An Architecture for Ontology-based Semantic Reasoning Using LLMs in Healthcare Domain

Müge Oluçoğlu
Department of Computer Engineering
Izmir Bakircay University
Izmir, Turkey
muge.olucoglu@bakircay.edu.tr

Okan Bursa
Department of Computer Engineering
Izmir Bakircay University
Izmir, Turkey
okan.bursa@bakircay.edu.tr

***Abstract*—This study presents research that involves examining various methods used in the field of health services and proposing a new architecture for more precise diagnosis from health records. Traditional and modern methods such as Electronic Health Records (EHR), Clinical Decision Support Systems (CDSS), Natural Language Processing (NLP)-based analytics, and Machine Learning (ML) techniques are discussed, highlighting their advantages, disadvantages, and usage areas. Based on these evaluations, a new method involving ontologies and Large Language Models (LLMs) has been developed to provide a more effective solution for healthcare informatics. The proposed approach is a candidate solution to achieve higher accuracy, speed, and flexibility by integrating ontologies, LLM and reasoners.**

Keywords-semantic reasoning; ontology; healthcare; knowledge graph; large language model.

I. INTRODUCTION

Health informatics is currently undergoing a significant transformation with the integration of big data and artificial intelligence technologies. At the center of this transformation are Large Language Models (LLMs) and Knowledge Graphs (KGs). LLMs have made groundbreaking advances in Natural Language Processing (NLP) techniques and can extract meaningful and contextual information from large datasets [1]. KGs, on the other hand, improve knowledge integration and extraction processes by representing the relationships between data at a semantic level.

Healthcare is an emerging field where LLMs can provide significant advantages in data analysis and interpretation. These models can extract meaningful information from large and diverse data sources such as electronic health records, clinical notes and medical literature, and provide valuable insights to healthcare professionals in decision support systems [2]. However, the complexity and multidimensionality of health data require the use of advanced techniques to interpret and process these data accurately and efficiently. Knowledge graphs play an important role in addressing this complexity.

Knowledge graphs are structured data models used for semantic modeling and identifying relationships between data elements. In healthcare, knowledge graphs integrate different data sources such as patient data, genetic information, medical literature and clinical trials, and establish semantic links between them [3]. This integration enables healthcare professionals and researchers to better understand and analyze complex data relationships. On the other hand, this integrated representation of the knowledge and data makes the processing of the domain knowledge and claims of the rules of the domain from the data more efficient. This extraction is crucial to submerge the existing knowledge and rules from the data to provide more precise definitions for the healthcare domain.

The relationship between LLMs and KGs has great potential for the processing and analysis of health data. The aim is to create models that improve healthcare outcomes, obtain personalized and accurate findings, and support human decision-making processes [4][5]. When the NLP capabilities of LLMs are combined with knowledge graphs, it becomes possible to analyze and extract the semantics of health data more effectively. This integration enables the development of knowledge-based decision support systems in healthcare and potential to provide more accurate results in clinical decision processes [6].

In this regard, the aim of this study is to investigate semantic reasoning processes in the healthcare domain using LLMs and KGs. Existing approaches in the literature for semantic modeling, integration and analysis of health data will be summarized to create an architecture to merge not only knowledge graphs but also ontologies with the LLM to obtain a framework to build semantic-aware LLM applications. Furthermore, a new reasoning structure is introduced with the usage of Semantic Web Rule Language (SWRL) [23] for domain knowledge and Simple Protocol and Resource description framework Query Language [24] (SPARQL) to query the inferred entities.

In the following sections of the study, the literature review will be described in Section 2, and the methods, databases and ontologies currently used for selected studies in different healthcare fields will be explained in Section 3. In Section 4, the difference between the architecture created and the existing studies will be discussed. We conclude the article in Section 5.

II. LITERATURE REVIEW

In this part of the article, as a result of the literature review, studies on different health fields carried out using LLMs and KGs will be explained. Since current content was desired to be included, studies between 2020 and 2024 were in the literature review. The research summarized in these papers focuses on leveraging various artificial intelligence and semantic technologies to enhance healthcare predictions, diagnosis, and management, particularly in the domains of chronic diseases, mental health disorders, and Alzheimer's disease.

The study from [7] introduces dynamic and adaptive approaches, such as the dynamic fuzzy rule-based inference system for Alzheimer's disease diagnosis and the ostensive information architecture for enhancing semantic interoperability in healthcare information systems. The authors describe a real-world case study utilizing an Alzheimer's Disease (AD) diagnosis system that is built on fuzzy, dynamic, and semantic decision criteria. The study's dataset was compiled from medical records of patients in the USA and Canada. Semantic data, including genetics, screening findings, treatments, etc. from patient diagnostic results is used to develop a recommended system and do a comparative study. Machine learning and ontology-based systems were compared using the same data.

To handle uncertainty in medical information, particularly in relation to the diagnosis of mental health issues, the combination of fuzzy logic and ontologies is advocated. The complexity and uncertainty associated with mental health illnesses are addressed in this study by proposing a machine learning model that integrates fuzzy logic and ontology with Mamdani inference in Fuzzy Ontology Web Language (OWL) for Protegee for the diagnosis of Major Depressive Disorder (MDD) [8]. Nine input variables—such as mood, sleep, hunger and weight, joy and pleasure, fatigue, guilt, memory, and psychomotor impairment—are linked to symptoms of MDD. To account for the uncertainty in characterizing symptoms, each of these variables is mapped to Type-2 fuzzy sets with trapezoidal membership functions. Based on input symptoms, the suggested fuzzy inference system applies a set of rules to determine the degree of depression.

Many studies integrate ontologies, knowledge graphs, and semantic reasoning to improve the accuracy of clinical prediction models using Electronic Health Record (EHR) data. In another study [9] that integrates semantic reasoning using ontology-based decision support and recommendation systems for diabetes nutrition treatment and diagnosis, fuzzy medical rules are used. The study in [10], on the other hand, offers a nutrition recommendation system. The authors of [11] have developed a system using the symptoms and complaints of diabetic patients in Fast Health Interoperability Resources (FHIR). In another study, the authors created a prediction system with personal health information in the EHR system [12].

Recent developments in LLMs are increasing interest in their potential applications in medicine. Studies evaluate the performance of LLMs on a variety of health-related tasks, including clinical language understanding, new drug discovery, and health prediction, using wearable sensor data. Challenges such as hallucinations and the need for task-specific orientation strategies are also addressed. [13] developed an LLM-based query-answer system for drug discovery in cancer research. It validates gene-disease associations using machine learning techniques to analyze multimodal data [14]. [15] performs semantic reasoning using past patient records. The system proposed by the authors asks and receives informative questions and answers regarding the clinical scenarios at hand. The authors used National Center for Biotechnology Information (NCBI-Disease), BC5CDR-Chemical, i2b2 2010-Relation, SemEval 2013-DDI, BIOSSES, MedNLI, i2b2 2006-Smoking, BioASQ 10b-Factoid, BioASQ 10b-Factoid databases. LLMs have also been used in medical decision-making. The authors of [16] used decision trees to produce more reliable medical answers. The data attributes used by the authors of [17] are user demographics, health information, stress, readiness, fatigue, activity, calories, sleep quality, sleep disorder, anxiety, and depression. This variety of patient data attributes in [17] enables LLMs to make inferences about health based on semantic reasoning.

In addition to all these benefits that LLMs provide in the field of healthcare, situations that may cause data security, ethical handling, etc. and many other problems are discussed in [18] and [19]. It is mentioned how and what solutions will be produced against these problems.

The proposed architectural solution leverages ontologies to reason over the collected data, while also enabling queries with well-defined parameters. This approach has the potential to identify irregularities in the LLM's responses, leading to more accurate and precise answers for the patient over health records.

III. METHODOLOGY

In this section, existing methodologies in different healthcare fields will be discussed in detail.

In this part of the study, there are many different studies conducted in different fields of health services in the literature. The methods and ontological languages used in these studies are explained in more detail in Section 2.

In [7], dynamic and adaptive methodologies are introduced, such as a complete information architecture to improve semantic interoperability in health information systems and a dynamic fuzzy rule-based inference system for Alzheimer's disease detection. The dataset for the research is composed of medical records from patients in the USA and Canada, derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the World Health Organization (WHO). After patient information was collected, data preprocessing steps were used. In the Semantic Reasoning stage, neuropsychological cognitive data, Cerebrospinal Fluid (CSF) data, MRI-PET, physical examination, demographic data, and genetic data obtained from patients are divided into numerical and categorical classes. These classes are later accepted as semantic features for inference. Thus, semantic reasoning is derived by performing ontological reasoning. The results obtained here are made from fuzzy reasoning by applying fuzzification. This system, created by processing semantic data such as symptoms, allows early diagnosis of Alzheimer's disease.

In [8], a study was conducted for the diagnosis of MDD, which is complex and unclear, related to mental health and illness. An anonymous medical dataset of 90 people with depression between the ages of 25 and 65, selected from a mental health center in Iran, was used. 9 parameters were selected for diagnosis in line with the standards recognized by The Diagnostic and Statistical Manual of Mental Disorders (DSM). These parameters; (1) mood, (2) sleep, (3) change in appetite & weight, (4) joy & pleasure, (5) feeling tired, (6) feeling guilty, (7) memory, (8) the dominant force in suicide and (9) psychomotor disturbance. After the fuzzifier step is applied to the findings obtained, the knowledge base and rules base steps are taken for definitive inference. Defuzzification is applied to determine the severity of depression. The result of the decision is given to the application user and doctor. There are 4 modules in the recommendation system for the diagnosis phase of the study. Module 1 is the patient's electronic record repository, which contains the patient's history of health problems and personal information such as name, gender, and age. Module 2 is an ontological database where all variables of the patient's medical care (Dopamine, Cortisol, Growth Hormone, Norepinephrine, Thyroid) and psychological symptoms are semantically explained and stored. Module 3 is the implementation of the fuzzy inference system that allocates input variables to output variables based on rules and regulations defined in the ontology database. Module 4 is the control and decision-making layer.

In [9], a diagnostic and nutritional recommendation system for diabetes was designed. It creates a linguistic fuzzy rule base that integrates information from the EHR and domain experts with information obtained from training data and information extracted from the semantic model. The authors developed a fuzzy knowledge-based system on pre-processed data using machine learning algorithms. Mamdani inference engine was applied to the results of these pre-processing using fuzzy rules.

The authors of [11] have developed a system for detecting complaints and symptoms of diabetic patients. The authors started from the problems in implementing FHIR. A Semantic Engine with FHIR knowledge graph as the core was built using Neo4j to provide semantic interpretation and examples. The authors obtained the data in Medical Information Mart for Intensive Care (MIMIC) and diabetes datasets. In the architecture that had been created in [11], the components of the semantic engine are the FHIR knowledge graph and transformation components. FHIR knowledge graph consists of three layers. The first layer consists of local health information systems, which act as sources for collective health data. The second layer, known as the transformation layer, functions as an integrator between the first and third layers, utilizing a query processor, linker, and mapping connector to process the data. Once connected, the health data is transferred to the third layer, the semantic interpretation layer. This layer is responsible for interpreting the semantics by defining the lexical meaning of nodes and their relationships, as established in the second layer. As a result, the FHIR Knowledge Graph is constructed in the final layer.

In [13], the authors developed the Knowledge Graph Based Thinking (KGT) model by combining LLM's knowledge graphs to reduce real errors in reasoning. KGT uses LLMs to create the optimal subgraph based on important information extracted from the question.

TABLE I. LITERATURE REVIEW

Ref	***Year***	***Domain***	***Method***	***Ontology Language***	***Dataset***
[7]	2024	Alzheimer's disease diagnosis	Semantic features were created using ontological reasoning. Fuzzification was applied to the results	Fuzzy OWL	Alzheimer's Disease Neuroimaging Initiative (ADNI)
[8]	2023	Diagnosed with major depressive disorder	After the fuzzification process, knowledge base and rule base steps were used. Depression severity was clarified	Fuzzy OWL	Anonymous patients' records
[9]	2020	Treatment diagnosis for diabetes	Creating a linguistic fuzzy rule base for the information coming from the semantic model	Fuzzy OWL	Electronic Health Records (EHR)
[11]	2024	Local health information system diabetes	The components of the semantic engine consist of knowledge graph of the FHIR data and its transformation components	OWL	Medical Information Mart for Intensive Care (MIMIC)
[13]	2024	Drug discovery for cancer research	Creates the optimal subgraph using important information	-SynLethKG -SDKG -SOKG	--
[14]	2023	Gene-disease relationship	Cancer information is integrated into the ontology. Final fine-tuning with LLMs	Ontological rules	-OncoNet Ontology (ONO) -Scientific Article

It facilitates the discovery of new uses of existing drugs through drug-cancer relationships. The authors use the SmartQuerier Oncology Knowledge Graph (SOKG) ontology, a pan-cancer knowledge graph.

The Onco Net ontology (ONO) used in [14] is created with scientific articles and real information obtained from various sources. The obtained cancer-related information is integrated into the ontology. Final fine-tuning is being done with LLM's as it has a more comprehensive knowledge graph. Thus, it is aimed to confirm the gene-disease diagnosis relationship.

Table 1 provides information on current methods and practices used in healthcare in different fields for selected studies. It includes summary information about the year, method, ontology used, study area and databases of the studies. Studies in different healthcare fields, mostly published in 2023 and 2024, are included.

IV. ARCHITECTURE

In this chapter, we introduce an architecture for ontology-based semantic reasoning using LLMs in the healthcare domain. Various methods and ontologies used in health services have been examined to develop a robust framework. In previous studies, different datasets and analytical techniques were applied across diverse healthcare fields. For example, dynamic methodologies were developed for the early diagnosis of Alzheimer's using ADNI and WHO data, leveraging ontological reasoning. In another study on major depressive disorder (MDD) diagnosis, DSM parameters were applied to assess depression severity. For diabetes diagnosis and nutrition recommendation systems, a Mamdani inference engine utilizing fuzzy rules based on EHR data was implemented. A similar approach was used to identify diabetic symptoms through a semantic engine built with FHIR knowledge graph data. These efforts exemplify how reasoning models, such as the Reasoning-Based Thinking (KGT) model, integrate LLMs with knowledge graphs for enhanced diagnostic precision. The architecture presented here builds on these foundational works, aiming to offer a scalable and efficient solution for semantic reasoning in healthcare using ontologies and LLMs.

The architecture created in this study was designed as a system that determines possible treatment methods by performing semantic reasoning in the field of healthcare. The steps of the architecture created for the study are clearly seen in Figure 1.

The first step is to collect data from the patient. Patient data can be in text, image or video format. To process this multi-modal data, it is first tokenized into small pieces. On the other hand, the same data is also combined with the prompting process of a domain expert or a doctor defined with SWRL rules.

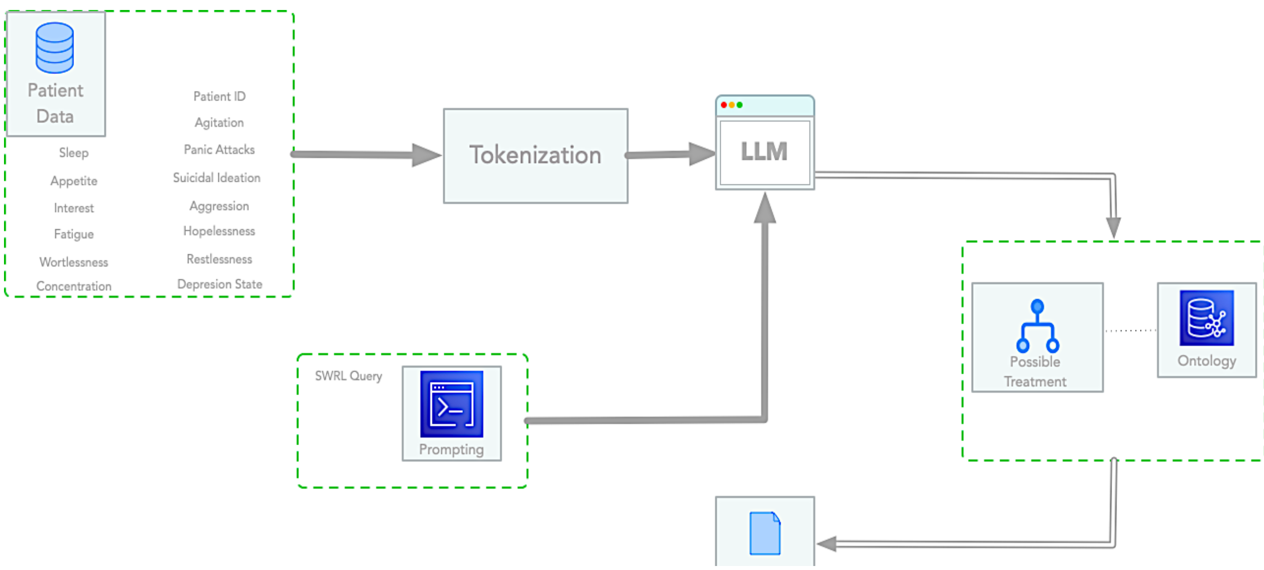

Figure 1. Proposed Model Architecture.

Figure 2 shows an example SWRL rule. It states that if a particular patient (person) uses certain substances and has certain symptoms (dry mouth and muscle tension), that patient should seek care from a doctor and receive psychotherapy treatment. The first step is to collect data from the patient. Patient data can be in text, image or video format. To process this multi-modal data, it is first tokenized into small pieces. On the other hand, the same data is also combined with the prompting process of a domain expert or a doctor defined with SWRL rules. SWRL is a rule definition language that allows us to create rules using entities in RDF datasets. These rules enable patient data to be made more meaningful and enriched with semantic features. Patient data and the data obtained by the prompting process are fed into an LLM. This model is trained with a broad knowledge base and language knowledge, allowing it to understand patient data more deeply.

```
PERSON(?PATIENT), MAKESUSE(?PATIENT,
?SUBSTANCE), DOCTOR(?DOCTOR),
PSYCHOTHERAPY(?TREATMENT),
SYMPTOMISASSOCIATEDWITH(?CASE, ?SYMPTOM),
DRYMOUTH(?SYMPTOM),
MUSCULARTENSION(?SYMPTOM) ->
TAKESCARE(?DOCTOR, ?PATIENT),
TREATS(?PATIENT, ?TREATMENT)
```

Figure 2. Example of SWRL.

Data analyzed by LLM is used to identify potential treatment methods. LLM evaluates patient data and applies SWRL rules to analyze tokenized healthcare information. Based on this evaluation, it generates and promotes the most appropriate treatment options tailored to the patient's specific needs. However, in order to further optimize the response, knowledge graphs are used to determine treatment methods. The response will be reconstructed within the knowledge graph to find the possible treatments are valid due to SWRL rules. In order to check the valid responses, knowledge graphs are queried similar to graphs to create a subgraph of the treatment response of the LLM. A valid response means that a sub-graph can be extracted, and this shows that the response is a known method in the healthcare domain.

The possible treatment methods identified by the LLM are determined as the final results to be presented to the user and the doctor. These results are presented to the user in an understandable format. It provides information about the details and applicability of treatment methods.

This architecture aims to make sense of patient data and determine the most appropriate treatment methods by performing semantic reasoning in the field of healthcare. The system consists of the steps of tokenizing patient data, enriching it with prompting and SWRL rules, analyzing it using LLM, and checking its validity with help of knowledge graphs.

The dataset chosen for this study is related to mental health. The dataset was obtained from Kaggle [20]. In the dataset, there are 13 personal attributes of information about the patients: frequency of sleep disturbance, change in appetite, loss of interest in activities, fatigue or low energy, feeling of worthlessness or guilt, difficulty concentrating, physical agitation, suicidal ideation, aggression, experiencing panic attacks. There is information about despair, restlessness and general depression. Except for the general depression state, the values of other parameters are between 1 and 6. These values are 1: Never, 2: Always, 3: Often, 4: Rarely, 5: Sometimes, 6: Not at all. Values for the General Depression State parameter are classified as: "No depression", "Mild", "Moderate", and "Severe".

The ontology constructed in the study [21] includes concepts and features used to risk classification in mental health. The study aims to determine the patient's risk level by scoring signs and symptoms. The ontology consists of 332 classes, 82 individuals and 37 properties.

V. CONCLUSION

This study examines the effectiveness of a new architecture built with Large Language Models (LLMs) using ontology-based semantic reasoning in the healthcare field. The mental health dataset and the ontology in this field were used as a sample study area. Dynamic and adaptive methodologies in the literature have been compared with systems used in the diagnosis and treatment of various health problems such as Alzheimer's disease diagnosis and major depressive disorder.

In the proposed architecture, patient data is collected and made analyzable by tokenization, and a prompting process enriched with SWRL rules is applied. This data is processed using LLM and potential treatment methods are identified through knowledge graphs. This approach enables more accurate and comprehensive decisions in healthcare, thanks to the combination of semantic reasoning and natural language processing techniques. Applications made on the mental health dataset can be used in the diagnosis of psychological disorders, especially depression and anxiety, and in the treatment recommendation system.

In the future, expanding and modifying this architecture to entail additional healthcare domains presents a possible avenue to enhance healthcare quality. Text-based data is the main focus of the current system. In order to achieve better prediction, image and video data will be integrated to create a multimodal structure. This makes it possible to assess patients' body language, facial expressions, and other visual cues to deliver more thorough and precise diagnostic and treatment suggestions. Sentiment analysis will be used to assess the emotional tone and substance of the patients' expressions in order to gather more detailed information on the mood and emotional state of the patients. The most suitable course of treatment will be selected by optimizing the system's learning capabilities by considering an integration of the patient's unique medical history, genetic information, and preferences.

REFERENCES

[1] M. P. Joachimiak et al., “The Artificial Intelligence Ontology: LLM-assisted construction of AI concept hierarchies”, arXiv preprint arXiv:2404.03044, 2024.

[2] B. Li, T. Meng, X. Shi, J. Zhai, and T. Ruan, “MedDM: LLM-executable clinical guidance tree for clinical decision-making”, arXiv preprint arXiv:2312.02441, 2023.

[3] S. Singh, “Natural language processing for information extraction”, arXiv preprint arXiv:1807.02383, 2018.

[4] D. -Q. Wang et al., “Accelerating the integration of Chat-GPT and other large-scale AI models into biomedical research and healthcare”, MedComm – Future Medicine, vol. 2(2), https://doi.org/10.1002/mef2.43, 2023.

[5] K. Singhal et al., “Towards Expert-Level Medical Question Answering with Large Language Models”, arXiv.2305.09617, 2023.

[6] Q. C. Ong et al., "Advancing Health Coaching: A Comparative Study of Large Language Model and Health Coaches", Available at SSRN 4784958, pp. 1-22, 2024.

[7] N. Shoaip, S. El-Sappagh, T. Abuhmed, and M. Elmogy, "A dynamic fuzzy rule-based inference system using fuzzy inference with semantic reasoning", Scientific Reports, 14(1), pp. 1-17, 2024.

[8] A. Ghorbani, F. Davoodi, and K. Zamanifar, "Using type-2 fuzzy ontology to improve semantic interoperability for healthcare and diagnosis of depression", Artificial Intelligence in Medicine, vol. 135, pp. 1-17, 2023.

[9] N. Shoaip, S. El-Sappagh, S. Barakat, and M. Elmogy, "A framework for disease diagnosis based on fuzzy semantic ontology approach", International Journal of Medical Engineering and Informatics, 12(5), pp. 475-488, 2020.

[10] D. Spoladore, M. Tosi, and E. C. Lorenzini, "Ontology-based decision support systems for diabetes nutrition therapy: A systematic literature review. Artificial Intelligence in Medicine", 102859, pp. 1-21, 2024.

[11] H. Guo, M. Scriney, and K. Liu, "An ostensive information architecture to enhance semantic interoperability for healthcare information systems", Information Systems Frontiers, 26(1), pp. 277-300, 2024.

[12] P. Jiang, C. Xiao, A. R. Cross, and J. Sun, "GraphCare: Enhancing Healthcare Predictions with Personalized Knowledge Graphs", arXiv preprint arXiv:2305.12788, 2023.

[13] Feng, Y. et al., "Knowledge Graph-based Thought: a knowledge graph enhanced LLMs framework for pan-cancer question answering", bioRxiv, pp. 1-10, April 2024.

[14] M. R. Karim et al., "From Large Language Models to Knowledge Graphs for Biomarker Discovery in Cancer", ArXiv, abs/2310.08365, 2023.

[15] Y. Wang, Y. Zhao, and L. Petzold, "Are Large Language Models Ready for Healthcare? A Comparative Study on Clinical Language Understanding", ArXiv, abs/2304.05368, 2023.

[16] B. Li, T. Meng, X. Shi, J. Zhai, and T. Ruan, "MedDM: LLM-executable clinical guidance tree for clinical decision-making", arXiv preprint arXiv:2312.02441, 2023.

[17] Y. Kim, X. Xu, D. McDuff, C. Breazeal, and H. W. Park, "Health-LLM: Large language models for health prediction via wearable sensor data", arXiv preprint arXiv:2401.06866, 2024.

[18] I. L. Alberts et al., "Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?", European journal of nuclear medicine and molecular imaging, 50(6), pp. 1549-1552, 2023.

[19] R. Poulain, H. Fayyaz, and R. Beheshti, "Bias patterns in the application of LLMs for clinical decision support: A comprehensive study", arXiv preprint arXiv:2404.15149, 2024.

[20] Kaggle (Accessed: September 2024), Available at: https://www.kaggle.com/datasets/hamjashaikh/mental-health-detection-dataset/data

[21] Bioportal (Accessed: September 2024), Available at: https://bioportal.bioontology.org/ontologies/ONTRISCAL

[22] ADNI (Accessed: September 2024) Available at: https://adni.loni.usc.edu/

[23] SWRL, (Accessed: September 2024) Available at: https://www.w3.org/submissions/SWRL/

[24] SPARQL, (Accessed: September 2024) Available at: https://www.w3.org/TR/sparql11-query/

 ISBN: 978-1-68558-190-9