dataset indicate that leveraging the knowledge graph allows for improved answers to various questions, mainly when the information is dispersed across multiple documents.

II. RELATED WORK

Extracting structured information from unstructured text is essential for knowledge discovery. Many research studies have focused on extracting structured data from unstructured scientific texts in the biomedical domain [10], including neuroscience [3] [11-14] [32], and also in other fields of science [22] [54]. The two most essential tasks in structured information extraction from the scientific literature are extracting domain-specific concepts, called entity, and their relations.

A. Entity and Relation Extraction

The existing research works for entity and relation extraction in the biomedical domain can be divided into the following categories [45]: rule-based, ontology, and dictionary-based, using machine learning and deep learning techniques, along with the use of large language models in more recent studies. In dictionary-based approaches, the Unified Medical Language System (UMLS) is most commonly used [10] [11] [12]. QuickUMLS [31], MaxMatcher [43], and DeepLife [40] use approximate string matching to match the unstructured text with the terms in UMLS to extract medical concepts. Ontology has been used for extracting structured information related to the treatment of spinal cord injury [32]. Conditional Random Field (CRF) is the most commonly used machine learning model for extracting entities from unstructured text [12].

In DL-based methods, many works use the Bidirectional Long Term Short Term Memory Networks (Bi-LSTM) model combined with CRF for entity extraction tasks [12] [39]. In BioBERT-CRF [13], the output of the BioBERT layer is used as input to the CRF layer to extract the brain region entities. It achieved an F1 score better than Bi-LSTM-CRF. BioSEPBERT [14] connects a position pointer module to the output of BERT to extract brain regions and connectivity relations among them. Word sense disambiguation (WSD) is a task related to entity recognition and relation extraction. Various approaches using the embedding vector of the words and the corresponding concepts from the knowledge base have been explored in [42] for biomedical WSD. A comparison between dictionary-based, CRF-based, and DL-based methods for entity recognition was

made in [12].

With the advent of LLMs, researchers are utilizing them for different types of NLP tasks in the biomedical domain [51]. BioNER-LLaMA [17] has fine-tuned LLAMA-7B models using instruction tuning for biomedical entity recognition. It has achieved performance comparable to the specialized PubMedBERT model. Zero-shot and few-shot learning approaches in LLMs have been explored for biomedical entity recognition and relation extraction [16] [21]. As shown in [16], the few-shot approach shows significant improvement but is still behind the specialized models such as PubMedBERT. In [47], in-context learning (ICL) in LLM is used to improve the accuracy of the information extracted from the clinical data. Using prompt engineering in GPT-4, [53] achieved an F1 score of 0.887 in biological entity recognition. In addition to the biomedical domain, LLMs are also used for information extraction from unstructured scientific texts of other domains [22] [54].

An entity may have many different surface forms [19]. Entity linking (EL) maps different non-standard surface forms of an entity to a standard label in the knowledge base [36]. The EL task includes two steps: surface form identification, referred to as mention detection (MD), and entity disambiguation (ED). The neural entity linking models use encoders to generate vector representations of entity mentions and candidate entities. Then, a ranking model is used to determine the matching score [56]. MedLinker [44] employs BiLSTM and CRF for embedding generation, utilizing cosine similarity to calculate the matching score. An end-to-end entity linking method is proposed in [55] to perform MD and ED jointly. It uses bidirectional LSTM to generate the embedding of entity mentions. In [46] and [48], the candidate entities are retrieved using an approximate nearest neighbor search. EL methods using n-gram are also proposed in the literature [52] [57].

Many studies have used entity-augmented or entity-focused retrieval to improve the performance in the information extraction task. CLEAR [24] utilizes entities from the input query to retrieve text from clinical notes, achieving superior performance compared to the baseline RAG. EnFoRe [34] identifies entities from the query and computes their importance score given the query and passages. These scores are then used to determine the query-passage relevancy. An entity-centric information retrieval approach is proposed in [37] in which the bridge entities connect the relevant paragraphs in multi-hop question answering.

Extracting the most relevant sentences or spans of text from the unstructured text is crucial in many NLP tasks. The Machine Reading Comprehension (MRC) task involves understanding unstructured text and utilizing the applicable span of tokens to answer questions [36]. The work in [33] combines term-based and neural semantic retrieval to extract a set of sentences to answer the query. In [50], the NER task is formulated as an MRC task that locates and extracts the relevant span. Clinical concepts and relation extraction as an MRC task have been explored in [58]. It uses BERT and LLM with prompt-based question answering to extract the pertinent text spans.

B. Biomedical Knowledge Graph

Domain-specific knowledge graph (KG) stores verified domain knowledge in a structured form. Researchers have created several knowledge graphs for the biomedical domain [39]. KGHC [10] is a knowledge graph for hepatocellular carcinoma created from structured and unstructured data. PrimeKG [41] is a multimodal knowledge graph constructed using 20 structured knowledge bases for precision medicine analysis. A KG for heart failure has been developed in [23] using LLM and prompt engineering. However, there are only a few works on knowledge graph construction in neuroscience. A knowledge graph development approach to discover knowledge from the Dementia research articles has been proposed in [3]. In the research work [11], a knowledge graph from unstructured research abstracts related to autism spectrum disorder has been constructed with minimum supervision. One primary use of KG is question answering. Various methods have been suggested to extract the relevant sub-graph from the KG to answer questions [29] [35]. KGAREVION [38] uses KG to verify and filter LLM-generated triplets containing the answer to the input query.

We identified the following research gaps in KG construction for the neuroscience domain. First, there is limited work on neuroscience entity extraction methods that work on unlabeled data and can accurately extract any neuroscience entity. Second, LLM's semantic understanding capability in neuroscience research texts has not been explored much. Lastly, to our knowledge, limited research exists on constructing knowledge graphs related to neuroscience.

III. METHODS

A. Extraction of Neuroscience Entities

Given an unstructured text document $D = \{w_1, w_2, \dots, w_n\}$, where D is a sequence of words w_i , and a vocabulary of entities $V = \{e_k\}_{k \in [1, |V|]}$ from the knowledge base, the entity extraction task, formulated as an entity-linking problem [19] [48], generates a list of $\{(w_{ij}, e_k)\}$ pairs [55], where w_{ij} is a sequence of words from D and $e_k \in V$.

We developed the vocabulary *V* of neuroscience entities by recursively traversing the NIFSTD [18] ontology tree, including the imported ontologies. We collected each class's labels, preferred and alternate labels, synonyms, and abbreviations whenever possible [12]. To further enrich the neuroscience vocabulary, we incorporated ontology classes from the "*Nervous System*" and "*Nervous System Diseases*" subtrees of the Medical Subject Headings (MeSH) ontology. The generated vocabulary underwent a review and cleaning process; some common English words were removed using the Python library NLTK. Finally, the vocabulary has ~400K neuroscience entity strings representing different surface forms of the entities.

The n-grams of a text document $D = \{w_1, w_2, \dots, w_n\}$ are the sequence of words $s_{ij} = \{w_i, \dots, w_j\} \ \forall i, j \text{ where} \ \{i \in \{i, \dots, i\}\}$

 $\{1,...,|D|-n+1\}$, j=(i+n-1) and $n\in\{1,...,\max(n)\}$. We generated all possible n-grams [52] [57] from a text document D for a specific maximum value of n and compared them against the entities in V using a similarity function [11] entsim to generate the list of $\{(w_{ij},e_k)\}$ pairs. The similarity function returns the following set:

$$E_{all} = \{ (w_{ij}, e_k) \mid entsim(w_{ij}, e_k) \ge \alpha \}$$
 (1)

where $\alpha \in [0, 1]$ is similarity threshold. The entities from set E_{all} that match the longest non-overlapping word sequences are ultimately considered as the entities present in document D.

We explored two similarity functions: 1) fuzzy matching utilizing Levenshtein Distance and 2) cosine similarity through the semantic retrieval model [19] [56]. The fuzzy matchingbased similarity check was implemented using the Python library the Fuzz. We employed two models for cosine similarity with the semantic retrieval model: NV-Embed-v2 [26] and a fine-tuned BAAI General Embedding (BGE) large model [59]. NV-Embed-v2 is an LLM-based embedding model, and it was one of the top-scored models in the embedding leaderboard [61] for retrieval tasks while we performed the experiments. For fine-tuning the BGE model, we generated two different strings for each entity name in the vocabulary, having a maximum of two edit distances from the actual names collected from the ontology [40]. We first conducted continual pre-training of the BGE model using all entity names, including actual names and their edit distance variants, to adapt the model specifically to neuroscience entity names. During fine-tuning, the modified names for an entity were treated as queries, while the corresponding actual name served as a positive sample for the retrieval task. Five negative samples were also hard-mined for each modified entity name.

We conducted experiments using three open-source LLama models: Llama-3.1-70B, Llama-3.3-70B, and Llama-3.1-405B, focusing on extracting neuroscience entities from unstructured research articles. We employed a one-shot prompt for this task, and the models generated outputs based on their parametric knowledge. Given an unstructured document *D*, the output of entity extraction task using LLMs can be formulated as follows:

$$f_{ex}(D) = \left\{ t \mid t \in D_{span}, e \in \mathcal{E}, P(t, e \mid D) \ge \tau \right\}$$
 (2)

where D_{span} denotes the text spans in document D, \mathcal{E} represents the set of neuroscience entities present in the parametric knowledge of the model, $P(t,e\mid D)$ is the joint probability of span t being a neuroscience entity e, τ is a threshold. Since t is a span of D, it must be mapped to a standard form of an entity, which is discussed in the entity disambiguation step below.

We found that LLM-based entity extraction methods achieve higher precision but lower recall. In contrast, the similaritybased retrieval using NV-Embed-v2 demonstrates higher recall but relatively lower precision. To benefit from both these methods, we developed a new approach combining the entities

Algorithm 1 Context Retrieval Using Entity Specific Spans

```
v_a: Normalized query embedding vector
C: All normalized context embedding vectors
S_c: All normalized embedding vectors of entity-specific
    spans of context C
threshold: Default value 0.05
get\_similar\_context(q) // q is the query
   S_1, S_2 \leftarrow top2values(v_q \cdot C^T) // S_1, S_2 are similarity scores
    C_1 \leftarrow \text{Context corresponding to similarity score } S_1
    C_2 \leftarrow \text{Context corresponding to similarity score } S_2
    score_diff \leftarrow S_1 - S_2 // S_1 > S_2
    if score\_diff < threshold
        ES_1 \leftarrow \max(v_q \cdot S_{c1}^T)
        ES_2 \leftarrow \max(v_a \cdot S_{c2}^T)
        span\_weight \leftarrow get span weight (score\_diff)
        S_1' \leftarrow S_1 * (1 - span\_weight) + ES_1 * span\_weight
        S_2' \leftarrow S_2 * (1 - span\_weight) + ES_2 * span\_weight
        S_1 \leftarrow \max(S_1, S_1')
        S_2 \leftarrow \max(S_2, S_2')
    if S_1 > S_2
        return C1
    else
        return C<sub>2</sub>
```

extracted by the LLM-based and NV-Embed-v2 similarity methods. We then filter this combined set again using a one-shot prompt and Llama-3.3-70B model to eliminate false positives. This method was finally used to extract entities in the subsequent steps.

B. Entity Disambiguation

A neuroscience entity can appear in various surface forms, including synonyms, abbreviations, and lexical variations such as singular and plural or hyphenated versions. It is essential to map these different representations to a standard name before using them in downstream tasks like knowledge graph (KG) construction or KG-based retrieval. When creating the vocabulary V of neuroscience entities from the NIFSTD [18] and MESH ontologies, we use the skos:prefLabel, if available, as the standard name for each entity; otherwise, rdfs:label is used. Additionally, we maintain a mapping table to map other names and abbreviations to the standard name. For entity names extracted by LLM-based methods, the cosine similarity with a high threshold value is used to map the document spans to standard names in vocabulary [24]; if no similar name is found, the span is treated as a new entity and added to the vocabulary.

C. Entity Centric Retrieval

We utilized PubMedQA [25] labeled and unlabeled QA instances to assess the entity-centric retrieval capabilities of large language models (LLMs). The PubMedQA dataset consists of question-context pairs derived from the abstracts of research articles. To focus specifically on neuroscience, we filtered the neuroscience question-context pairs from the PubMedQA dataset using the Llama-3.3-70B model in two steps. First, we provided a one-shot prompt to the Llama-3.3-70B model, asking it to identify and filter the neuroscience-specific contexts. Then, we prompted the model again to extract

TABLE I
WEIGHTAGE OF SPAN SIMILARITY SCORE

Score difference	Span weightage
<= 0.01	0.10
<= 0.02	0.15
<= 0.03	0.20
<= 0.04	0.25
<= 0.05	0.30

the neuroscience entities from the filtered contexts. Ultimately, we selected the contexts that contained at least one neuroscience entity for the subsequent tasks. Using these neuroscience-specific question-context pairs, we formulated a retrieval task in which, given a question, the corresponding context needs to be retrieved.

We used the Llama-3.3-70B model with a few-shot prompt to extract entity-centric texts from the contexts. The model is given a set of entities mentioned in a context and asked to summarize the entity-specific texts from the context paragraph [24]. We then used the NV-Embed-V2 model to generate the embedding for the questions, contexts, and entity-centric text spans.

We developed an algorithm for the retrieval task, as illustrated in "Algorithm 1," which utilizes a weighted combination of similarity scores of complete context text and entity-centric text spans. The embedding of the entire context captures the overall semantics, while the embedding of entityspecific spans focuses on the semantics relevant to particular entities. When two contexts closely match a given question, we calculate the overall similarity score by incorporating the similarity scores of the entity-specific texts, thereby enhancing retrieval performance. We conducted experiments to determine how varying the weight of text span similarity affects the overall similarity score calculation. Our findings revealed that the performance of the retrieval task improves when the weight of the entity-specific text span similarity score is gradually increased, particularly as the difference in the full context similarity score grows. The final weight assigned to the entityspecific text spans is presented in Table 1.

D. KG Construction and Q&A Using KG

We used Entrez Programming Utilities to extract research articles from the PubMed Central (PMC) archive. The PMC archive was searched using various subcategories of the *Neurosciences* MeSH category. However, this search yielded limited full-text neuroscience publications, so we collected several hundred additional high-quality articles from reputable neuroscience journals. More detailed information about the selection process and the selected articles can be found in the Supplementary material. The above process yielded 1,764 full-text articles and 1,484 abstract-only research articles, which were utilized to construct the prototype knowledge graph. The texts of these research articles were downloaded and organized into different paragraphs corresponding to the various sections of each article.

Algorithm 2 Entity Augmented Sub-graph Retrieval

 v_a : Normalized query embedding vector

$get_relevant_paragraphs(q)$ // q is the query

Extract the neuroscience entities Ent_q from q, where $Ent_q = \{ent_q | ent_q \text{ is a neuroscience entity in } q\}$ $V_{rec} = \{v_{rec} | v_{rec} \in V_r \text{ and entity name}(v_{rec}) \in Ent_r\}$

 $V_{ent} = \{v_{ent} \mid v_{ent} \subseteq V_e \text{ and } entity_name(v_{ent}) \in Ent_q \}$

if $|V_{ent}| == 1$

 V_{ent} has one vertex v_{ent}

Collect all relevant edge and paragraph node pairs P_{ep} where,

 $P_{ep} = \{(e_{ep}, v_p) \mid e_{ep} = (v_{ent} \times v_p) \in E_{ep} \text{ and } v_p \in V_p$ Collect the relation descriptions and performs toyle T, when

Collect the relation descriptions and paragraph texts T_{ep} where $T_{ep} = \{(relation_text(e_{ep}), para_text(v_p) \mid (e_{ep}, v_p) \in P_{ep}\}$ else if $|V_{ent}| == 2$

 V_{ent} has two vertices v_{ent}^1 and v_{ent}^2

Collect all the edges $E' \subseteq E_{ee}$ in the paths between v_{ent}^1 and v_{ent}^2 where max path length = 2

Collect the relation descriptions and paragraph texts T_{ep} where $T_{ep} = \{(relation_text(e_{ee}), para_text(e_{ee}) \mid e_{ee} \in E'\}$ else

Extract two most important entities from q using LLM

 v_{ent}^1 , $v_{ent}^2 \leftarrow$ vertices corresponding to the important entities $v_{ent}^{other} \leftarrow$ vertices corresponding to the other entities

Collect all the edges $E_1' \subseteq E_{ee}$ in the paths between v_{ent}^1 and v_{ent}^{other} where max path length = 2

Collect all the edges $E_2' \subseteq E_{ee}$ in the paths between v_{ent}^{other} and v_{ent}^2 where max path length = 2

Collect the relation descriptions and paragraph texts T_{ep} where $T_{ep} = \{(relation_text(e_{ee}), para_text(e_{ee}) \mid e_{ee} \in E_1' \cup E_2'\}$

Re-rank the texts in T_{ep} using semantic similarity score against v_q $top_k \leftarrow$ top k paragraphs according to rank return top_k

We extracted 22547 neuroscience entities from these research articles. The relationships among the entities were identified using the Llama-3.3-70B model through a few-shot prompt. The model was given a paragraph along with a set of entities from that paragraph and was asked to extract all text spans describing the relationships between pairs of entities [58]. We also extracted the relation between an entity and the containing paragraph. These extracted text spans, which convey the relationships in natural language, were stored as relation attributes in the knowledge graph. In total, we extracted 612272 relations. The constructed knowledge graph can be defined as G = (V, E), where $V = V_e \cup V_p$ and $E = E_{ee} \cup E_{ep}$, such that

- $V_e = \{v_e\}$, where $v_e = (id, entity_name)$ represent a node for a neuroscience entity.
- $V_p = \{v_p\}$, where $v_p = (id, paragraph_path)$ represent a node for a paragraph from research articles.
- $E_{ee} = \{e_{ee}\} \subseteq V_e \times V_e$, where $e_{ee} = (id, relation_text, paragraph_path)$ represents an entity-entity relationships "RELATED_TO".
- $E_{ep} = \{e_{ep}\} \subseteq V_e \times V_p$, where $e_{ee} = (id, relation_text)$ represents an entity-paragraph relationships "DESCRIBES".

To evaluate the effectiveness of the knowledge graph, we synthetically generated 493 neuroscience questions based on the neuroscience entities present in the graph. A subset of entities was selected according to their importance, determined by the degree of the corresponding node. We chose 102 entities

TABLE II
NEUROSCIENCE TOPICS FOR ENTITY EXTRACTION EVALUATION

Neuroscience topics	Research articles
	covering the topic
Neuropathology	1
Neurodegenerative Disease	1
Regenerative Neuroscience	1
Cellular and Molecular Neuroscience	2
Developmental Neuroscience	1
Neuroimaging	1
Neuroplasticity	1
Adult Neurogenesis	1
Clinical Neuroscience	2
Cognitive Neuroscience	2
Neuroanatomy	2
Systems Neuroscience	1
Neurochemistry	1
Neurophysiology	1

with diverse degrees; for example, the *hippocampus*, *astrocytes*, and *amygdala* are significant nodes in the graph with degrees greater than 2000, whereas entities such as the *superior longitudinal fasciculus* and *chronic traumatic encephalopathy* have degrees around 50. From this set of 102 entities, we randomly selected two or three that do not have a direct connection in the graph. These entities are then provided to the gemini-2.0-flash-thinking-exp-01-21 model, having enhanced thinking and reasoning capabilities, to generate a neuroscience question that can be used to explore the relationships or interactions among them.

We developed an entity-augmented retrieval algorithm [24], as outlined in "Algorithm 2," to extract the subgraph relevant to a given question. The algorithm returns a set of paragraphs that can be used to answer the question. The knowledge graph (KG) was created using Neo4j, and we utilized Cypher queries when necessary to query the graph database. To compare the paragraphs retrieved by our algorithm against those from the baseline RAG, we retrieved another set of paragraphs by comparing the cosine similarity of the paragraph and query embedding vectors generated by the NV-Embed-v2 model. The gemini-2.0-flash-thinking-exp-01-21 model was then tasked with determining which set of paragraphs is best suited to answer the question. To mitigate the position bias [30], we asked the model to evaluate the sets of paragraphs twice by switching their positions in the prompt. Additionally, we ensured both sets contained the same number of paragraphs to minimize verbosity bias [30].

E. Analysis of the Reference Fabrication Issue in LLMs

We generated 10 neuroscience questions to investigate the issue of reference fabrication in large language models (LLMs). Three LLMs, Llama-3.1-405B, Llama-3.3-70B, and gemini-2.0-flash-thinking-exp-01-21, were asked to answer these questions and provide the corresponding reference publications. Each model was required to include the title and

TABLE III
PERFORMANCE OF ENTITY EXTRACTION APPROACHES (MACRO-AVERAGED VALUES FROM ALL SELECTED ABSTRACTS)

Approach	Model	Precision	Recall	$\mathbf{F_1}$
	Llama-3.1-405B	0.973	0.630	0.747
One-shot prompt	Llama-3.1-70B	0.917	0.498	0.625
	Llama-3.3-70B	0.950	0.587	0.708
Entity Linking Heing	NV-Embed-v2	0.612	0.656	0.627
Entity Linking Using Cosine similarity	Finetuned bge-	0.824	0.296	0.417
	large-en-v1.5			
Entity Linking Using		0.855	0.476	0.596
Fuzzy similarity				
One-shot prompt + Entity	Llama-3.3-70B,	0.897	0.806	0.842
Linking + Filtering	NV-Embed-v2			

TABLE IV
PERFORMANCE OF PUBMEDQA CONTEXT RETRIEVAL TASK

Threshold	Question Type	No. of Questions	Span Weightage	Average Precision@1
	All	8579	0	93.43
			1	93.15
0.05 -			Variable	93.85
	Top 2 context similarity < 0.05	1358	0	62.96
			1	61.19
			Variable	65.61
0.01 -	All	8579	0	93.43
			1	93.50
			Variable	93.61
	Top 2 context similarity < 0.01	312	0	43.59
			1	45.51
			Variable	48.72

DOI of the referenced publications.

Llama-3.1-405B cited 40 references in its answers, while Llama-3.3-70B and gemini-2.0-flash-thinking-exp-01-21 used 38 and 41 references, respectively. To verify the authenticity of these references, we searched for the publications using their DOIs and titles (for the closest title match) in the Semantic Scholar database, utilizing the Semantic Scholar Academic Graph API [60]. If the search returned any result, we manually compared the titles to verify whether the cited reference exists and relevant.

IV. RESULTS AND DISCUSSION

A. Extraction of Neuroscience Entities

We selected 12 neuroscience research papers that cover a diverse range of topics, as illustrated in Table II, to evaluate the entity extraction methods. Entities were extracted from the abstracts of these articles using different extraction techniques, resulting in a total of 1,209 neuroscience entities. We manually reviewed the entities extracted from each abstract and checked for correct, incorrect, and missed entities. The performance of different unsupervised entity extraction methods is shown in Table III.

The LLM-based methods have high precision but lower recall as they fail to extract many entities. Also, LLM-based methods extract the exact text spans from the given text, which often has different surface forms, e.g., "peripheral neuron" and "peripheral neurons". It's necessary to map them to the standard form before any downstream task. As expected, the performance improves among LLMs as the number of parameters of the model increases. However, the hardware cost of running the models also rises with the increased parameter count. We ran Llama-3.1-70B and Llama-3.3-70B models using 4 A100 GPUs on an on-premises NVIDIA DGX server. For the Llama-3.1-405B model, we utilized the NVIDIA NIM APIs.

The entity extraction methods that utilize neuroscience vocabulary for entity linking require less expensive hardware. These methods map the span of the source texts to specific vocabulary terms. As a result, the generated entities exhibit fewer lexical variations; however, these methods tend to have lower precision than those based on large language models (LLMs). They also show lower recall rates, except for the entity linking using the NV-Embed-v2 model.

Entity linking with NV-Embed-v2 has the highest recall. As a result, the combined approach of the one-shot method using Llama-3.3-70B and the extraction method based on NV-Embed-v2 shows a significantly higher F1 score of 0.84 than the others. This combined approach was used for large-scale entity extraction from the neuroscience papers. Combining the NV-Embed-v2 method with Llama-3.1-405B or larger models could further enhance performance, though this would also increase costs for large-scale entity extraction.

TABLE V
MOST AND LEAST EXPLORED ENTITIES IN KG

Neuroscience terms	Node degree
Hippocampus	3002
Astrocytes	2595
Amygdala	2220
Schizophrenia	1436
Pyramidal neuron	656
Oligodendrocyte	168
Precursor Cell	
5-HT1A receptor	61
Stria medullaris	1
Moebius syndrome	1

TABLE VI COMPARISON OF THE ANSWERS GENERATED FROM KG

Question Type	Position of Sub-graph Retrieved Texts in Prompt		
7.1	First	Second	
Total Evaluated	493	493	
Sub-graph Retrieved	277	270	
Texts Deemed Best	(56.19%)	(54.77%)	
Best Texts Common in Both Positions	202		

B. Entity Centric Retrieval

The results of the PubMedQA retrieval task with different weights for entity-centric spans are shown in Table IV. This table highlights how the entity-centric spans and their weightage in the similarity calculation impact the accuracy of the retrieval process. The retrieval task involving the PubMedQA dataset is challenging because the context section containing the actual answer to the query has been removed. When span weightage is zero, retrieval is performed solely based on the embedding vectors generated by NV-Embed-V2, achieving an impressive accuracy of 93.43%. Improving this accuracy further is even more challenging. However, as we can see in Table IV, the overall accuracy is further enhanced to 93.85% by using the varying weightage of the entity-centric spans in the similarity score calculation.

Notably, the enhancement in accuracy is even more pronounced when we focus on questions where the top two retrieved contexts exhibit a minimal difference in their similarity scores. In these cases, the accuracy improves by 2.65% and 5.13% when the differences are less than 0.05 and 0.01, respectively. This increase indicates that the entity-centric spans extracted by LLM hold significant semantic relevance.

On the other hand, when the embedding of the entire context is considered, it retains the semantics of the complete text. If the entity-centric spans are given a maximum weight of 1 in the overall similarity calculation, the retrieval accuracy decreases to 93.15%. This result emphasizes the importance of the complete text's semantics. Accuracy improves only when the final similarity calculation combines context-level and entity-centric semantics.

TABLE VII

ANALYSIS OF LLM'S REFERENCE FABRICATION PROBLEM

Search By	Reference Type	gemini- 2.0-flash- thinking- exp-01-21	Llama- 3.1-405B	Llama- 3.3-70B
	Total Cited References	41	40	38
	Not Found	16	18	16
DOI	Title Mismatch	25	22	22
υоι -	Incorrect DOIs	41	40	38
	(%)	(100%)	(100%)	(100%)
Title	Not Found	27	32	23
Title - (Closest - Match)	Title Mismatch	11	3	1
	Incorrect Titles	38	35	24
	(%)	(92.68%)	(87.50%)	(63.16%)

C. Q&A Using KG

The knowledge graph contains valuable information from the research articles and can provide new insights, such as identifying the most and least explored research areas. Table V presents some of the most and least mentioned neuroscience entities, determined by the degree of their corresponding nodes in the knowledge graph. The *Hippocampus* and *Astrocytes* are among the most frequently discussed entities. *Moebius syndrome* and *Stria medullaris* are the least mentioned, suggesting that these concepts have not been widely explored in the selected research articles.

Table VI shows the relevance of the paragraphs retrieved by the entity-augmented sub-graph retrieval algorithm for answering the synthetically generated neuroscience questions. The results indicate that the LLM preferred the paragraphs retrieved using the knowledge graph over those retrieved by the baseline RAG system for more than 50% of the questions. When paragraphs retrieved by the proposed algorithm were presented first in the LLM prompt, the retrieved texts for 277 out of 493 questions, or 56.19%, were deemed the best. Conversely, when paragraphs were placed second, texts for 270 out of 493 questions, or 54.77%, were rated the best. These findings underscore the effectiveness of the KG in providing answers to questions.

Additionally, we investigated the overlap between the two sets of questions for which the paragraphs retrieved using KG were identified best by the LLM, and 202 were common to both sets, confirming that the paragraphs retrieved using KG were superior for a significant percentage of the questions. Since the questions involve entities from different articles or parts of articles, our analysis shows that the RAG system fails to retrieve the relevant intermediate texts. In contrast, the proposed algorithm can discover and retrieve the same using KG.

D. Reference Fabrication Issue in LLMs

We analyzed the reference fabrication issue present in the answers generated by LLM for the neuroscience questions. The results of our study are summarized in Table VII. The DOIs are incorrect in all the references for all three models. Specifically,

the DOIs were either fabricated and could not be located, or the titles of the references generated by the models did not match the titles of the actual publications associated with those DOIs.

When we searched for the references using their titles, we found that a significant percentage of them either could not be found or that the closest matched title was different. Among the three models, the gemini-2.0-flash-thinking-exp-01-21 model had the highest rate of incorrect titles at 92.68%, while Llama-3.3-70B had the lowest at 63.16%.

Our analysis shows that LLMs generate spurious neuroscience references when the answers are generated from their parametric knowledge. Also, we only checked for the existence of the references, but it still needs to be analyzed if the references are relevant to the generated answers. This analysis highlights the significance of knowledge graphs in providing evidence-based answers in specialized fields like neuroscience.

V. CONCLUSION

In this work, we present our research contributions focused on methods for knowledge discovery from neuroscience research articles. We developed an unsupervised entity extraction method that leverages LLMs and entity linking with specialized neuroscience vocabulary. This method works without requiring annotated data and can accurately extract neuroscience entities from research articles. Next, we explored the capability of LLMs to extract entity-centric relevant texts from neuroscience literature. Our findings indicate that the texts extracted by the LLM, which are focused on specific entities, possess semantic relevance. We utilized this text extraction capability to identify relationships between entities mentioned in neuroscience texts and subsequently constructed a knowledge graph.

We demonstrated the effectiveness of the knowledge graph by applying an entity-augmented information retrieval algorithm that answers neuroscience questions by revealing new knowledge. This KG was constructed from just a few thousand neuroscience research articles, while millions of articles have been published. More research articles should be added to enhance the KG's effectiveness in discovering neuroscience knowledge. However, this will increase the complexity of the KG, making it important to re-evaluate the performance of the retrieval algorithm as the complexity grows.

REFERENCES

- A. W. K. Yeung, T. K. Goto, and W. K. Leung, "The changing landscape of neuroscience research, 2006–2015: A bibliometric study," Front. Neurosci., vol. 11, 2017.
- [2] A. Paul, M. Segreti, P. Pani, E. Brunamonti, and A. Genovesio, "The increasing authorship trend in neuroscience: A scientometric analysis across 11 countries," IBRO Neurosci. Rep., vol. 17, pp. 52–57, 2024.
- [3] K. Fahd, Y. Miao, S. J. Miah, S. Venkatraman, and K. Ahmed, "Knowledge graph model development for knowledge discovery in dementia research using cognitive scripting and next-generation graphbased database: a design science research approach," Soc. Netw. Anal. Min., vol. 12, no. 1, 2022.
- [4] T. F. A. França, "Exploring undiscovered public knowledge in neuroscience," Eur. J. Neurosci., vol. 60, no. 5, pp. 4723–4737, 2024.
- [5] K. Narasimhan, "Scaling up neuroscience," Nat. Neurosci., vol. 7, no. 5, pp. 425–425, 2004.A. Harrison, private communication, May 1995.

- [6] V. Karpukhin et al., "Dense passage retrieval for open-domain question answering," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.
- [8] H. Lee, S. Yang, H. Oh, and M. Seo, "Generative Multi-hop Retrieval," in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 1417–1436.
- [9] W. Xiong et al., "Answering complex open-domain questions with multihop dense retrieval," in International Conference on Learning Representations 2021.
- [10] N. Li et al., "KGHC: a knowledge graph for hepatocellular carcinoma," BMC Med. Inform. Decis. Mak., vol. 20, no. S3, 2020.
- [11] J. Yuan et al., "Constructing biomedical domain-specific knowledge graph with minimum supervision," Knowl. Inf. Syst., vol. 62, no. 1, pp. 317–336, 2020.
- [12] M. Shardlow et al., "A text mining pipeline using active and deep learning aimed at curating information in computational neuroscience," Neuroinformatics, vol. 17, no. 3, pp. 391–406, 2019.
- [13] X. Chai et al., "Deep learning-based large-scale named entity recognition for anatomical region of mammalian brain," Quant. Biol., vol. 10, no. 3, pp. 253–263, 2022.
- [14] X. Chai et al., "Knowledge mining of brain connectivity in massive literature based on transfer learning," Bioinformatics, vol. 40, no. 12, 2024.
- [15] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," IEEE Trans. Knowl. Data Eng., vol. 34, no. 1, pp. 50– 70, 2022.
- [16] Y. Labrak, M. Rouvier, and R. Dufour, "A zero-shot and few-shot study of instruction-finetuned Large Language Models applied to clinical and biomedical tasks," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024.
- [17] V. K. Keloth et al., "Advancing entity recognition in biomedicine via instruction tuning of large language models," Bioinformatics, vol. 40, no. 4, 2024.
- [18] F. T. Imam, S. D. Larson, A. Bandrowski, J. S. Grethe, A. Gupta, and M. E. Martone, "Development and use of ontologies inside the neuroscience information framework: A practical approach," Front. Genet., vol. 3, 2012.
- [19] W. Shen, Y. Li, Y. Liu, J. Han, J. Wang, and X. Yuan, "Entity linking meets deep learning: Techniques and solutions," IEEE Trans. Knowl. Data Eng., pp. 1–1, 2021.
- [20] G. Aaron et al., "The Llama 3 herd of models," arXiv [cs.AI], 2024.
- [21] R. Fornasiere, N. Brunello, V. Scotti, and M. Carman, "Medical information extraction with Large Language Models," ICNLSP, pp. 456– 466, 2024.
- [22] J. Dagdelen et al., "Structured information extraction from scientific text with large language models," Nat. Commun., vol. 15, no. 1, 2024.
- [23] T. Xu, Y. Gu, M. Xue, R. Gu, B. Li, and X. Gu, "Knowledge graph construction for heart failure using large language models with prompt engineering," Front. Comput. Neurosci., vol. 18, 2024.
- [24] I. Lopez et al., "Clinical entity augmented retrieval for clinical information extraction," NPJ Digit. Med., vol. 8, no. 1, pp. 1–11, 2025.
- [25] Q. Jin, B. Dhingra, Z. Liu, W. Cohen, and X. Lu, "PubMedQA: A Dataset for Biomedical Research Question Answering," 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), 2019.
- [26] C. Lee et al., "NV-Embed: Improved techniques for training LLMs as generalist embedding models," arXiv [cs.CL], 2024.
- [27] J. Gravel, M. D'Amours-Gravel, and E. Osmanlliu, "Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions," Mayo Clinic Proceedings: Digital Health, vol. 1, no. 3, pp. 226–234, 2023.
- [28] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020, pp. 9459–9474.
- [29] A. O. M. Saleh, G. Tur, and Y. Saygín, "SG-RAG: Multi-hop question answering with large Language Models through Knowledge Graphs," in Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024), 2024, pp. 439–448.

- [30] L. Zheng et al., "Judging LLM-as-a-judge with MT-bench and Chatbot Arena," in Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023, pp. 46595–46623.
- [31] L. Soldaini and N. Goharian, "QuickUMLS: a fast, unsupervised approach for medical concept extraction," in MedIR Workshop at SIGIR, 2016.
- [32] B. Paassen et al., "Ontology-based extraction of structured information from publications on preclinical experiments for spinal cord injury treatments," in Proceedings of the Third Workshop on Semantic Web and Information Extraction, 2014, pp. 25–32.
- [33] Y. Nie, S. Wang, and M. Bansal, "Revealing the importance of semantic retrieval for machine reading at scale," 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), 2019, pp. 2553–2566.
- [34] J. Wu and R. Mooney, "Entity-focused dense passage retrieval for outside-knowledge visual question answering," in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 8061–8072.
- [35] A. Asai, K. Hashimoto, H. Hajishirzi, R. Socher, and C. Xiong, "Learning to retrieve reasoning paths over Wikipedia graph for question answering," in International Conference on Learning Representations, 2020.
- [36] D. Yifan, Y. Michael, and W. Tim, "Span-Oriented Information Extraction -- A unifying perspective on information extraction," arXiv [cs.CL], 2024.
- [37] R. Das et al., "Multi-step entity-centric information retrieval for multi-hop question answering," in Proceedings of the 2nd Workshop on Machine Reading for Question Answering, 2019, pp. 113–118.
- [38] X. Su et al., "KGAREVION: AN AI AGENT FOR KNOWLEDGE-INTENSIVE BIOMEDICAL QA," in International Conference on Learning Representations, 2025.
- [39] L. Murali, G. Gopakumar, D. M. Viswanathan, and P. Nedungadi, "Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study," J. Biomed. Inform., vol. 143, no. 104403, p. 104403, 2023.
- [40] X. Dai, X. Yan, K. Zhou, Y. Wang, H. Yang, and J. Cheng, "Convolutional Embedding for Edit Distance," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 599–608.
- [41] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," Sci. Data, vol. 10, no. 1, pp. 1–16, 2023.
- [42] A. Sabbir, A. Jimeno-Yepes, and R. Kavuluru, "Knowledge-based biomedical word sense disambiguation with neural concept embeddings," in 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE), 2017, pp. 163–170.
- [43] X. Zhou, X. Zhang, and X. Hu, "MaxMatcher: Biological concept extraction using approximate dictionary lookup," in Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1145–1149.
- [44] D. Loureiro and A. M. Jorge, "MedLinker: Medical entity linking with neural representations and dictionary matching," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 230–237.
- [45] N. Perera, M. Dehmer, and F. Emmert-Streib, "Named entity recognition and relation detection for biomedical information extraction," Front. Cell Dev. Biol., vol. 8, 2020.
- [46] D. Gillick et al., "Learning dense representations for entity retrieval," in Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), 2019.
- [47] D. Li, A. Kadav, A. Gao, R. Li, and R. Bourgon, "Automated clinical data extraction with knowledge conditioned LLMs," in Proceedings of the 31st International Conference on Computational Linguistics: Industry Track, pages 149–162, Abu Dhabi, UAE. Association for Computational Linguistics., 2024.
- [48] L. Wu, F. Petroni, M. Josifoski, S. Riedel, and L. Zettlemoyer, "Scalable zero-shot entity linking with dense entity retrieval," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [49] P. Ernst, A. Siu, D. Milchevski, J. Hoffart, and G. Weikum, "DeepLife: An entity-aware search, analytics and exploration platform for health and life sciences," in Proceedings of ACL-2016 System Demonstrations, 2016, pp. 19–24.

- [50] X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A unified MRC framework for named entity recognition," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.
- [51] N. Bhagat, O. Mackey, and A. Wilcox, "Large language models for efficient medical information extraction," AMIA Summits Transl. Sci. Proc., vol. 2024, pp. 509–514, 2024.
- [52] D. Banerjee, D. Chaudhuri, M. Dubey, and J. Lehmann, "PNEL: Pointer network based end-to-end entity linking over knowledge graphs," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 21–38.
- [53] S. J. Jung, H. Kim, and K. S. Jang, "LLM based biological named entity recognition from scientific literature," in 2024 IEEE International Conference on Big Data and Smart Computing (BigComp), 2024, pp. 433–435.
- [54] M. P. Polak and D. Morgan, "Extracting accurate materials data from research papers with conversational language models and prompt engineering," Nat. Commun., vol. 15, no. 1, pp. 1–11, 2024.
- [55] N. Kolitsas, O.-E. Ganea, and T. Hofmann, "End-to-end neural entity linking," in Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018, pp. 519–529.
- [56] Ö. Sevgili, A. Shelmanov, M. Arkhipov, A. Panchenko, and C. Biemann, "Neural entity linking: A survey of models based on deep learning," Semant. Web, vol. 13, no. 3, pp. 527–570, 2022.
- [57] D. Sorokin and I. Gurevych, "Mixing context granularities for improved entity linking on question answering data across entity categories," in Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, 2018, pp. 65–75.
- [58] C. Peng, X. Yang, Z. Yu, J. Bian, W. R. Hogan, and Y. Wu, "Clinical concept and relation extraction using prompt-based machine reading comprehension," J. Am. Med. Inform. Assoc., vol. 30, no. 9, pp. 1486– 1493, 2023.
- [59] S. Xiao, Z. Liu, P. Zhang, N. Muennighoff, D. Lian, and J.-Y. Nie, "C-pack: Packed resources for general Chinese embeddings," in Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024, vol. 33, pp. 641–649.
- [60] R. Kinney et al., "The semantic Scholar open data platform," arXiv [cs.DL], 2023.
- [61] N. Muennighoff, N. Tazi, L. Magne, and N. Reimers, "MTEB: Massive Text Embedding Benchmark," arXiv [cs.CL], 2022.