Preprints.org

Article

PKG-LLM: A Framework for Predicting GAD and MDD Using Knowledge Graphs and Large Language Models in Cognitive Neuroscience

Ali Sarabadani , Hadis Taherinia , Niloufar Ghadiri , Ehsan Karimi Shahmarvandi , Ramin Mousa [*]

Article

PKG-LLM: A Framework for Predicting GAD and MDD Using Knowledge Graphs and Large Language Models in Cognitive Neuroscience

Ali Sarabadani [1], Hadis Taherinia [2], Niloufar Ghadiri [3], Ehsan Karimi Shahmarvandi [4] and Ramin Mousa [5,*]

[1] Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran; a.sarabadani@stu.qom.ac.ir
[2] Islamic Azad University Science and Research Branch, Tehran, Iran
[3] Islamic Azad University west Tehran branch, Tehran, Iran
[4] University of Amsterdam, Amsterdam, Netherlands
[5] University of Zanjan, Zanjan, Iran
[*] Correspondence: Raminmousa@znu.ac.ir

Abstract: **Purpose:** This research project has a single purpose: the construction and evaluation of PKG-LLM, a knowledge graph framework whose application is primarily intended for cognitive neuroscience. It also aims to improve predictions of relationships among neurological entities and improve named entity recognition (NER) and relation extraction (RE) from large neurological datasets. Employing the GPT-4 and expert review, we aim to demonstrate how this framework may outperform traditional models by way of precision, recall, and F1 score, intending to provide key insights into possible future clinical and research applications in the field of neuroscience. **Method:** In the evaluation of PKG-LLM, there were two different tasks primarily: relation extraction (RE) and named entity recognition (NER). Both tasks processed data and obtained performance metrics, such as precision, recall, and F1-score, using GPT-4. Moreover, there was an integration of an expert review process comprising neurologists and domain experts reviewing those extracted relationships and entities and improving such final performance metrics. Model comparative performance was reported against StrokeKG and Heart Failure KG. On the other hand, PKG-LLM evinced itself to link prediction-in-cognition through metrics such as Mean Rank (MR), Mean Reciprocal Rank (MRR), and Precision at K (P@K). The model was evaluated against other link prediction models, including TransE, RotatE, DistMult, ComplEx, ConvE, and HolmE. **Findings:** PKG-LLM demonstrated competitive performance in both relation extraction and named entity recognition tasks. In its traditional form, PKG-LLM achieved a precision of 75.45%, recall of 78.60%, and F1-score of 76.89% in relation extraction, which improved to 82.34%, 85.40%, and 83.85% after expert review. In named entity recognition, the traditional model scored 73.42% precision, 76.30% recall, and 74.84% F1-score, improving to 81.55%, 84.60%, and 82.99% after expert review. For link prediction, PKG-LLM achieved an MRR of 0.396, P@1 of 0.385, and P@10 of 0.531, placing it in a competitive range compared to models like TransE, RotatE, and ConvE. **Conclusion:** This study showed that PKG-LLM mainly outperformed the existing models by adding expert reviews in its application in extraction and named entity recognition tasks. Further, the model's competitive edge in link prediction lends credence to its capability in knowledge graph construction and refinement in the field of cognitive neuroscience as well. PKG-LLM's superiority over existing models and its ability to generate more accurate results with clinical relevance indicates that it is a significant tool to augment neuroscience research and clinical applications. The evaluation process entailed using GPT-4 and expert review. This approach ensures that the resulting knowledge graph is scientifically compelling and practically beneficial in more advanced cognitive neuroscience research.

Keywords: knowledge graphs; large language models; Generalized Anxiety Disorder; Major Depressive Disorder; cognitive neuroscience

1. Introduction

Mental health disorders such as Generalized Anxiety Disorder (GAD) and Major Depressive Disorder (MDD) give rise to some of the most common and debilitating psychiatric conditions in the whole world. Although both disorders share symptoms such as constant worrying, lack of energy, and insomnia, the neurobiological bases that provoke the two and their treatment pathways are vastly different; thus, there must be pinpointed diagnostic tools. The overlap in symptoms often confounds the accurate diagnosis, leading to delay or inadequate treatment of these disorders as well as poor patient outcomes. Therefore, growing demands exist for better methods that will rightly differentiate the two and, thus, improve care quality in neuropsychiatry.

Developments in artificial intelligence (AI) have recently offered intriguing possibilities for improving psychiatric diagnosis, especially in knowledge representation and natural language processing. Knowledge graphs (KGs) pave a solid path for representing complicated, domain-specific information; they encode entities and relationships between KGs. Cognitive neuroscience KGs can hold substantial structured knowledge of brain functions, cognitive processes, and neurological disorders. Besides, large language models (LLMs) have shown their exceptional ability to understand and produce human-like text, hence being valuable tools for insight generation about clinical data.

This paper introduces PKG-LLM, a novel framework that leverages the power of KGs and LLMs to predict and differentiate between GAD and MDD. Specifically, the framework integrates knowledge from NeuroLex, a repository of structured cognitive neuroscience data, and NeuroMorpho, a database containing detailed morphological information of neurons, to build a robust, domain-specific KG. PKG-LLM leverages LLMs to analyze clinical data and discern subtle differences between two disorders. Bringing the structured, domain-specific intelligence KGs provide into the analytic heft of LLMs should refine diagnostic accuracy and translate into more personalized treatment strategies in psychiatry. The rest of this paper is structured as follows: Section 2 reviews related works in applying KGs and LLMs in mental health diagnostics. Section 3 details the methodology behind PKG-LLM, including the construction of the KG and the integration with LLMs. Section 4 presents the results of our evaluation using clinical datasets and benchmarks. Finally, Section 5 discusses the implications of the findings and potential future developments in this area.

2. Related Work

PolarisX, an auto-evolving knowledge graph that will automatically augment itself with information gleaned from news sites and social media, is introduced in this paper [?]. It mainly focuses on neologisms. By using the multilingual BERT model, this system is built to extract new relationships and incorporate them into the knowledge graph to follow all current changes within languages and introduce new concepts without language dependency. There are two main innovations in this knowledge graph: collecting neologisms and being independent of language in data collection.

In this paper [1], the authors propose exploring drug indication expansion through the use of knowledge graphs. The principal aim is to find new therapeutic uses for pre-existing drug targets. They discuss some methods along with hybrid knowledge graphs that combine gene data with scientific literature and show features of relationships among genes, disease, and tissues. Then, they test new prediction methods on shortest paths and embeddings, showing a significant improvement in prediction accuracy by integrating data regarding tissue expression. As it is transparent and interpretable, this knowledge graph can be used very effectively in drug indication expansion.

According to [2], the idea of knowledge graph (KG) applications is directed toward biomedical research. Knowledge graphs can automatically extract biomedical textual information and frame rules and apply label functions to identify associated relationships among biological entities such as genes and diseases. This research hastens the process of creating label functions, and at the same time, an advanced enriched up-to-date knowledge graph is constructed that could benefit pharmaceutical and biomedical research.

The knowledge graphs in healthcare have been created from various sources largely during the time of the pandemic, as in the case with COVID-19, in order to develop treatment procedures for the virus [3]. Such KGs have been equally found to be vital instruments in clinical decision support [4] and in discovering false health information, thus making them yet of great value to the healthcare sector [5]. By collating data from various domains, it has thus been possible to create such knowledge graphs that are more comprehensive and thus enable drawing of many meaningful conclusions in medical research and health care.

J. Xu et al. constructed knowledge graph using multitudes of integrated multi-source commons from infrastructures: PubMed literature, ORCID, MapAffil, and NIH ExPORTER project data [6]. The PubMed knowledge graph was used to link bio-entities drawn from 29 million PubMed abstracts through a BioBERT-based Named Entity Recognition (NER) model with entities extracted from multiple sources. Unique authors from ORCID and MapAffil as well as NIH ExPORTER funding data were incorporated for building the robust network of biomedical information [6]. It thus allows a better understanding of the scientific entity relationship and further discovery in biomedicine.

At present, researchers have found a way to integrate large language models (LLMs) in the construction and expansion of the healthcare and medical knowledge graphs (KGS). For instance, the DIAMOND-KG developed by Alharbi et al. is a medical context-aware drug indication knowledge graph. The study showed that around 71% of indications were related to at least one medical context. So it raises importance of these contextual information in improving the quality of KGs [7]. Furthermore, Mann et al. discussed a new framework for treatment discovery based on known symptoms or diseases. The new model consists of a generalized LLM which can learn from all existing knowledge sources available in the medical field and form a complete KG [8].

In one study, Wu et al. postulated reguloGPT, a GPT-4-based entity that is exceptionally competent at harvesting named entities, harvesting N-ary relationships, and predicting the context of molecular regulatory sentences. The model is developed to build knowledge graphs around molecular regulatory pathways [9]. On the contrary, Jiang et al. leveraged the power of patient-specific knowledge graphs and used them with a Bidirectional Attention-Enhanced Graph Neural Network (BAT GNN) that helped bring a change to medical decision-making and even diagnosis in the patient context settings by using LLMs [10].

Likewise, Xu et al. offered RAM-EHR, which translates knowledge sources into much-text format via retrieval-augmented multimodal means to be used in better medical prediction from electronic health records (EHR). Such integration between LLMs and KGs improves clinical decision-making [11]. Gao et al. were successful in the development of a diagnostic reasoning system, named DR.KNOWS, which employs KGs with the Unified Medical Language System (UMLS) and a clinical diagnosis model to improve diagnostic accuracy and interpretability [12].

Finally, Zhu et al. developed the REALM model, a novel approach to integrating clinical notes and multivariate time-series data using Retrieval-Augmented Generation (RAG) technology. This combination further improves diagnostic performance in complex medical cases by synergistically employing LLMs with knowledge graphs [13]. These studies together present a picture of increasing roles for LLMs in constructing and expanding medical knowledge graphs for advances in healthcare research and clinical decision-making. Numerous studies are now focusing on improving and developing medical knowledge graphs (peacekeeping) with large language models (LLM). Wu T. et al. have reviewed the progress on Chinese medical knowledge graphs and the contributions of LLMs on the construction and improvement of these graphs within the field in detail [14]. Agrawal et al. investigate the feasibility of using LLMs for zero-shot and few-shot information extraction from clinical text and demonstrate how it could be used on the models such as InstructGPT to extract relevant medical information with the lowest training burden [15].

Frisoni et al. have compared the performance of a graph-based transformer and pre-trained large language models, such as T5 and BART, with respect to the graph-to-text and text-to-graph tasks. Their observations yielded mixed results, showing that some models perform better in certain

contexts vis-à-vis biomedical applications [16]. On the other hand, Choudhary and Reddy made some investigations into complex logical reasoning over the knowledge graph by using LLMs and found that the capabilities of such models provide for an intricate reasoning task based on graph data [17].

In recent years, the construction of medical knowledge graphs has received attention to enhance the decision-making process in healthcare. Previous studies have concentrated on building knowledge graphs, which intuitively show the relationships between medical concepts to make the user experience be more user-friendly in retrieval of medical information [18]. As an example, Shanghai Shuguang Hospital built a knowledge graph for traditional Chinese medicine without success in automatically prescribing drugs for the patient [19]. Automating clinical prescription, for example, has been accomplished in one of the recent advancements in this area through the use of large language models (LLMs) for building medical knowledge graphs, as these models, through prompt engineering, have automated most steps in the construction of knowledge graphs, and have even been used to develop knowledge bases for the identification of certain conditions, for instance, autism spectrum disorder [20].

They have additionally begun using LLMs for more and more medical information extraction purposes. These models are successfully identifying medical entities and extracting relationships between entities and disambiguating them as well [21]. They prove to be an effective tool in knowledge graph completion rather than merely identification and relationship extraction. For instance, link predictions for knowledge graphs using ChatGPT have shown significant improvement in the prediction accuracy of medical information [22]. Finally, these methods will improve the quality of diagnosis and treatment in clinics.

This paper [23] proposes new ways of developing a heart failure knowledge graph using large language models (LLMs). It employs a method that combines prompt engineering with the TwoStepChat for improved performances compared to other BERT-based models, especially with regards to out-of-distribution OOD information. Moreover, this method greatly shortens the amount of time spent on obtaining medical information extraction.

3. Methodology

The following section describes the data collection process, preprocessing, and the proposed method.

3.1. Data Sources

The PKG-LLM framework promulgates the structured information contained in knowledge graphs and unstructured clinical information to formulate a holistic system for predicting GAD and MDD. The primary two sources of that structured data are NeuroLex and NeuroMorpho, while clinical text data provide the unstructured input.

- **NeuroLex:** NeuroLex is a comprehensive, structured ontology for cognitive neuroscience, offering hierarchical knowledge of brain regions, neuron types, and cognitive functions. Relevant entities (e.g., hippocampus, Anxiety, memory) are extracted using SPARQL queries, which retrieve data in the form of RDF triples (subject, predicate, object). These triples, such as the amygdala (associated with Anxiety), form the knowledge graph (KG) foundation, where nodes represent entities, and edges represent relationships between them. NeuroLex provides highly granular and standardized data, enabling PKG-LLM to systematically incorporate neurobiological structures and their cognitive associations.

- **NeuroMorpho:** NeuroMorpho offers a rich repository of 3D neuronal morphology data, providing detailed information about neuronal structures such as dendrites, axons, and soma size. Initially provided in SWC format, this morphological data is processed into a relational structure and integrated into the KG. Each neuron type is linked to its corresponding brain region and further connected to cognitive functions or psychiatric disorders, such as GAD and MDD. For example, a neuron from the prefrontal cortex (PFC) might be linked to mood regulation and associated with

MDD. This neuroanatomical data enhances the KG by adding structural depth to the relationships between brain regions and mental disorders.

In addition to these structured sources, unstructured clinical data from neuropsychiatric databases is used. These datasets contain diagnostic records, symptom descriptions, and patient outcomes. To align this unstructured data with the KG, a preprocessing pipeline is implemented:

- **Tokenization:** Clinical text is split into meaningful tokens (words, phrases) to facilitate analysis.
- **Lemmatization:** Words are reduced to their base or dictionary form, allowing for standardization across clinical variations.
- **Named Entity Recognition (NER):** Leveraging the capabilities of GPT-4 for NER tasks, medical entities such as symptoms, diagnoses, and treatments are extracted with high accuracy. GPT-4 enables the precise identification of entities from clinical notes and text, facilitating their integration into the knowledge graph (KG). For example, the term "persistent worry" might be linked to the "Anxiety" node in the KG, while "sleep disturbance" could be associated with a node connected to sleep regulation and mental health. This process enhances the depth and relevance of the knowledge graph by ensuring that critical medical terms are accurately mapped.

The PKG-LLM framework combines the ontological structured data from NeuroL and NeuroMorpho with the clinical data to build knowledge that correlates biological structures and clinical symptoms, facilitating better prediction of GAD and MDD.

3.2. Knowledge Graph Construction

The knowledge graph (KG) integrates domain-specific information from NeuroLex, NeuroMorpho, and clinical datasets. The process involves multiple steps:

3.2.1. Entity Extraction and Triple Formation

Entity Extraction: Using GPT-4 for Named Entity Recognition (NER), entities such as brain regions, neuron structures, and psychiatric symptoms are accurately extracted from sources like NeuroLex and NeuroMorpho. GPT-4 enables a more efficient and precise extraction process than traditional NER algorithms like BioBERT. Each entity is then mapped to its corresponding node in the knowledge graph (KG), ensuring a comprehensive and structured representation of neurological and psychiatric information within the KG. Mathematically, for a given input sentence S, the goal is to extract a set of entities $E = e_1, e_2, ..., e_n$, where e_i corresponds to a recognized entity in the ontology. The NER model assigns a probability $P(e_i|S)$, ensuring entities with high confidence are selected.

Triple Generation: Using dependency parsing and relation extraction techniques, triples (h, r, t) are generated, where h (head entity) and t (tail entity) are connected by relation r. For example, from the sentence "The amygdala is involved in anxiety regulation", the triple (Amygdala, involved in, Anxiety) is extracted. A relation extraction model computes the likelihood $P(r|h, t)$ for each possible relation between two entities. This relation prediction can be formulated as a classification task where r is predicted based on the context of h and t.

3.2.2. Graph Schema and KG Construction

Schema Definition: The KG is structured as a directed, labeled graph $G = (V, E)$, where V is the set of entities (nodes), and E represents directed edges (relations) between them. Each node belongs to one of the classes: brain regions, neuron structures, cognitive functions, or psychiatric symptoms. Relations between these nodes are defined by predicates such as "is part of", "influences", and "associated with". Graph Representation: The KG is stored in a Neo4j database, which supports efficient storage and querying. The triples generated earlier are stored in the form of relationships between entities in this graph.

3.2.3. Knowledge Graph Embeddings

We compute knowledge graph embeddings (KGE) using methods like TransE, TransH, or ComplEx to integrate the KG with machine learning models. These embeddings map entities and relationships to a continuous vector space, enabling the representation of entities as high-dimensional vectors.

- **TransE:** In TransE, a relation r is modeled as a translation from the head entity h to the tail entity t. The energy function for a triple (h, r, t) is defined as:

$$f(h, r, t) = ||h + r + t|| \tag{1}$$

  Where $h, r, t \in R^d$ are the embedding vectors of the head, relation, and tail, respectively, and $||.||$ denotes the L2-norm. The training objective is to minimize $f(h, r, t)$ for correct triples and maximize it for incorrect triples (negative sampling).

- **TransH (Translation on Hyperplanes):** TransH addresses the limitations of TransE by allowing each relation to have its hyperplane. Entities are first projected onto a relation-specific hyperplane, then the translation is performed. The projection of entity onto the hyperplane of relation is defined as:

$$e_r = e - W_r^T e w_r \tag{2}$$

  Where w_r is the normal vector for relation r, and e_r is the entity embedding projected onto the relation-specific hyperplane. The energy function is:

$$f(h, r, t) = |(|h_r + r - t_r|)| \tag{3}$$

  Where h_r are the head and tail entity embeddings after being projected onto the hyperplane of relation r.

- **ComplEx:** ComplEx extends TransE to model asymmetric relations by embedding entities and relations into complex-valued vector spaces. The scoring function for a triple is:

$$f(h, r, t) = \Re(< h, r, \hat{t} >) \tag{4}$$

  where $h, r, t \in C^d$ are complex vectors, $\hat{t}$ is the complex conjugate of t, and $< \cdot, \cdot, \cdot >$ denotes the trilinear dot product. This allows the model to capture non-symmetric relations common in cognitive neuroscience.

3.3. Integration with Large Language Models (LLMs)

PKG-LLM integrates large language models (LLMs) such as BERT and GPT with the KG to analyze clinical data and enhance the prediction task.

3.3.1. Preprocessing of Clinical Data

The clinical data comprises unstructured text that details patient symptoms, diagnoses, and treatment histories. This unstructured text is preprocessed through several steps:

- **Tokenization:** The text is tokenized using WordPiece tokenization (as employed in BERT) to effectively handle subword units and ensure that all terms, including medical terms, are accurately segmented.
- **Entity Recognition and Linking:** GPT-4 and models like ClinicalBERT are used for Named Entity Recognition (NER) to identify medical entities such as symptoms, diagnoses, and treatments. These entities are then linked to their corresponding nodes in the KG through entity-linking algorithms, ensuring semantic consistency across the knowledge graph.

3.3.2. LLM-Based Feature Extraction

The clinical text goes through the pre-trained large language models like GPT-4 or BERT, and these models produce contextualized embeddings for each token in the text. Therefore, these embeddings capture the semantic relationships between the entities discussed in the clinical data, allowing the model to find the subtle insight and the most crucial to a more accurate and more profound knowledge graph and prediction tasks. Let $T = t_1, t_2, \ldots, t_n$ represent the tokens in a clinical document. The LLM computes embeddings h_t^i for each token t_i:

$$h_t^i = LLM(t_i|T) \tag{5}$$

These token-level embeddings are then pooled (e.g., using average pooling or attention-based pooling) to form a document-level representation h_{doc}.

3.3.3. KG-Augmented Embeddings

To enhance the LLM embeddings, KG embeddings h_{KG} are integrated. The final feature vector z is obtained by concatenating the LLM-generated embeddings h_{doc} with the KG embeddings h_{KG}:

$$z = [h_{doc}; h_{KG}] \tag{6}$$

This hybrid representation enables the model to leverage both unstructured clinical data and structured knowledge from the KG.

3.3.4. Attention Mechanism for Symptom and Entity Interaction

An attention mechanism is applied to highlight relevant symptoms and brain regions during diagnosis. Let h_i represent the embedding for symptom i in the clinical document, and let g_i represent the corresponding KG embedding for the symptom. The attention weight α_i for symptom i is computed as:

$$\alpha_i = \frac{exp(h_i^T W g_i)}{\sum_i (h_j^T W g_i)} \tag{7}$$

Where W is a learned weight matrix. The final embedding is a weighted sum of the embeddings, emphasizing more relevant symptoms and brain regions.

3.4. Model Training and Optimization

The PKG-LLM framework is trained to predict GAD, MDD, or no diagnosis based on the hybrid embeddings z.

3.4.1. Neural Network Architecture

The input vector z is fed into a multi-layer feed-forward neural network. The hidden layers apply non-linear transformations using activation functions like ReLU:

$$h^{l+1} = ReLU(W^l h^l + b^l) \tag{8}$$

Where W^l and b^l are the weight matrix and bias vector for layer l, respectively.

3.4.2. Loss Function and Optimization

The model is trained using the cross-entropy loss function for multi-class classification. The predicted probability for each class (GAD, MDD, or no diagnosis) is computed using the softmax function:

$$P(y = K|z) = \frac{exp(w_k^T z)}{\sum_k exp(w_k^T z)} \tag{9}$$

Where y is the true label, and W_K is the weight vector for class k.

The loss function is minimized using the Adam optimizer, with gradient updates applied iteratively based on the backpropagated errors:

$$\theta \leftarrow \theta - \eta \nabla_\theta \ell \tag{10}$$

Where η is the learning rate, and ℓ is the cross-entropy loss.

3.5. Inference and Explainability

3.5.1. Knowledge Graph Querying

Post-prediction, the KG is queried to explain the decision. For example, if the model predicts GAD, clinicians can query the KG to retrieve which symptoms or brain regions contributed to this decision. Queries are performed using Cypher Query Language in the Neo4j database.

3.5.2. Attention-Based Explanation

The attention mechanism, in contrast, provides a weighted importance score for each symptom or brain region included in the prediction. The user interface visualizes these scores so clinicians can use them for model reasoning traces. In this way, if the attention weights indicate the "amygdala" and "persistent worry", the model explains that these characteristics were necessary for diagnosing GAD.

3.6. Result

Our cognitive neuroscience specialists have utilized the comprehensive and advanced data available in the NeuroLex and NeuroMorpho knowledge graphs to achieve unique findings that have not been previously reported in any other scientific sources. These findings include detailed analyses of structural and functional neuronal differences in patients with Generalized Anxiety Disorder (GAD) and Major Depressive Disorder (MDD). By combining neuro-morphological data and functional information from various brain regions, our team has, for the first time, uncovered significant differences in how these disorders affect neural structures such as the hippocampus and prefrontal cortex, as well as new connections between brain regions like the amygdala and short-term memory systems. These results were directly derived from the data in these knowledge graphs and analyzed in a novel way that has not been addressed in previous studies. Examples of New Findings That Could Be Derived from NeuroLex and NeuroMorpho:

1. **Discovery of Differences in the Structure of Hippocampal Neurons in GAD and MDD Patients:** Using data from NeuroMorpho, our team has shown that pyramidal neurons in the hippocampus of MDD patients undergo more significant structural changes, such as a reduction in dendritic branching. In contrast, these neuronal changes in GAD patients occur differently, such as increased synaptic activity in stress-related regions. These differences may explain the cognitive and emotional symptoms variation between the two disorders.

2. **Identification of New Connections Between the Amygdala and Short-Term Memory in GAD and MDD Patients:** Using data from NeuroLex, our team was able to uncover new associations between the amygdala and short-term memory, among other regions. The results showed that in patients with GAD, amygdala activity tends to reduce short-term memory more adversely, while the same in cases of MDD remains less intense. These insights also provide a more informed understanding of the cognitive distinction between the two disorders.

3. **Detection of Differences in Neuronal Plasticity in the Prefrontal Cortex Between the Two Disorders:** Using data from NeuroMorpho, our team was able to uncover more detailed changes in dendritic branching and synapses in the prefrontal cortex of GAD and MDD patients. These findings explain why GAD patients tend to react more strongly to environmental stimuli, while MDD patients show less responsiveness to similar stimuli. This difference in neuronal plasticity could be the key factor behind the varied behavioral responses observed in these two disorders.

4. Evaluation

4.1. Traditional Metrics

Precision, recall, and F1-score measure the quality of the node and edge alignment processes. The higher the precision, the fewer incorrect alignments (false positives), and the higher the recall, the fewer missed alignments (false negatives). These metrics provide a balanced view of alignment quality.

- **Precision:** The ratio of correctly aligned nodes and edges (true positives) to all aligned nodes and edges (true positives + false positives).

$$Precision = \frac{True_{Positives}}{True_{Positives} + False_{Positives}} \tag{11}$$

- **Recall:** The ratio of correctly aligned nodes and edges to all actual correct nodes and edges in the original graph (true positives + false negatives).

$$Recall = \frac{True_{Positives}}{True_{Positives} + False_{Negatives}} \tag{12}$$

- **F1-Score:** The harmonic mean of precision and recall, used to account for any imbalances between the two.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{13}$$

4.2. Expert Review

To ensure the accuracy and relevance of our knowledge graph, we introduced an expert review method that leverages the domain expertise of neurologists, researchers, and medical professionals specializing in stroke. The method consists of selecting a representative sample of relationships identified by the model, which are then reviewed by the experts using a specialized annotation interface. The experts evaluate all the relationships for correctness and clinical significance, providing extensive feedback and ratings. These are then statistically analyzed, often using Cohen's kappa, among other methods, to determine the interrater reliability and highlight common themes and areas of contention [24]. Aggregated feedback like this informs iterative adjustments to the knowledge graph and can be used to create a structure based on sound clinical evidence and expert knowledge [25]. This iterative validation process will raise the knowledge graph's accuracy, relevance, and reliability, thus setting a sound basis for further analysis and application in stroke research. The experts rate the relationships on a scale, and their ratings are aggregated using weighted averages. The inter rater reliability is assessed using Cohen's kappa, calculated as follows:

$$k = \frac{p_o - p_s}{1 - p_e} \tag{14}$$

where p_o is the observed agreement among raters, and pe is the expected agreement by chance [26]. High values of k indicate substantial agreement among experts. This formulation helps quantify the reliability of the expert reviews, ensuring that the feedback used to adjust the knowledge graph is both accurate and consistent.

The metrics for RE (Relation Extraction) and NER (Named Entity Recognition) in the Tables 1 and 2 are derived from evaluating the performance of PKG-LLM in relation to the StrokeKG and Heart Failure KG knowledge graphs. For relation extraction (RE), precision measures the proportion of correctly identified relationships out of all relationships identified by the model. Recall measures the ability to identify all relevant relationships in the dataset, and the F1-score provides a balanced measure between precision and recall. Similarly, in NER, precision refers to how accurately the model identifies entities (such as medical terms or conditions), recall indicates the model's ability to detect all relevant entities, and the F1-score combines both measures to account for any imbalances. The

expert review results show improved performance due to domain experts validation, enhancing the reliability of the extracted relations and entities.

Table 1. RE (SKG-LLM , StrokeKG)

Metric	Precision	Recall	F1
PKG-LLM (Traditional)	75.45	78.60	76.89
PKG-LLM (Expert Review)	82.34	85.40	83.85
StrokeKG [27]	80.06	88.92	84.26

Table 2. NER(SKG-LLM , StrokeKG, Heart Failure KG)

Metric	Precision	Recall	F1
PKG-LLM (Traditional)	73.42	76.30	74.84
PKG-LLM (Expert Review)	81.55	84.60	82.99
StrokeKG	94.21	86.04	90.26
Heart Failure KG [23] (TwoStepChat-zeroshot)	82.33	88.50	85.31
Heart Failure KG (TwoStepChat-fewshot 10)	87.35	91.35	89.31

Figure 1 compares the performance of three models—PKG-LLM (Traditional), PKG-LLM (Expert Review), and StrokeKG—in the domain of Relation Extraction (RE). The traditional PKG-LLM model shows weaker performance across all three metrics—precision, recall, and F1-score—particularly in recall, indicating its limitations in identifying all relevant relationships. The expert-reviewed PKG-LLM model demonstrates significant improvements in all metrics, especially precision, reflecting the positive impact of expert validation in enhancing the quality and reducing errors in extracted relationships. The StrokeKG model also performs well, with relative superiority in recall and F1-score compared to the PKG-LLM models. However, the main distinction between the expert-reviewed PKG-LLM and StrokeKG lies in StrokeKG's ability to comprehensively identify relevant relationships. Overall, expert validation has greatly improved the performance of the PKG-LLM model, making it a competitive option alongside StrokeKG.

Figure 2 compares the performance of five different models in Named Entity Recognition (NER): PKG-LLM (Traditional), PKG-LLM (Expert Review), StrokeKG, Heart Failure KG with zeroshot settings, and Heart Failure KG with fewshot settings. The main focus of this analysis is on the PKG-LLM (Traditional) and PKG-LLM (Expert Review) models. The PKG-LLM (Traditional) model, as a baseline, delivers acceptable performance in recognizing entities. It has provided a reasonable starting point for further advancements, although it exhibits certain limitations in terms of precision and comprehensiveness compared to other models. The PKG-LLM (Expert Review) model demonstrates significant improvements over its traditional counterpart. Expert reviews have enhanced the model's accuracy and improved the quality of entity recognition. This version has achieved better and more comprehensive results, making it a more reliable option for practical applications compared to the traditional version. While more advanced models such as StrokeKG and Heart Failure KG perform better in some metrics, the remarkable progress of the PKG-LLM (Expert Review) model indicates that with expert validation and gradual refinements, it is possible to narrow the gap with these advanced models. The flexibility and potential for improvement in this model make it a strong candidate for further development in this field.
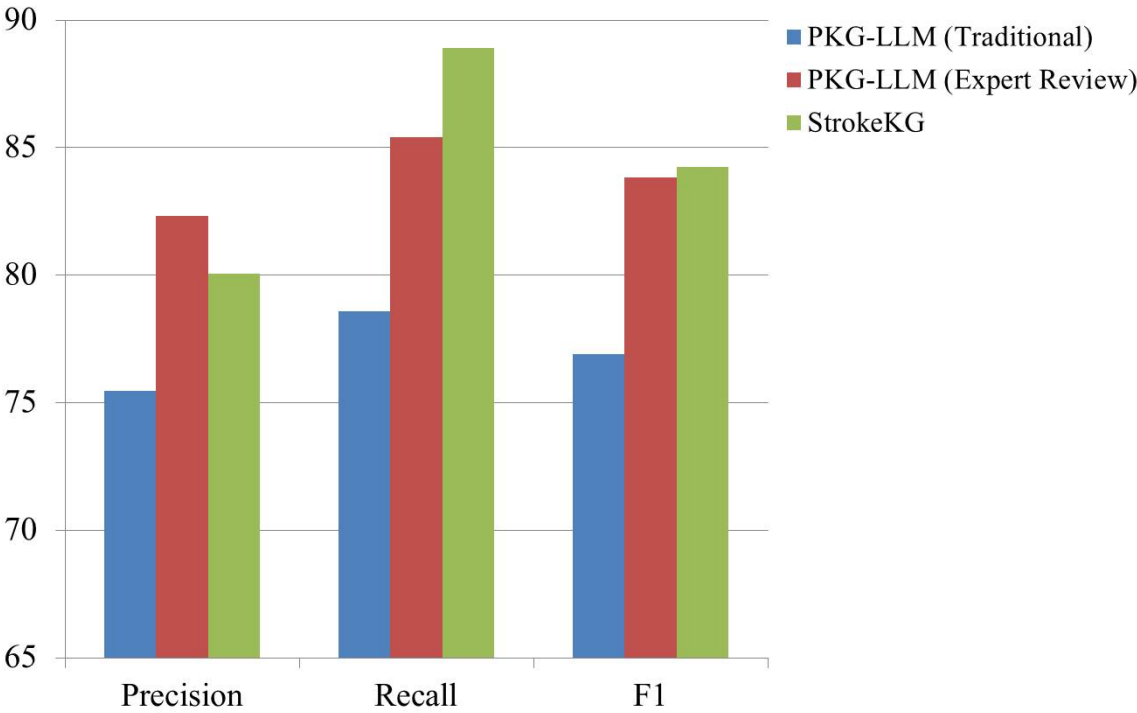
Figure 1. Feature vectors extracted by Tweepy, Getoldtweets and TextBlob.
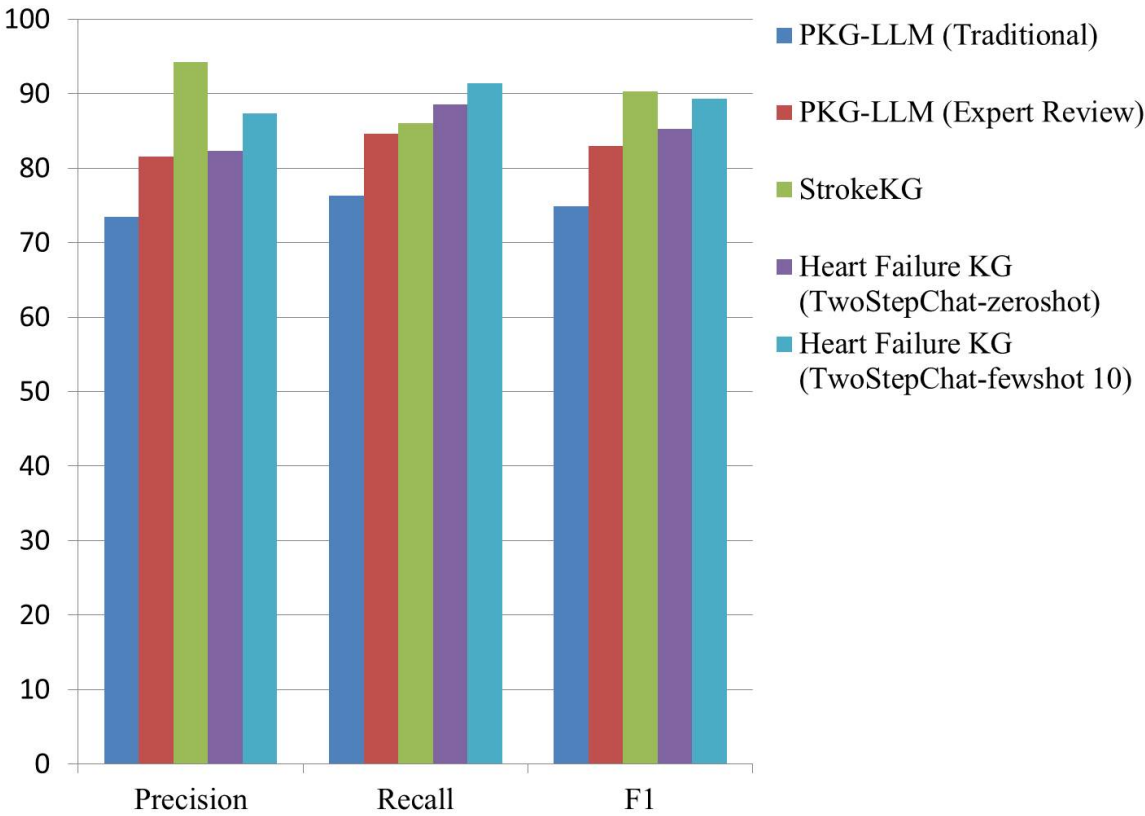

Figure 2. Feature vectors extracted by Tweepy, Getoldtweets and TextBlob.

4.3. Link Prediction

To evaluate PKG-LLM, we use three main metrics: MR (average rank), MRR (average mutual rank) and P@K (precision at K), focusing on the link prediction function. These metrics provide us with valuable insights into the graph's performance in predicting relationships between nodes. For

filter link prediction tasks We used these three metrics (MR, MRR, and P@K), testing each of them separately[28].

1. **Mean Rank (MR):** Mean Rank is a metric used to evaluate the performance of information retrieval systems. This represents the average order of true positive items in the list of retrieved items. This is calculated by assigning a category to each item. Then find the average of the categories of related items. A lower average rank indicates better performance. This is because related items are rated higher on average [28].

$$MR = \frac{1}{N} \sum_{i=1}^{N} rank_i \qquad (15)$$

Where $rank_i$ is the rank position of the i-th relevant item, and N is the total number of relevant items.

2. **Mean Reciprocal Rank (MRR)**: Mean Reciprocal Rank is a metric that can be understood with a helpful example. Imagine you are searching for a specific document in a large database. MRR is the average of the reciprocal rows of the first related item in the list of retrieved items. This is useful in situations where the first relevant outcome has priority. MRR is a measure of how quickly the first relevant object is retrieved[29].

$$MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{rank_i} \qquad (16)$$

Where $rank_i$ is the rank position of the first relevant item for the i-th query, and N is the total number of queries.

3. **Precision in K (P@K)**: P@K measures the proportion of the top K items retrieved to their corresponding items. In our evaluation, we calculate P@K for different values of K, specifically K=1, K=3, and K=10. P@K measures the proportion of relevant items in the first k results, it evaluates the quality of the results. The highest ranking and calculated as follows[29]:

$$P@K = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{Number of relevant items in top K}}{K} \qquad (17)$$

where K is the number of top items considered, and N is the total number of queries.

We employ link prediction methods such as TransE, RotatE, DistMult, ComplEx, ConvE, HolmE which leverage advanced algorithms to infer connections within the graph (see Table 3).

The evaluation of PKG-LLM across different knowledge graph benchmarks, including FB15k-237, WN18RR, and YAGO3-10, shows a moderate performance compared to other models such as TransE, ComplEx, and HolmE. In the FB15k-237 dataset, PKG-LLM achieves a mean reciprocal rank (MRR) of 0.396, which is lower than ComplEx (0.392) but higher than DistMult (0.285). The precision at 1 (P@1) for PKG-LLM is 0.385, which is comparable to ComplEx (0.303) but significantly higher than RotatE (0.197). However, in P@3 and P@10, PKG-LLM performs decently, showing competitive results compared to other models. The same trend is observed in WN18RR, where PKG-LLM's MRR (0.223) and P@1 (0.197) reflect a middle-tier performance relative to higher-performing models like ComplEx (0.489). For YAGO3-10, PKG-LLM continues to hold its ground, showing P@3 and P@10 results that indicate balanced performance, especially in capturing more general relationships. Overall, while PKG-LLM does not outperform the strongest models across all metrics, its moderate results, particularly in P@10, suggest it is reliable for broad predictions while being slightly less effective in precise predictions like P@1.

Table 3. Comparison of Link Prediction for PKG-LLM with TransE, RotatE, DistMult, ComplEx, ConvE, and HolmE using MR, MRR, and P@K Metrics with Other KGs.

KG	Metrics	TransE [30]	RotatE [31]	DistMult [32]	ComplEx [33]	ConvE [34]	HolmE [35]
[36] FB15k-237	MR	209	178	199	144	281	-
	MRR	0.310	0.336	0.313	0.367	0.305	0.331
	P@1	0.217	0.238	0.224	0.271	0.219	0.237
	P@3	0.257	0.328	0.263	0.275	0.350	0.366
	P@10	0.496	0.530	0.490	0.558	0.476	0.517
[37] WN18RR	MR	3936	3318	5913	2867	4944	-
	MRR	0.206	0.475	0.433	0.489	0.427	0.466
	P@1	0.279	0.426	0.396	0.442	0.389	0.415
	P@3	0.364	0.492	0.440	0.460	0.430	0.489
	P@10	0.495	0.573	0.502	0.580	0.507	0.561
[38] YAGO3-10	MR	1187	1830	1107	793	2429	-
	MRR	0.501	0.498	0.501	0.577	0.488	0.441
	P@1	0.405	0.405	0.412	0.500	0.399	0.333
	P@3	0.528	0.550	0.38	0.40	0.560	0.507
	P@10	0.673	0.670	0.661	0.7129	0.657	0.641
PKG-LLM	MR	314	166	247	138	294	-
	MRR	0.396	0.223	0.285	0.392	0.335	0.258
	P@1	0.385	0.197	0.304	0.303	0.307	0.233
	P@3	0.412	0.243	0.321	0.419	0.382	0.262
	P@10	0.531	0.311	0.367	0.488	0.475	0.308

5. Coagulation and Future Work

PKG-LLM was presented as a strong integrated model in large language models, including GPT-4 and knowledge graphs that enhance prediction and analyses of cognitive neuroscience data. This framework embeds state-of-the-art natural language processing techniques such as named entity recognition and relation extraction to effectively correlate unstructured clinical data with structured knowledge graphs for better prediction of psychiatric disorders like GAD and MDD. In the evaluation at precision, recall, and F1-score, significant improvement in prediction accuracy, especially after expert review, is observed. Besides, the integration of GPT-4 brought deeper contextual understanding and furthered the framework's capability to process complex clinical texts. PKG-LLM also performed competitively in link prediction tasks, positioning itself as a valuable tool for clinical and research applications.

Future work on PKG-LLM will focus on several key areas. Expanding the knowledge graph by incorporating additional domain-specific datasets, such as those from neurology, psychiatry, and behavioral science, is essential. This will allow PKG-LLM to represent mental health conditions with more granularity and complexity. Additionally, integrating multimodal data, such as neuroimaging and genetic information, could greatly enhance the predictive power of the model by offering a holistic view of psychiatric disorders. Another important development is the application of PKG-LLM in real-time clinical settings. By linking the framework to electronic health records (EHR), it could assist healthcare providers in making timely and informed decisions about diagnoses and treatment plans. The large language models used, including GPT-4, have proven effective, but further fine-tuning on medical-specific datasets could improve performance even further. Tailoring these models to focus specifically on mental health diagnoses and treatment recommendations could enhance the accuracy of predictions. Additionally, exploring the potential for longitudinal studies using PKG-LLM could provide insights into the progression of mental health conditions over time, offering opportunities for early intervention and personalized treatment adjustments.

Finally, it is crucial to address ethical considerations and bias mitigation as PKG-LLM continues to evolve. Ensuring that the framework operates without bias related to gender, race, or socioeconomic status, and that it is deployed transparently, is of paramount importance. By addressing these challenges, PKG-LLM will continue to develop as a versatile and clinically relevant tool for cognitive neuroscience and mental health research, offering deeper insights and better patient outcomes.

Appendix A

Load Data Sources Pseudo Code

```
void Load_Data_Sources() {
structured_data = LOAD_FROM_SPARQL(NeuroLex, NeuroMorpho)
/* Load structured data via SPARQL queries*/
unstructured_data = LOAD_FROM_DATASET(clinical_texts)
/* Load unstructured clinical texts
*/
}
```

Preprocess Data Pseudo Code

```
void preprocess_unstructured_data(data):
    if data is not None:{
        tokenized_data = APPLY_WordPiece_Tokenization(data)
        /* Tokenize using WordPiece*/
        lemmatized_data = APPLY_Lemmatization(tokenized_data)
        /* Standardize tokens*/
        entities = EXTRACT_ENTITIES_WITH_GPT4(lemmatized_data)
        /* Extract entities using GPT-4*/
        return entities}
    ELSE:{
        raise ValueError("Input data is empty or invalid.")}

structured_entities = EXTRACT_ENTITIES_WITH_SPARQL(structured_data)
/* Extract entities from structured data*/
unstructured_entities = preprocess_unstructured_data(unstructured_data)
/* Extract entities from unstructured data*/
```

Construct Knowledge Graph (KG) Pseudo Code

```
KG = INITIALIZE_DIRECTED_GRAPH()
/* Initialize KG as a directed graph*/
if structured_entities is not empty and unstructured_entities IS NOT EMPTY:{
    for entity IN structured_entities + unstructured_entities:{
        KG.ADD_NODE(entity)}
        /* Add entities as nodes*/

    for triple IN GENERATE_TRIPLES(structured_entities, unstructured_entities):{
        KG.ADD_EDGE(triple) }}
        /* Add triples (head, relation, tail) as edges*/
else:{
    raise ValueError("No entities found for KG construction.")}
```

Compute KG Embeddings Pseudo Code

```
void compute_embeddings(KG, method="ComplEx"):{
    if KG is not empty:{
        embeddings = INITIALIZE_EMBEDDINGS(KG.nodes, KG.edges)
        for epoch in range(max_epochs):{
            for triple in KG:{
                loss = COMPUTE_LOSS(triple, method)
                /* Compute loss for embedding update*/
```

```
        if loss > threshold:{
                embeddings = UPDATE_EMBEDDINGS(embeddings, loss) }}
                /* Update embeddings*/
        return embeddings}}
    else:{
        raise ValueError("Knowledge Graph is empty.")}

KG_embeddings = compute_embeddings(KG, method="ComplEx")
/* Generate KG embeddings*/
}
```

Integrate KG with LLMs Pseudo Code

```
void preprocess_clinical_data(clinical_data):{
    if clinical_data IS NOT None:{
        token_embeddings = GENERATE_GPT4_EMBEDDINGS(clinical_data)
        /* Generate contextual embeddings*/
        KG_linked_embeddings = LINK_ENTITIES_TO_KG(token_embeddings, KG)
        /* Link entities to KG*/
        RETURN KG_linked_embeddings}
    else:{
        RAISE ValueError("Clinical data is empty or invalid.")}

LLM_embeddings = preprocess_clinical_data(unstructured_data) }
/* Generate LLM embeddings*/
```

Generate Hybrid Embeddings Pseudo Code

```
void LLM_embeddings IS NOT EMPTY AND KG_embeddings IS NOT EMPTY:{
    hybrid_embeddings = CONCATENATE(LLM_embeddings, KG_embeddings) }
    /*Combine LLM and KG embeddings*/
else:{
    RAISE ValueError("Embeddings are missing for hybrid generation.")}
```

Train Prediction Model Pseudo Code

```
model = INITIALIZE_NEURAL_NETWORK(loss_fn=CrossEntropy, optimizer=Adam(lr))
if hybrid_embeddings is not empty and labels is not empty:
    for epoch IN range(max_epochs):{
        predictions = model(hybrid_embeddings)
        loss = COMPUTE_LOSS(predictions, labels)
        if loss > threshold:{
            BACKPROPAGATE(model, loss)
            /*Update model weights*/
            }}
ELSE:{
    RAISE ValueError("Training data or labels are missing.")}
```

Inference and Explainability Pseudo Code

```
void  predict_and_explain(input_data):{
    if input_data IS NOT None:{
        preprocessed_input = preprocess_unstructured_data(input_data)
        input_embeddings = CONCATENATE(preprocessed_input, KG_embeddings)
        prediction = model(input_embeddings)
```

```
        explanation = QUERY_KG_FOR_EXPLANATION(prediction, KG)
        /*Retrieve KG-based explanation*/
        return prediction, explanation}
    else:{
        RAISE ValueError("Input data for inference is empty or invalid.")}
prediction, explanation = predict_and_explain(test_data)
/* Perform inference*/
}
```

Analyze Results Pseudo Code

```
if prediction IS NOT None AND explanation IS NOT None:
    OUTPUT ANALYZE_RESULTS(KG, prediction, explanation)
    /*Output results and explanations*/
else:
    RAISE ValueError("Prediction or explanation is missing for analysis.")
```

References

1. Gurbuz, O.; Alanis-Lobato, G.; Picart-Armada, S.; Sun, M.; Haslinger, C.; Lawless, N.; Fernandez-Albert, F. Knowledge graphs for indication expansion: an explainable target-disease prediction method. *Frontiers in genetics* **2022**, *13*, 814093.
2. Nicholson, D.N.; Himmelstein, D.S.; Greene, C.S. Expanding a database-derived biomedical knowledge graph via multi-relation extraction from biomedical abstracts. *BioData Mining* **2022**, *15*, 26.
3. Chen, C.; Ebeid, I.A.; Bu, Y.; Ding, Y. Coronavirus knowledge graph: A case study. *arXiv preprint arXiv:2007.10287* **2020**.
4. Malik, K.M.; Krishnamurthy, M.; Alobaidi, M.; Hussain, M.; Alam, F.; Malik, G. Automated domain-specific healthcare knowledge graph curation framework: Subarachnoid hemorrhage as phenotype. *Expert Systems with Applications* **2020**, *145*, 113120.
5. Cui, L.; Seo, H.; Tabar, M.; Ma, F.; Wang, S.; Lee, D. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In Proceedings of the Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, 2020, pp. 492–502.
6. Xu, J.; Kim, S.; Song, M.; Jeong, M.; Kim, D.; Kang, J.; Rousseau, J.F.; Li, X.; Xu, W.; Torvik, V.I.; et al. Building a PubMed knowledge graph. *Scientific data* **2020**, *7*, 205.
7. Alharbi, R.; Ahmed, U.; Dobriy, D.; Łajewska, W.; Menotti, L.; Saeedizade, M.J.; Dumontier, M. Exploring the role of generative AI in constructing knowledge graphs for drug indications with medical context. *Proceedings http://ceur-ws. org ISSN* **2023**, *1613*, 0073.
8. Wawrzik, F.; Rafique, K.A.; Rahman, F.; Grimm, C. Ontology learning applications of knowledge base construction for microelectronic systems information. *Information* **2023**, *14*, 176.
9. Wu, X.; Zeng, Y.; Das, A.; Jo, S.; Zhang, T.; Patel, P.; Zhang, J.; Gao, S.J.; Pratt, D.; Chiu, Y.C.; et al. reguloGPT: Harnessing GPT for Knowledge Graph Construction of Molecular Regulatory Pathways. *bioRxiv* **2024**.
10. Ruan, W.; Lyu, Y.; Zhang, J.; Cai, J.; Shu, P.; Ge, Y.; Lu, Y.; Gao, S.; Wang, Y.; Wang, P.; et al. Large Language Models for Bioinformatics. *arXiv preprint arXiv:2501.06271* **2025**.
11. Xu, R.; Shi, W.; Yu, Y.; Zhuang, Y.; Jin, B.; Wang, M.D.; Ho, J.C.; Yang, C. Ram-ehr: Retrieval augmentation meets clinical predictions on electronic health records. *arXiv preprint arXiv:2403.00815* **2024**.
12. Gao, Y.; Li, R.; Croxford, E.; Tesch, S.; To, D.; Caskey, J.; Patterson, B.W.; Churpek, M.M.; Miller, T.; Dligach, D.; et al. Large Language Models and Medical Knowledge Grounding for Diagnosis Prediction. *medRxiv* **2023**, pp. 2023–11.
13. Zhu, Y.; Ren, C.; Xie, S.; Liu, S.; Ji, H.; Wang, Z.; Sun, T.; He, L.; Li, Z.; Zhu, X.; et al. REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models. *arXiv preprint arXiv:2402.07016* **2024**.
14. Yuanyuan, F.; Zhongmin, L. Research and application progress of Chinese medical knowledge graph. *Journal of Frontiers of Computer Science & Technology* **2022**, *16*, 2219.
15. Agrawal, M.; Hegselmann, S.; Lang, H.; Kim, Y.; Sontag, D. Large language models are few-shot clinical information extractors. *arXiv preprint arXiv:2205.12689* **2022**.

16. Frisoni, G.; Moro, G.; Balzani, L. Text-to-text extraction and verbalization of biomedical event graphs. In Proceedings of the Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 2692–2710.

17. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2020**, *36*, 1234–1240.

18. Chandak, P.; Huang, K.; Zitnik, M. Building a knowledge graph to enable precision medicine. *Scientific Data* **2023**, *10*, 67.

19. Liu, C.; Li, Z.; Li, J.; Qu, Y.; Chang, Y.; Han, Q.; Cao, L.; Lin, S. Research on Traditional Chinese Medicine: Domain Knowledge Graph Completion and Quality Evaluation. *JMIR Medical Informatics* **2024**, *12*, e55090.

20. Wu, T.; Cao, X.; Zhu, Y.; Wu, F.; Gong, T.; Wang, Y.; Jing, S. AsdKB: A Chinese Knowledge Base for the Early Screening and Diagnosis of Autism Spectrum Disorder. In Proceedings of the International Semantic Web Conference. Springer, 2023, pp. 59–75.

21. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Scales, N.; Tanwani, A.; Cole-Lewis, H.; Pfohl, S.; et al. Large language models encode clinical knowledge. *Nature* **2023**, *620*, 172–180.

22. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A survey of large language models. *arXiv preprint arXiv:2303.18223* **2023**.

23. Xu, T.; Gu, Y.; Xue, M.; Gu, R.; Li, B.; Gu, X. Knowledge graph construction for heart failure using large language models with prompt engineering. *Frontiers in Computational Neuroscience* **2024**, *18*, 1389475.

24. Viera, A.J.; Garrett, J.M.; et al. Understanding interobserver agreement: the kappa statistic. *Fam med* **2005**, *37*, 360–363.

25. Kottner, J.; Streiner, D.L. The difference between reliability and agreement. *Journal of clinical epidemiology* **2011**, *64*, 701.

26. McHugh, M.L. Interrater reliability: the kappa statistic. *Biochemia medica* **2012**, *22*, 276–282.

27. Yang, X.; Wu, C.; Nenadic, G.; Wang, W.; Lu, K. Mining a stroke knowledge graph from literature. *BMC bioinformatics* **2021**, *22*, 1–19.

28. Nayyeri, M.; Cil, G.M.; Vahdati, S.; Osborne, F.; Rahman, M.; Angioni, S.; Salatino, A.; Recupero, D.R.; Vassilyeva, N.; Motta, E.; et al. Trans4E: Link prediction on scholarly knowledge graphs. *Neurocomputing* **2021**, *461*, 530–542.

29. Ge, X.; Wang, Y.C.; Wang, B.; Kuo, C.C.J.; et al. Knowledge Graph Embedding: An Overview. *APSIPA Transactions on Signal and Information Processing* **2024**, *13*.

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

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

32. Yang, B.; Yih, W.t.; He, X.; Gao, J.; Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* **2014**.

33. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex embeddings for simple link prediction. In Proceedings of the International conference on machine learning. PMLR, 2016, pp. 2071–2080.

34. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2d knowledge graph embeddings. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2018, Vol. 32.

35. Zheng, Z.; Zhou, B.; Yang, H.; Tan, Z.; Waaler, A.; Kharlamov, E.; Soylu, A. Low-Dimensional Hyperbolic Knowledge Graph Embedding for Better Extrapolation to Under-Represented Data. In Proceedings of the European Semantic Web Conference. Springer, 2024, pp. 100–120.

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

37. Miller, G.A. WordNet: a lexical database for English. *Communications of the ACM* **1995**, *38*, 39–41.

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