Transforming Personal Health AI: Integrating Knowledge and Causal Graphs with Large Language Models

Zhongqi Yang[1], Iman Azimi[1], Mohammed J. Zaki[2], Manas Gaur[3], Oshani Seneviratne[2], Deborah L. McGuinness[2], Sabbir M. Rashid[2], and Amir M. Rahmani[1,4]
[1]*Department of Computer Science, University of California, Irvine*
[2]*Department of Computer Sciences, Rensselaer Polytechnic Institute*
[3]*Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County*
[4]*School of Nursing, University of California, Irvine*
{zhongqy4, azimii, a.rahmani}@uci.edu, {zaki, dlm}@cs.rpi.edu, manas@umbc.edu, {senevo, rashis3}@rpi.edu

***Abstract*—Large Language Models (LLMs) hold considerable promise for healthcare applications, leveraging vast, diverse datasets to deliver insights across a broad range of tasks. However, their effectiveness in personal health settings is currently limited by their dependence on unstructured information, leading to issues concerning accuracy, trustworthiness, and personalization. In this perspective, we propose a transformative framework integrating Knowledge Graphs (KGs) and Causal Graphs (CGs) with LLMs to tackle these challenges. KGs contribute structured, verifiable knowledge that grounds LLM outputs in validated information, while CGs delineate causal relationships that are crucial for precise health assessments and intervention planning. We illustrate this framework with a practical example in diabetes management to show its real-world application. We indicate that integrating KGs and CGs into LLMs is a pivotal advancement for addressing key challenges in personal healthcare. This integration directly tackles issues of trustworthiness, truthfulness, and personalization by anchoring LLM in such structured knowledge. Practical solutions can be deployed using this integrated approach with the support of personal health data.**

***Index Terms*—Personalized Healthcare, Large Language Model, Knowledge Graph, Causal Graph, Causal Inference, Reasoning, Truthfulness, Privacy, Bias, Explainability**

I. Introduction

Large language models (LLMs) have rapidly advanced as powerful tools capable of understanding and generating human-like text, driven by vast networks of parameters and extensive training on diverse text corpora [1]–[3]. Their ability to encode and process vast amounts of information positions them as transformative assets across various domains, particularly in healthcare, where they hold the potential to significantly enhance patient care, research, and clinical decision-making [4].

Several approaches have been developed to integrate medical knowledge into LLMs for healthcare applications. Fine-tuning involves adapting the model's parameters using domain-specific datasets to improve accuracy and relevance in medical contexts [5]–[8]. Prompt engineering, especially, Retrieval-Augmented Prompting (RAP), incorporates relevant external data into prompts in real-time, enhancing the contextual accuracy of the model's responses during medical discussions [9]–[12]. RAP first retrieves relevant information using either sparse methods (like TF-IDF [13] and BM25 [14]), or dense methods (such as DPR [15], ART [16], ColBERT [17]) and then integrates the retrieved information into the prompts to ensure more contextually accurate outputs. LLMs can also serve as agents by engaging with their environment to tackle complex healthcare tasks with precision. Rather than simply generating responses, they use structured frameworks like ReAct [18] to reason, plan, and take deliberate actions in a step-by-step process. These approaches predominantly rely on unstructured data sources, such as text from electronic health records (EHRs), clinical notes, medical academic papers, and general domains like Wikipedia, which underpin much of the current work in integrating LLMs into healthcare.

Despite the growing interest in applying LLMs to healthcare, prior work has largely leveraged text, and thus has focused less on the valuable content in structured forms. The significant challenges include the issues of truthfulness, limited task evaluation, and gaps in explainability, transparency, and privacy [19]. These problems are exacerbated by the reliance on unstructured knowledge, which often fails to capture the full complexity of medical information [20], [21] often without much provenance information. For instance, free-text clinical notes are difficult to process, may be acronymn heavy, and can result in the loss of critical connections within the data [22]–[25]. This is particularly problematic in personal health, where accurate and personalized information is essential. With unstructured knowledge, LLMs struggle to provide the necessary detail and context, limiting their effectiveness in personal health applications [26], [27].

Integrating advanced knowledge representation techniques, such as Knowledge Graphs (KGs) and Causal Graphs (CGs), into LLMs presents a promising solution to these challenges by addressing the complex relationships and causalities inherent in medical information [28]–[36]. KGs provide a structured, interconnected understanding of entities and their relationships, enabling LLMs to maintain the critical connections often lost in unstructured data. CGs, on the other hand, map out causal relationships between variables, enhancing LLMs' ability to reason and make informed decisions based on nuanced causalities. For instance, leveraging a medical KG can improve an LLM's ability to reason and explain diagnostic conclusions, while constructing a causal graph from a user's personal data can capture nutritional effects unique to the individual, enabling more personalized recommendations [29], [35], [37], [38]. These structured representations can elevate the effectiveness of LLMs by aligning external knowledge with the rigorous demands of healthcare applications.

In this perspective paper, we explore how the integration of KGs and CGs can overcome the limitations of current LLMs in personal health applications. We envision a future where these graph-based technologies become integral to LLMs, bringing much-needed structure to enhance truthfulness, trustworthiness, and personalization in healthcare AI. Our discussion is centered on several key contributions:

1) Investigating the unique roles and contributions of KGs, CGs, and LLMs, as well as the benefits of their integration for personal health.
2) Exploring how the integration of KGs, CGs, and LLMs can address specific challenges in personal health.
3) Providing a framework and a practical demonstration of how this approach could be applied in personal diabetes management.

II. Knowledge Graphs and Causal Graphs

A KG is a structured representation of information that interlinks entities, such as people, places, and events, through relationships [39]–[41]. Each entity in a KG is a node, and the relationships between them are edges. This representation allows for the organization and retrieval of factual data in a way that captures both the entities themselves and context in which they exist.

For example, Google Health Knowledge Graph is an early implementation of how structured information is organized in the healthcare domain [42]. It connects various entities, such as diseases, symptoms, treatments, risk factors, and associated conditions, into an interconnected web of knowledge. For instance, in the context of diabetes, the Knowledge Graph might include entities like "Diabetes," "Insulin," "Blood Glucose," and "Diet," with relationships represented by edges such as "is treated with" (e.g., "Diabetes is treated with Insulin") or "affects" (e.g., "Diabetes affects Blood Glucose levels"). This interconnected structure allows healthcare providers and patients to access relevant medical information more quickly and accurately.

Causal Graphs (CGs) add more expressivity options as well as complexity by using directed edges to clearly represent the cause-and-effect relationships between different entities or variables. Rooted in the principles of causal inference, CGs focus on understanding how changes in one node (the cause) can directly influence another (the effect) [43]–[45]. This representation of power makes CGs particularly valuable in healthcare for predicting disease progression or evaluating the impact of interventions, where understanding cause-and-effect relationships is crucial.

In a healthcare setting, for instance, in the study of temporomandibular disorders (TMDs) [44], the nodes in the causal graph represent factors such as "Facial Injury," "Pressure Pain Threshold," "Age," "Gender," "Psychological Stress," and "Genetic Predisposition." The directed edges illustrate causal relationships, such as "Facial Injury affects Pressure Pain Threshold," and "Pressure Pain Threshold influences the risk of developing TMDs." This causal graph can help researchers understand how addressing one factor, like mitigating facial injury, could potentially alter the pressure pain threshold and thereby reduce the likelihood of TMDs.

The primary difference between KGs and CGs lies in the types of relationships they model. KGs excel at organizing and retrieving factual information based on associative relationships, but they typically do not include causality explicitly. In contrast, CGs are specifically designed to map out cause-and-effect dynamics, which are crucial for applications that require predictive and prescriptive analytics. Rather than representing all possible relationships between data entities, a causal graph focuses on the relationships that directly connect causes and effects.

III. Challenges in Applying Large Language Models to Personal Health

Despite the notable advancements of LLMs in healthcare, they encounter substantial obstacles in personal health applications. Achieving high-standard personalized health responses, which is the core objective of personal health applications, presents unique challenges. These challenges arise from the complexities of ensuring data availability, accurately analyzing individual patterns, contextualizing findings within broader health metrics, integrating population norms, and delivering tailored responses. Specifically, the limitations of LLMs in this domain can be broadly categorized into two key areas: the need for deeper personalization and the imperative for enhanced trustworthiness.

A. Trustworthiness

Trustworthiness for LLMs in personal health is a complex and multifaceted issue [46]–[48]. Our discussion focuses on three critical factors of trustworthiness: Privacy, Bias, and Explainability.

1) Privacy: Privacy concerns are raised since achieving deeper personalization involves accessing detailed personal health data, which includes sensitive information such as medical history, genetic data, and personal identifiers. Exposing personal data to LLMs raises critical risks of data breaches and unauthorized access, potentially leading to severe privacy violations [49]–[55].

2) Explainability: These limitations also impact the model's capacity for explainability, which involves transparently describing the reasoning process, enabling users to understand how conclusions were derived. In personal health, where decisions can profoundly affect patient outcomes, the ability to audit and validate these decisions is crucial [56], [57]. Processing complex personal data without a clear mechanism to explain outputs complicates the trust between the technology and its users—both patients and healthcare providers—and further limits the utility of LLMs in sensitive health applications [58]–[62].

3) Bias: Moreover, bias in LLM responses, especially when handling personal data, manifests as unequal and unjust treatment. For example, an LLM might provide high-quality responses for certain populations but deliver less accurate or effective recommendations for individuals from underrepresented groups [63]–[65]. This often results from biased training corpora and the use of sensitive personal information like demographics, which can reinforce existing disparities [66]–[70].

B. Truthfulness

Truthfulness refers to the accurate representation of information, facts, and results. This is a critical concern in healthcare because errors can lead to misdiagnoses, inappropriate treatments, and overall patient harm [71]–[73]. LLMs often struggle to provide truthful responses when relying solely on their internal knowledge [74]. A major challenge is the phenomenon of "hallucination," where LLMs generate incorrect but convincingly presented outputs. This issue stems from misinformation or outdated information in the training data, along with the inherent generative nature of LLMs. To mitigate this, the generative process should be guided by factual, validated knowledge sources to improve accuracy. Research shows that LLMs augmented with external knowledge outperform state-of-the-art models on benchmark datasets [75].

C. Deeper Personalization

Deeper personalization in personal health requires responses to be better tailored to individuals. This involves a detailed consideration and extraction of nuanced information from each patient's multifaceted personal data, such as patient histories, preferences, and specific health conditions, which is crucial for the truthfulness and safety of the health advice provided [76]–[78]. Some existing studies attempt to enable LLMs to provide personalized responses by incorporating raw personal data. However, a deeper level of personalization required for personal health often demands more sophisticated processing of complex personal data [49], [50], [79], [80]. For example, a diabetic patient's personal health data may include structured clinical information such as glucose readings, medication dosage, and lab results, along with unstructured data like patient-reported symptoms (e.g., fatigue, stress) and lifestyle information from wearables or smartphones (e.g., daily physical activity, sleep patterns). Providing a deeply personalized response requires sophisticated data processing across multiple aspects of a patient's health [79], [81]. For instance, in response to a diabetic patient asking, "How can I manage my blood sugar better?", the system would need to check the availability of recent glucose readings, analyze trends in blood sugar levels over time, and identify any anomalies. Additionally, it would assess lifestyle factors such as diet and exercise from wearable devices, correlate these findings with medication schedules, and contextualize them against both personal health history and population norms. LLMs can function as agents by planning and executing steps to generate tailored recommendations, integrating both structured data (e.g., glucose levels) and unstructured inputs (e.g., patient-reported symptoms) [29], [82], [83]. However, they still face challenges in navigating unstructured text, complex medical terminology, and diverse data formats, lacking the advanced mechanisms needed for fully personalized healthcare.

IV. Contributions of Knowledge Graphs, Causal Graphs, and Large Language Models in Personal Health

Understanding the impact of combining KGs, CGs, and LLMs in personal health starts by recognizing the individual strengths of each. This section examines the distinct roles of KGs, CGs, and LLMs and highlights their contributions to overcoming the challenges of personal healthcare, as Figure 1 illustrates.

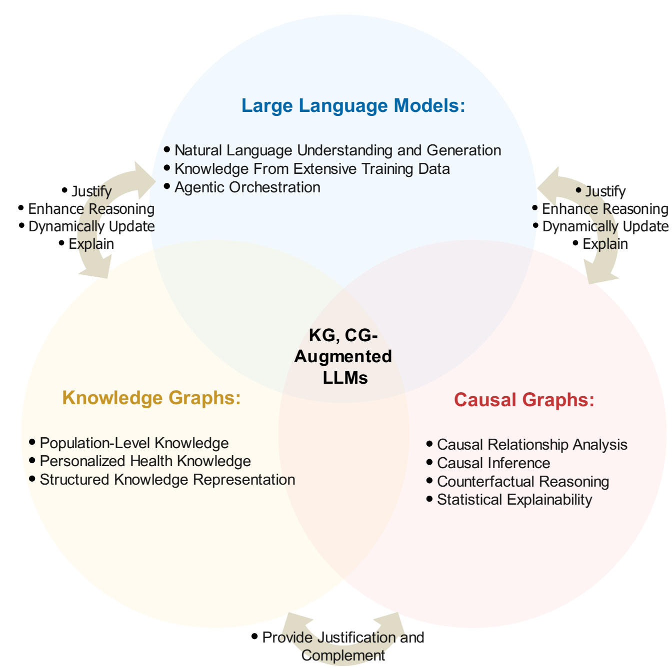

Fig. 1. Overview of Integrating Knowledge Graphs, Causal Graphs, and Large Language Models in Personal Health

A. What Can Knowledge Graphs Provide?

A Knowledge Graph organizes data through a structured network of entities (such as objects, events, or concepts) interconnected by edges that represent their relationships [84]. This graphical format allows for representing and querying complex relationships and attributes in an integrated manner, which is particularly advantageous in healthcare where the connections between symptoms, diseases, treatments, and patient characteristics are intricately linked.

1) Comprehensive Population-Level Knowledge: KGs can compound a vast array of health-related information across different populations, including epidemiological data, treatment outcomes, demographic statistics, etc. [85]. This integration provides a holistic view that can inform public health decisions and research [86]. By capturing diverse health scenarios within the population, KGs can help a personal health service generate responses that are contextually appropriate and informed by real-world, broad-spectrum validated knowledge.

2) Personalized Health Knowledge: KGs provide detailed representations of an individual's health profile by integrating diverse personal data. This data includes medical history, current medications, genetic information, lifestyle factors, familial health patterns, and personal health goals. By organizing these elements into a knowledge graph, personal KGs offer a holistic view of a patient's health, reflecting the various factors that contribute to their overall well-being [87]. This comprehensive perspective can provide more precise tailoring of medical recommendations to meet the patient's unique health needs and circumstances.

3) Structured Knowledge Representation: KGs systematically organize information, enabling efficient retrieval and analysis of data. This structured approach is essential for comprehending and generating information relevant to specific personal health queries. It facilitates effective navigation through complex medical terminology and patient data. More specifically, by explicitly representing the relationships between various data types—such as symptoms, diagnoses, treatments, and outcomes—KGs allow for precise querying and retrieval of specific information. This organization enhances the relevance of the resulting analyses and makes it easier to derive meaningful insights from complex healthcare data.

B. What Can Causal Graphs Provide?

A Causal Graph is a directed graph that illustrates causal relationships between variables. It is fundamental for those systems where understanding the causal relationships founded on statistical principles is essential [88]. In the healthcare scenario, CGs play a vital role by mapping how changes in one factor can causally influence others. This aids in the determination of the causal impacts of treatments, lifestyle choices, or genetic factors on health outcomes.

1) Causal Relationship Analysis: CGs provide a structured representation of causal relationships among various health factors. This can help LLMs understand disease progression, treatment effects, and potential side effects. As a result, decision-making becomes more informed, considering the full health scenario of patients. In complex medical cases where multiple variables interact, CGs enable users to manage potential complications or treatment benefits effectively.

2) Support for Causal Inference and Counterfactual Reasoning: CGs encode causal assumptions that are essential for using statistical methods to distinguish true causation from mere correlation. With CGs, it is possible to conduct causal inference to determine whether specific treatments directly affect health outcomes or if observed effects arise from other confounding factors. Additionally, CGs facilitate counterfactual reasoning by exploring "what-if" scenarios, which allows anticipation of the outcomes of various treatment options before their actual implementation. This ability provides insights into potential effects and side effects of treatments and offers a depth of understanding beyond the capabilities of LLMs alone.

3) Statistical Explainability: This is based on established statistical theory and principles, is allowing LLMs to provide reasoning that is transparent and verifiable, which is essential for explainability in healthcare applications [89]. CGs outline how specific health outcomes can be causally linked to certain treatments or conditions. This capability is particularly valuable when LLMs need to justify their outputs and build user trust in their recommendations.

4) Personalized Causal Insights: Personalized CGs can be constructed to include individual patient data to offer detailed insights into the potential impacts of specific health interventions based on each patient's unique health profile. This level of personalization extends beyond generic medical advice, and allows for better customized recommendations that reflect an individual's specific health conditions and predispositions. By incorporating personalized CGs, healthcare delivery can become more responsive and precisely tailored to meet the unique needs of individual patients.

C. What Can Large Language Models Provide?

Large Language Models represent the next generation of tools for natural language processing and generation, built upon vast amounts of training data, and are increasingly powerful in personal health applications due to their advanced decision-making capabilities.

1) Natural Language Understanding and Generation: Large Language Models excel in understanding and generating human language, making them invaluable for interpreting complex medical inquiries and providing coherent, contextually relevant responses. Their ability to comprehend nuanced language enables them to interpret patient questions, extract relevant information from diverse textual sources, and generate detailed, patient-specific advice [72], [90]–[94]. This proficiency bridges the gap between technical medical knowledge and everyday patient communication, ensuring clarity and understanding.

2) Knowledge From Extensive Training Data: Trained on a vast corpus spanning a wide range of topics, LLMs possess a rich repository of both general and specialized knowledge. This extensive training enables them to recall and synthesize information from diverse medical literature, guidelines, and case studies. Techniques such as zero-shot prompting, few-shot prompting, and chain-of-thought reasoning can be applied to invoke this inherent knowledge, guiding LLMs to generate accurate and contextually relevant responses [95]–[97]. In healthcare applications, this empowers LLMs to provide insights that are informed by the latest medical research and established best practices.

3) Agentic Orchestration: LLMs can act as central orchestrators in the framework to manage complex tasks that extend beyond simple language processing. They have the ability to plan, execute, and coordinate multi-step processes by utilizing KGs and CGs to collect, process, and analyze patient data across various stages. This structured, iterative approach allows LLMs to manage healthcare more effectively.

Specifically, LLMs serve as orchestrators by interfacing between patients and different healthcare tools. They autonomously choose the most suitable resources—such as retrieving information from KGs or performing predictive analysis with CGs—to respond to complex health queries. For example, in chronic disease management, an LLM can continuously adjust treatment plans by integrating real-time data from wearable devices with patient histories stored in KGs. Through this orchestration across data sources and tools, LLMs are able to facilitate personalized healthcare by making adaptive, well-informed decisions.

D. How Can Knowledge Graphs, Causal Graphs, and Large Language Models Collaborate?

The integration of KGs, CGs, and LLMs establishes a comprehensive framework that strengthens the functionality of each component within personal health applications, as demonstrated in Figure 1. This collaboration enables LLMs to utilize structured and validated knowledge from KGs and CGs, resulting in more precise and reliable healthcare solutions tailored to individual needs.

1) Justification and Error Mitigation: KGs provide a structured factual base that LLMs can use to validate their responses. Aligning outputs with verified information from KGs reduces biases and errors that might arise from the training data. This approach ensures that the health advice generated by LLMs adheres to established medical knowledge and guidelines. CGs offer insights into the causal relationships and potential outcomes of various health interventions, enhancing the factual and contextual relevance of responses. When combined, KGs provide the factual foundation that supports the causal relationships identified by CGs, while CGs clarify the implications of these facts. This synergy builds a solid basis for accurate and reliable health information.

2) Dynamic Knowledge and Causality Integration: KGs and CGs serve as dynamic guides that refine the decision-making process of LLMs. KGs provide up-to-date factual information, while CGs offer insights into the causal mechanisms and expected outcomes of health interventions. Regular updates to these graphs incorporate new knowledge and causal insights, which allows LLMs to access the most current information. This approach avoids the need for frequent retraining or fine-tuning of the models. Consequently, LLMs can produce outputs that are aligned with the latest medical practices and adaptable to the changing landscape of patient data and advancements in medical science.

3) Enhanced Large Language Models' Reasoning: Integrating KGs and CGs into LLMs advances their reasoning capabilities beyond what training or prompt engineering alone can achieve. KGs enrich LLM reasoning by providing a vast repository of factual knowledge for the generation of well-reasoned responses that are traceable within the graph. This traceability creates a clear, logical pathway from question to answer, linking each response to specific data points and relationships.

Additionally, CGs enhance the ability of LLMs to understand and predict health outcomes by elucidating causal relationships within medical data. This integration allows LLMs to accurately anticipate outcomes based on symptoms and treatments, and to grasp the underlying mechanisms driving these outcomes. Thus, by combining structured knowledge from KGs and causal insights from CGs, LLMs can produce responses that are both reasoned and grounded in a robust understanding of medical contexts.

4) Improved Explainability: KGs provide a structured representation of medical knowledge, delineating relationships between concepts. This empowers LLMs to not only generate responses but also reference the specific knowledge underpinning them. This transparency builds trust, allowing healthcare professionals and patients to understand the rationale behind the LLM's suggestions. CGs, on the other hand, excel at revealing causal relationships between health variables. By integrating CGs, LLMs can explain not just the answer itself, but the causal chain leading to it. This unveils the underlying logic behind the LLM's reasoning, fostering trust and enabling informed decision-making within the context of a patient's specific health data.

5) Bias Mitigation and Privacy Protection: KGs filter and ground the information used by LLMs in validated, unbiased facts. Sensitive data, such as ethnicity, can be included in KGs when relevant, but used selectively based on context [98]. For example, ethnicity may be applied only when it has a clinically significant role, like when certain populations show genetic predispositions to specific health conditions. This approach balances the need for bias mitigation with the necessity of including clinically relevant data so that sensitive attributes are considered only where appropriate, without reinforcing demographic biases in unrelated contexts.

CGs take things a step further. They equip LLMs with the ability to understand cause-and-effect relationships that influence health outcomes. This is achieved through statistical analysis, bypassing potentially biased factors like ethnicity or demographics. CGs take bias mitigation a step further by focusing on the cause-and-effect relationships that truly influence health outcomes, rather than relying on simple correlations. For instance, in the context of cardiovascular disease, a CG could model the causal relationships between factors such as diet, exercise, and heart health while excluding potentially biased variables like economic status. The focus on causality over correlations helps mitigate bias in the LLM's responses. Additionally, CGs can act as an intermediary layer, further protecting sensitive personal information.

V. Framework For Integration of Knowledge Graphs, Causal Graphs, and Large Language Models

A. Proposed Framework

Integrating KGs, CGs, and LLMs can effectively tackle key challenges in healthcare. This synergy improves trustworthiness, truthfulness, and personalization in medical applications. We propose a framework that integrates Knowledge Graphs, Causal Graphs, and Large Language Models to address this challenge. The framework integrates five key components to deliver personalized health management: Personal Health Data, Population Knowledge Graphs, Personal Knowledge Graphs, Personal Causal Graphs, and LLMs. The system synthesizes data from an individual's medical history, genetic profile, lifestyle habits, and biomarkers with population-level medical knowledge, personal health trends, and causal relationships to generate evidence-based personal health responses. The proposed framework is illustrated in Figure 2.

Personal Health Data forms the cornerstone of our framework, providing a comprehensive view of an individual's health status. This component includes medical history, genetic profile, lifestyle habits, physical activity patterns, vital signs, etc. It captures past and present health conditions, inherited traits, daily behaviors, exercise routines, and real-time physiological measurements. This detailed personal health information serves as the foundation for generating tailored health insights and recommendations. By continuously updating this diverse dataset, the system maintains an accurate representation of an individual's health state.

Population Knowledge Graphs are comprehensive collections of medical information, combining insights from research, clinical trials, and health statistics. These graphs connect various health concepts—such as diseases, symptoms, and treatments—showing how they relate to each other. This structure helps the system understand the complex interplay between different aspects of health. For example, a Population Knowledge Graph might show how a particular medicine relates to its effects, proper usage, and interactions with diet and exercise.

Personal Knowledge Graphs create a detailed picture of an individual's health. They are constructed from Personal Health Data to form a complete health profile. Personal Knowledge Graphs organize this personal information in a way that mirrors the structure of Personal Knowledge Graphs. This allows for a thorough understanding of a person's health history and current condition. For instance, a Personal Knowledge Graph might link someone's exercise habits, eating patterns, and sleep routines to their blood pressure over time. It could also include details about their job, family health history, and recent health changes, providing a full view of their health situation.

Personal Causal Graphs map out how different factors affect an individual's health. Personal Causal Graphs show how changes in one area of health might influence others, based on the individual's unique body and past experiences. For example, a Personal Causal Graph could predict how changes in diet or exercise might affect a person's cholesterol or mood, considering how they've responded to similar changes before. This helps the system anticipate how different health choices or treatments might affect the individual, allowing for more precise and personalized health advice.

An **LLM** serves as the intelligent core of the framework, interpreting user queries and translating complex medical information into actionable advice. It processes health-related questions with a nuanced understanding. The LLM accesses and integrates information from all three graph types: drawing broad medical knowledge from Population Knowledge Graphs, individual health profiles from Personal Knowledge Graphs, and predictive insights from Personal Causal Graphs. This comprehensive approach allows the LLM to generate personalized health recommendations, balancing general health guidelines with an individual's unique health history, current status, and predicted responses to interventions. The LLM communicates these tailored suggestions in understandable language, explaining the reasoning behind its advice. As new medical knowledge emerges and individual health data accumulates, the LLM continuously refines its responses to provide the most relevant and up-to-date guidance for health management.

B. Enhancing Trustworthiness

1) Safeguarding Privacy: Privacy is a fundamental right that ensures that patients' personal and medical information is protected and kept confidential. In personal health applications involving LLMs, privacy concerns are paramount, especially as personalization often requires access to sensitive personal data. Safeguarding this privacy is critical to maintaining patient trust and ensuring that sensitive information is neither exposed nor misused.

KGs and CGs can elevate privacy protection by serving as an abstraction layer, as demonstrated in Figure 4. These graphs extract and represent connections among symptoms, diagnoses, treatments, and outcomes while obscuring direct access to personal data [99]–[101]. CGs model causal relationships between treatments and health outcomes based on aggregated data. This approach allows KGs and CGs to safeguard privacy by structuring personal details in a format that prevents direct exposure of sensitive information.

For example, a KG can be constructed from personal data to represent a patient's medical condition and associated treatments. This structured representation allows LLMs to access and utilize necessary healthcare insights without linking directly to identifiable personal data. This process protects patient privacy while still enabling the LLM to provide personalized and accurate medical recommendations.

2) Advancing Explainability: In personal health applications, where decisions can directly influence patient outcomes, explainability is paramount [102]. Explainability refers to the ability to deconstruct and articulate the rationale underlying LLMs' outputs. Enhancing explainability in LLMs is vital for fostering trust among healthcare professionals and patients, and supporting informed decision-making [56], [103].

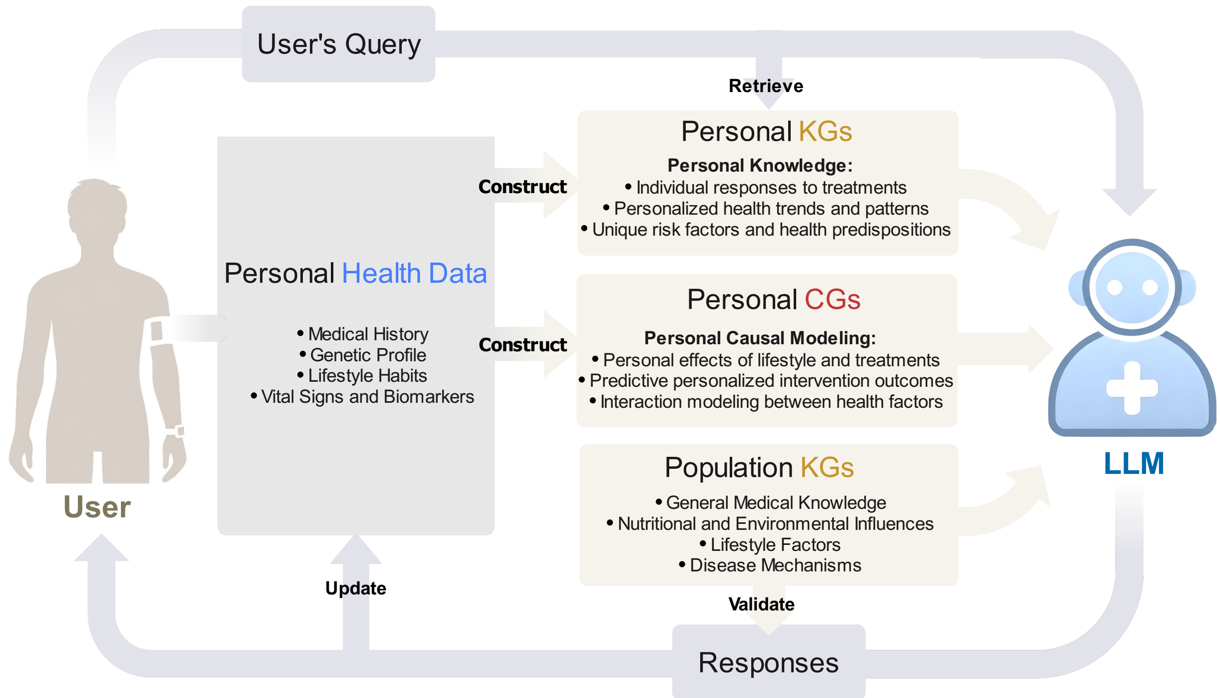

Fig. 2. KG and CG-Augmented LLM Framework for Personal Health

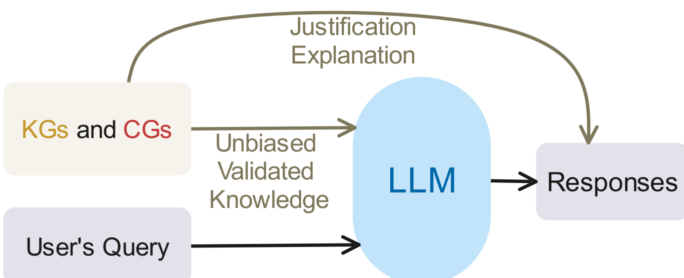

Fig. 3. Illustration of Enhancing Truthfulness, Explainability, and Reducing Bias through KGs and CGs. The validated KGs and CGs provide unbiased, validated knowledge to the LLM. The LLM processes this knowledge alongside user queries to generate responses. The model also offers justification or explanation for its outputs, improving transparency and trustworthiness.

KGs and CGs provide a clear and structured representation of medical knowledge, delineating relationships between various healthcare concepts. This structured approach allows LLMs to generate responses that are both clear and traceable, as demonstrated in Figure 3. For example, KGs created from electronic medical records (EMRs) offer a high-quality overview of medical knowledge [42], [104], [105]. These KGs mirror clinical reasoning and decision-making processes, thus addressing issues of explainability in LLMs. CGs enhance explainability by mapping out causal relationships between different health variables. This allows LLMs to explain not just the answers but also the causal pathways leading to those answers. For instance, when an LLM suggests a particular medication for hypertension, a CG can illustrate how this recommendation is derived from the patient's specific symptoms, health history, and potential interactions with other treatments.

In practice, when KGs and CGs are integrated with LLMs, they allow the models to clearly articulate the basis of their decisions. This integration enables LLMs to reference specific nodes and connections within the graphs when explaining their outputs. For example, an LLM might reference a KG to explain that a treatment recommendation is based on the latest clinical trials and safety warnings, while a CG could show how the treatment will affect the patient's health outcomes based on their personal data.

Studies have shown that the structured and validated knowledge provided by KGs and CGs significantly improves the transparency and reliability of LLM-driven healthcare recommendations [87], [106], [107]. This approach prioritizes explainability, thereby equipping users with the ability to grasp the reasoning behind the model's recommendations. This transparency strengthens trust in the model's outputs and empowers users to make informed decisions.

3) Mitigating Bias: Mitigating bias in personal health LLM applications is crucial to ensure equitable and accurate healthcare recommendations for all patients. Bias in LLMs can lead to unequal treatment based on patient demographics such as race, gender, or socioeconomic status, which can have serious implications for patient care and outcomes.

A key advantage of this integration lies in the structured, curated knowledge that KGs and CGs provide as shown in Figure 3. When carefully constructed, these graphs contain factual and comprehensive medical information, allowing LLMs to leverage this unbiased knowledge base. By grounding LLM-generated insights in well-curated medical facts, the system reduces the risk of bias typically associated with unstructured data sources, ensuring that health recommendations are more accurate and fair.

Even when KGs and CGs contain some inherent bias, they still play a vital role in reducing bias, particularly for underrepresented populations. These graphs can be specifically designed to capture diverse medical data across various demographic groups, including those traditionally underrepresented in healthcare datasets, such as specific genders, ethnicities, or age groups. KGs serve as structured, objective repositories of knowledge, standardizing information regardless of patient demographics or other potential sources of bias. For example, a KG can be constructed to facilitate balanced representation across racial and gender groups for conditions like heart disease or diabetes. Moreover, existing techniques [108]–[111] can help mitigate bias within KGs broadening the range of patient scenarios considered and better addressing the healthcare needs of diverse populations.

Similarly, CGs can model causal relationships grounded in robust statistical evidence, rather than relying on unexamined demographic correlations. This approach helps eliminate the risk of perpetuating historical biases, ensuring that LLM outputs are grounded in factual, unbiased data.

In practice, an LLM integrated with these graphs can draw on this structured and balanced information to generate responses. For instance, if a patient inquires about treatment options for diabetes, the LLM can refer to a KG that includes data from diverse patient populations and a CG that models the causal impact of various treatments without demographic bias. This integration allows for more accurate and personalized healthcare recommendations, ensuring that these recommendations are free from biases that might have arisen from unrepresentative training data.

C. Ensuring Truthfulness

Truthfulness in personal health LLM applications involves ensuring that these systems operate without causing unintended harm to patients. This includes ensuring that the LLM's advice, diagnoses, and treatment recommendations are accurate, reliable, and consistent with current medical standards. Prioritizing the truthfulness of LLMs is crucial to avoid misdiagnoses, inappropriate treatments, and overall patient harm to help build trust in healthcare applications.

Integrating KGs and CGs with LLMs greatly improves truthfulness in healthcare. As illustrated in Figure 3, KGs organize medical information in a structured way, making sure that LLM decisions are based on validated, up-to-date, and precise medical knowledge. This structure helps reduce diagnostic and treatment errors by aligning LLM outputs more closely with established medical standards and practices [112]–[114].

In practical applications, KGs and CGs guide LLMs by providing access to the latest treatment guidelines and research findings. For example, when recommending a treatment, a KG makes sure that the LLM considers the most current clinical trials and truthfulness warnings, lowering the risk of recommending outdated or unsafe treatments. Similarly, CGs map out the potential consequences of a recommendation by showing the causal relationships between treatments and outcomes. This allows LLMs to predict and avoid adverse reactions or ineffective treatments. For instance, a CG might show how a particular food could impact health outcomes in patients, enabling the LLM to recommend safer alternatives based on personalized causal analysis.

KGs and CGs also serve as review mechanisms to highlight potential errors or inconsistencies in LLM responses, preventing unsafe advice. For example, if an LLM generates a treatment recommendation, the integration with KGs and CGs allows for an automatic cross-checking process where the generated advice is validated against the structured knowledge within the graphs. An example of this would be an LLM suggesting a recipe containing an ingredient that is risky for a patient with diabetes; the KG could flag this issue based on stored medical profiles, preventing the spread of potentially harmful advice.

D. Deepening Personalization

Deeper personalization in personal health involves tailoring healthcare responses to meet each individual's unique needs and circumstances. This approach considers various patient-specific data, such as medical history, genetic predispositions, lifestyle choices, and personal preferences. This leads to improved patient outcomes, greater patient satisfaction, and increased trust in these advanced healthcare solutions.

Integrating KGs and CGs with LLMs marks a pivotal step toward deeper personalization in personal health management. Figure 4 outlines the methodological framework for this process. By constructing personal KGs from an individual's health records and developing personal CGs that model the unique causal effects of interventions, such as medications and lifestyle changes, LLMs gain access to a structured and information-rich knowledge base. This enriched knowledge allows LLMs to deliver more effective and personalized responses.

When an LLM equipped with these integrated graphs receives a health query, it can efficiently access relevant, personalized information from the KGs and CGs. Consider a diabetic patient who asks, "What are the potential risks of following a high-protein diet, given my current medications and health conditions?" An LLM integrated with Knowledge Graphs (KGs) and Causal Graphs (CGs) would first reference the patient's personal KG to retrieve critical information, such as their current medications (e.g., Metformin), known allergies (e.g., to dairy products), and dietary restrictions (e.g., low carbohydrate intake). Next, using the personal CG, the LLM can evaluate how the high-protein diet might affect the patient's blood glucose levels. For instance, the CG might indicate that consuming large amounts of protein could lead to increased gluconeogenesis, potentially causing elevated blood glucose levels, especially in the presence of certain medications like Metformin, which alters glucose metabolism.

Figure 4 illustrates this process. For example, a personal KG might compile comprehensive details about a patient's previous health conditions, current medications, and family medical history, highlighting potential health risks and effective treatment options [87], [106], [107]. Simultaneously, a personal CG could show how specific lifestyle adjustments, such as changes in diet or exercise routines, could causally influence this patient's risk factors for conditions like diabetes or heart disease.

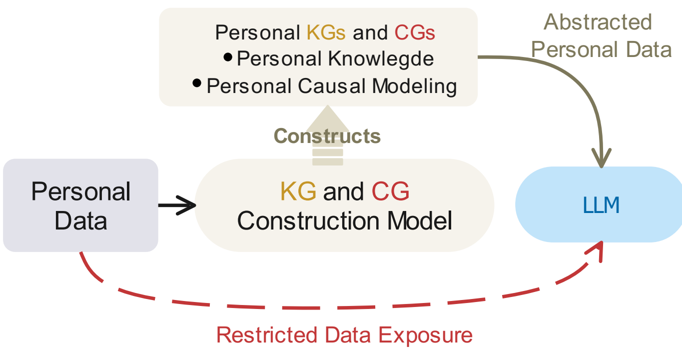

Fig. 4. Illustration of Deeper Personalization and Privacy Preservation through KGs and CGs. Personal data is first used to construct KGs and CGs via a construction model. These graphs capture personal knowledge and causal relationships. The resulting personal KGs and CGs are then abstracted, and only this abstracted personal data is passed to the LLM. Raw personal data is not directly input to the LLM.

VI. Practical Example: Personal Diabetes Management

This section showcases a practical application of integrated KGs, CGs, and LLMs to empower diabetic patients in their self-care routines. By combining personal health data with broader medical knowledge, the system constructs a personalized view of the patient's health. This integrated approach allows the LLM to understand the complex interplay between factors like diet, exercise, and medication, enabling it to generate evidence-based recommendations tailored to each individual's unique needs. An illustration of the workflow is shown in Figure 5.

A. The Inputs

1) Personal Health Data: The process begins by collecting and structuring the patient's personal health data. Effective diabetes management relies on continuous monitoring of blood glucose levels, careful dietary planning, regular physical activity, and adherence to medication protocols [115]–[118]. Wearable and mobile sensors/devices can be employed to track these metrics accurately. For example, continuous glucose monitors provide real-time data on blood glucose levels, fitness trackers monitor physical activity, and smart scales track weight changes. Additionally, dietary intake can be logged through mobile apps that track nutrition and meal timing. By continuously monitoring and recording a wide range of health metrics, the personal health data establishes a strong foundation for personalized diabetes management.

2) Population Knowledge Graphs: Population knowledge graphs refer to the population-level medical knowledge. They are utilized to offer a broader context to individual patient data. These graphs compile extensive health-related information across diverse populations, including epidemiological data, treatment outcomes, and demographic statistics. Integrating population knowledge graphs empowers the system to leverage this validated, comprehensive medical knowledge.

B. Personal Knowledge Graph and Causal Graph Construction

1) Personal Knowledge Graph: This is constructed from patient data to convert raw information into structured, actionable knowledge. This transformation is crucial for organizing and understanding the complex interactions between various health factors and their impacts on diabetes management [34]. The Personal KG integrates information on dietary habits, such as carbohydrate intake and meal timing, and lifestyle choices, including physical activity levels and sleep patterns. It links these factors to their impacts on blood glucose control and insulin sensitivity. Real-time health metrics, like continuous glucose monitoring data, are represented to provide insights into the patient's glycemic patterns and trends over time. This interconnected knowledge demonstrates how different treatments, lifestyle modifications, and genetic factors influence diabetes management.

2) Personal Causal Graph: The Personal Causal Graph maps out the causal relationships between different factors influencing the patient's health. Understanding these causal pathways is essential for predicting the outcomes of various health interventions and making informed decisions about diabetes management. The Personal CG models the causal pathways between various health determinants such as diet, physical activity, and medication adherence [29]. For instance, the CG elucidates how different foods impact blood glucose levels, how varying intensities of physical activity influence insulin sensitivity, and how consistent medication adherence affects overall glucose control. It enables predictive analysis, allowing healthcare providers to estimate the potential outcomes of different self-care strategies. For example, the CG can predict the effects of adjusting carbohydrate intake or increasing physical activity on the patient's glucose levels.

C. Personalized Diabetes Management Recommendations via LLMs

1) Query Processing: The process begins when the LLM receives a query from a patient regarding their diabetes management. For instance, a patient might ask, "What should I eat for breakfast to maintain stable blood sugar levels?" The LLM interprets this query, identifies the key aspects, and determines the specific information required to provide an accurate response.

2) Knowledge Retrieval: The LLM retrieves relevant information from both personal and population knowledge graphs. The personal knowledge graph provides insights into the patient's medical history, dietary preferences, and current health

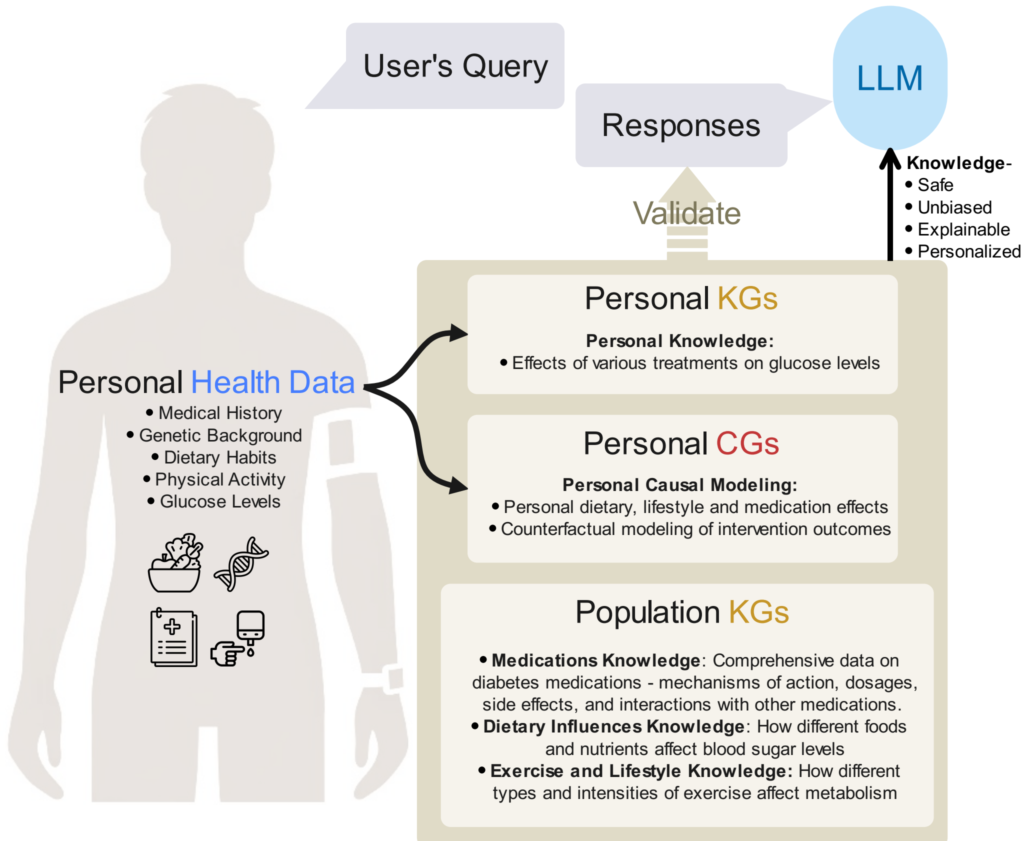

Fig. 5. An Example of Personalized Diabetes Management Using KG and CG-Augmented LLM.

metrics. The population knowledge graph offers validated data on the impacts of various dietary choices on blood sugar levels across different populations.

3) Causal Analysis: Using the causal graph, the LLM assesses the potential impact of various breakfast options. By examining the causal relationships between different foods and their predicted effects on blood glucose, the LLM predicts how specific dietary choices will influence the patient's glucose levels based on their unique health profile.

4) Trace Explanations from the Graphs: To maintain explainability and build trust, the LLM traces its reasoning and explanations back to the data and relationships within the knowledge and causal graphs. This involves outlining the causal pathways and factual bases for its recommendations and providing patients with clear and understandable rationales for the advice given.

5) Response Generation: Based on the insights derived from the knowledge and causal graphs, the LLM generates personalized, evidence-based dietary recommendations. For example, it might suggest a breakfast high in fiber and low in simple carbohydrates to help maintain stable blood sugar levels. This recommendation is tailored to the patient's individual health needs and supported by both personal and population-level medical knowledge.

6) Response Validation: Before presenting the response to the patient, it undergoes a rigorous validation process to prevent potential inaccuracies. The LLM cross-references the generated response with the knowledge and causal graphs to ensure it aligns with the patient's health condition and medical knowledge, verifying that no contraindications or allergies are overlooked. The LLM evaluates the response against the causal graph to confirm the predicted positive outcomes and check for any potential negative impacts. This process involves modeling the causal pathways to verify the truthfulness and effectiveness of the recommendations. Any discrepancies or potential improvements identified during this modeling prompt the LLM to refine the response.

D. Continuous Monitoring and Adaptation

1) Real-Time Data Integration: Incorporating real-time data from glucose monitors and wearable devices is essential. This process involves updating the knowledge and causal graphs with the latest health metrics from the patient. By using current information, the system can provide accurate and relevant recommendations tailored to the patient's immediate needs.

Real-time data integration ensures the system's responses are based on the latest health status. For example, if a glucose monitor detects a spike in blood sugar levels, this information updates the graphs immediately, allowing the system to adjust its advice accordingly.

2) Feedback Loop: A dynamic feedback loop allows the system to refine its recommendations continually. By collecting user feedback and new health data, the system can update its understanding and improve its advice over time. The feedback loop is vital for learning from the patient's unique responses to previous recommendations. This iterative process allows the responses to be more personalized and effective with each cycle.

For instance, if a dietary recommendation leads to unexpected glucose fluctuations, the system analyzes this feedback and adjusts future advice to better suit the patient's needs. This may involve modifying the causal graph to incorporate new insights about the patient's reactions to specific foods or activities.

VII. Technical Challenges and Opportunities

A. Bridging Structured Knowledge and Large Language Models

One of the key challenges in integrating KGs and CGs into Large LLMs lies in the fundamental differences in how they represent knowledge. KGs store information in a structured, symbolic form that explicitly maps relationships between entities. In contrast, LLMs operate using high-dimensional embeddings, where knowledge is encoded implicitly through statistical patterns in their parameters.

While early efforts, such as fine-tuning and prompt engineering, have begun to explore potential solutions, they remain limited. For example, even when factual information from graphs is included in the prompt, the performance of LLMs is highly sensitive to the prompt template itself, sometimes resulting in incorrect responses [29], [119]. Bridging the gap between these two forms of knowledge requires more advanced embedding techniques [120]. These techniques must translate the structured data from KGs into a format that is compatible with the statistical nature of LLMs. Achieving this hybrid representation is complex, as it needs to preserve the explicit relationships inherent in KGs while enabling LLMs to process and reason over the integrated data efficiently.

B. Computational Challenges

1) Training Costs: Integrating graphs into LLMs significantly increases computational demands. The model must query, process, and merge structured medical knowledge data from KGs and CGs with the language understanding tasks performed by LLMs. This heightens the need for processing power. Techniques such as Low-Rank Adaptation [121], IA3 [122], etc., have been introduced to reduce these computational costs. As the size of KGs, CGs and LLMs grows, improving the efficiency of these integration approaches becomes essential to maintain scalability.

2) Real-Time Inference: Incorporating KGs and CGs into real-time inference introduces additional complexity, which can lead to latency, especially when the model needs to access information from these graphs to construct prompts. Querying large-scale graphs during inference is resource-intensive and can slow down response times, making real-time applications less feasible [30]. The challenge is to efficiently query and process structured data without causing delays. To maintain responsiveness at scale, optimizing graph query mechanisms and minimizing the overhead of integrating KG and CG knowledge with LLM outputs are essential [123].

C. Graph Querying Costs

When querying large graphs, the computational cost can quickly escalate due to the vast number of nodes, edges, and relationships that must be processed and retrieved. As the size of KGs or CGs expands, querying and retrieving information efficiently without overwhelming computational resources becomes increasingly challenging [124]. Efficient query optimization techniques are critical for maintaining scalability, reducing query times, and addressing the challenges posed by the scale and complexity of highly interconnected graphs [125], [126]. More advanced solutions are needed to manage the growing size and intricate relationships within these graph structures.

D. Knowledge Updating

Keeping LLMs, KGs, and CGs aligned with new information as it emerges is complex [108]. Updating LLMs usually requires fine-tuning, which is computationally expensive since knowledge is embedded in their parameters. While KGs and CGs are easier to update with new facts, scaling KGs and handling large amounts of data in CGs can also become costly [42]. Another challenge is maintaining consistency between these systems as they are updated [35], [108]. Real-time updates for both graphs and LLMs are still in the early stages of development and need more efficient solutions to manage this process effectively [127].

E. Domain-Specific Adaptation

LLMs are generally trained for broad language understanding, while KGs and CGs are often tailored to specific domains, such as healthcare. Adapting LLMs to effectively utilize domain-specific knowledge from KGs and CGs, without losing their general versatility, is challenging. This requires advanced fine-tuning and prompting strategies that allow the model to incorporate specialized medical knowledge while still handling a wide range of queries [128]–[130].

References

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[3] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

[4] Jakob Nikolas Kather, Dyke Ferber, Isabella C Wiest, Stephen Gilbert, and Daniel Truhn. Large language models could make natural language again the universal interface of healthcare. *Nature Medicine*, pages 1–3, 2024.

[5] Hao Yang, Jiaxi Li, Siru Liu, Lei Du, Xiali Liu, Yong Huang, Qingke Shi, and Jialin Liu. Exploring the potential of large language models in personalized diabetes treatment strategies. *medRxiv:2023.06.30.23292034*, pages 2023–06, 2023.

[6] Chengsheng Mao, Jie Xu, Luke Rasmussen, Yikuan Li, Prakash Adekkanattu, Jennifer Pacheco, Borna Bonakdarpour, Robert Vassar, Li Shen, Guoqian Jiang, et al. Ad-bert: Using pre-trained language model to predict the progression from mild cognitive impairment to alzheimer's disease. *Journal of Biomedical Informatics*, 144:104442, 2023.

[7] Sara Montagna, Stefano Ferretti, Lorenz Cuno Klopfenstein, Antonio Florio, and Martino Francesco Pengo. Data decentralisation of llm-based chatbot systems in chronic disease self-management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, pages 205–212, 2023.

[8] SR Ali, TD Dobbs, HA Hutchings, and IS Whitaker. Using chatgpt to write patient clinic letters. lancet digit health 5, e179–e181, 2023.

[9] Li Wang, Xi Chen, XiangWen Deng, Hao Wen, MingKe You, WeiZhi Liu, Qi Li, and Jian Li. Prompt engineering in consistency and reliability with the evidence-based guideline for llms. *NPJ Digital Medicine*, 7(1):41, 2024.

[10] Thomas Savage, Ashwin Nayak, Robert Gallo, Ekanath Rangan, and Jonathan H Chen. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *NPJ Digital Medicine*, 7(1):20, 2024.

[11] Jason Holmes, Zhengliang Liu, Lian Zhang, Yuzhen Ding, Terence T Sio, Lisa A McGee, Jonathan B Ashman, Xiang Li, Tianming Liu, Jiajian Shen, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. *Frontiers in Oncology*, 13, 2023.

[12] Valentin Liévin, Christoffer Egeberg Hother, Andreas Geert Motzfeldt, and Ole Winther. Can large language models reason about medical questions? *Patterns*, 5(3), 2024.

[13] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

[14] Stephen E Robertson and Steve Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pages 232–241. Springer, 1994.

[15] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen Tau Yih. Dense passage retrieval for open-domain question answering. In *2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020*, pages 6769–6781. Association for Computational Linguistics (ACL), 2020.

[16] Devendra Singh Sachan, Mike Lewis, Dani Yogatama, Luke Zettlemoyer, Joelle Pineau, and Manzil Zaheer. Questions are all you need to train a dense passage retriever. *Transactions of the Association for Computational Linguistics*, 11:600–616, 2023.

[17] Omar Khattab and Matei Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 39–48, 2020.

[18] Shunyu Yao, Jeffrey Zhao, Dian Yu, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022.

[19] Manas Gaur and Amit Sheth. Building trustworthy neurosymbolic ai systems: Consistency, reliability, explainability, and safety. *AI Magazine*, 45(1):139–155, 2024.

[20] Ping Yu, Hua Xu, Xia Hu, and Chao Deng. Leveraging generative ai and large language models: a comprehensive roadmap for healthcare integration. In *Healthcare*, volume 11, page 2776. MDPI, 2023.

[21] Alaa Abd-Alrazaq, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Padraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damseh, Sadam Alabed Alrazak, Javaid Sheikh, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Medical Education*, 9(1):e48291, 2023.

[22] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. Large language models in medicine. *Nature medicine*, 29(8):1930–1940, 2023.

[23] Jenelle A Jindal, Matthew P Lungren, and Nigam H Shah. Ensuring useful adoption of generative artificial intelligence in healthcare. *Journal of the American Medical Informatics Association*, 31(6):1441–1444, 2024.

[24] Jesutofunmi A Omiye, Haiwen Gui, Shawheen J Rezaei, James Zou, and Roxana Daneshjou. Large language models in medicine: the potentials and pitfalls: a narrative review. *Annals of Internal Medicine*, 177(2):210–220, 2024.

[25] Scott Gottlieb and Lauren Silvis. How to safely integrate large language models into health care. In *JAMA Health Forum*, volume 4, pages e233909–e233909. American Medical Association, 2023.

[26] Sicen Liu, Xiaolong Wang, Yongshuai Hou, Ge Li, Hui Wang, Huiqin Xu, Yang Xiang, and Buzhou Tang. Multimodal data matters: Language model pre-training over structured and unstructured electronic health records. *IEEE Journal of Biomedical and Health Informatics*, 27:504–514, 2022.

[27] Dongdong Zhang, Changchang Yin, Jucheng Zeng, Xiaohui Yuan, and Ping Zhang. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Medical Informatics and Decision Making*, 20, 2020.

[28] Jiri Hron, Laura A Culp, Gamaleldin Fathy Elsayed, Rosanne Liu, Jasper Snoek, Simon Kornblith, Alex Rizkowsky, Isabelle Simpson, Jascha Sohl-Dickstein, Noah Fiedel, Aaron T Parisi, Alexander A Alemi, Azade Nova, Ben Adlam, Bernd Bohnet, Gaurav Mishra, Hanie Sedghi, Izzeddin Gur, Jaehoon Lee, John D Co-Reyes, Kathleen Kenealy, Kelvin Xu, Kevin Swersky, Igor Mordatch, Lechao Xiao, Maxwell Bileschi, Peter J Liu, Roman Novak, Sharad Vikram, Tris Warkentin, and Jeffrey Pennington. Training language models on the knowledge graph: Insights on hallucinations and their detectability. In *First Conference on Language Modeling*, 2024.

[29] Zhongqi Yang, Elahe Khatibi, Nitish Nagesh, Mahyar Abbasian, Iman Azimi, Ramesh Jain, and Amir M Rahmani. Chatdiet: Empowering personalized nutrition-oriented food recommender chatbots through an llm-augmented framework. *Smart Health*, 32:100465, 2024.

[30] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Chatgpt is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling. *arXiv preprint arXiv:2306.11489*, 2023.

[31] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[32] Hao Fei, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in bioinformatics*, 22(3):bbaa110, 2021.

[33] Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel Ni, Heung-Yeung Shum, and Jian Guo. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In *The Twelfth International Conference on Learning Representations*, 2024.

[34] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

[35] Yanjun Gao, Ruizhe Li, John Caskey, Dmitriy Dligach, Timothy Miller, Matthew M Churpek, and Majid Afshar. Leveraging a medical knowledge graph into large language models for diagnosis prediction. *arXiv preprint arXiv:2308.14321*, 2023.

[36] Amelie Gyrard, Seyedali Mohammadi, Manas Gaur, and Antonio Kung. Iot-based preventive mental health using knowledge graphs and standards for better well-being. *arXiv preprint arXiv:2406.13791*, 2024.

[37] Oshani Seneviratne, Amar K Das, Shruthi Chari, Nkechinyere N Agu, Sabbir M Rashid, Jamie McCusker, Jade S Franklin, Miao Qi, Kristin P Bennett, Ching-Hua Chen, et al. Semantically enabling clinical decision support recommendations. *Journal of Biomedical Semantics*, 14(1):8, 2023.

[38] Shruthi Chari. An ontology-enabled approach for user-centered and knowledge-enabled explanations of ai systems. In *DC@ ISWC*, 2023.

[39] Dieter Fensel, Umutcan Şimşek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, Alexander Wahler, Dieter Fensel, et al. Introduction: what is a knowledge graph?

Knowledge graphs: Methodology, tools and selected use cases, pages 1–10, 2020.

[40] Bilal Abu-Salih, Muhammad Al-Qurishi, Mohammed Alweshah, Mohammad Al-Smadi, Reem Alfayez, and Heba Saadeh. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data*, 10(1):81, 2023.

[41] Xiaohui Tao, Thuan Pham, Ji Zhang, Jianming Yong, Wee Pheng Goh, Wenping Zhang, and Yi Cai. Mining health knowledge graph for health risk prediction. *World Wide Web*, 23:2341–2362, 2020.

[42] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1):5994, 2017.

[43] Judea Pearl. Causal inference. *Causality: objectives and assessment*, pages 39–58, 2010.

[44] A. Akinkugbe, Sonia Sharma, R. Ohrbach, G. Slade, and Charles Poole. Directed acyclic graphs for oral disease research. *Journal of Dental Research*, 95:853 – 859, 2016.

[45] M. Etminan, G. Collins, and M. Mansournia. Using causal diagrams to improve the design and interpretation of medical research. *Chest*, 158 1S:S21–S28, 2020.

[46] Mahyar Abbasian, Elahe Khatibi, Iman Azimi, David Oniani, Zahra Shakeri Hossein Abad, Alexander Thieme, Ram Sriram, Zhongqi Yang, Yanshan Wang, Bryant Lin, et al. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative ai. *NPJ Digital Medicine*, 7(1):82, 2024.

[47] Avishek Choudhury and Zaira Chaudhry. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. *Journal of Medical Internet Research*, 26:e56764, 2024.

[48] Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hanguang Li. Trustworthy llms: a survey and guideline for evaluating large language models' alignment. *ArXiv*, abs/2308.05374, 2023.

[49] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A Hale. The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nature Machine Intelligence*, pages 1–10, 2024.

[50] Xitong Guo, Yongqiang Sun, Jinqiao Yuan, Ziyu Yan, and Nan Wang. Privacy-personalization paradox in adoption of mobile health service: The mediating role of trust. In *16th Pacific Asia Conference on Information Systems, PACIS 2012*. Pacific Asia Conference on Information Systems, 2012.

[51] Stephen Armstrong. Data, data everywhere: the challenges of personalised medicine. *BMJ*, 359, 2017.

[52] E. Cherif, Nora Bezaz, and M. Mzoughi. Do personal health concerns and trust in healthcare providers mitigate privacy concerns? effects on patients' intention to share personal health data on electronic health records. *Social science & medicine*, 283:114146, 2021.

[53] Catherine L. Anderson and Ritu Agarwal. The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information. *Inf. Syst. Res.*, 22:469–490, 2011.

[54] Xudong Pan, Mi Zhang, S. Ji, and Min Yang. Privacy risks of general-purpose language models. *2020 IEEE Symposium on Security and Privacy (SP)*, pages 1314–1331, 2020.

[55] Laura Weidinger, John F. J. Mellor, M. Rauh, C. Griffin, J. Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, B. Balle, A. Kasirzadeh, Z. Kenton, S. Brown, W. Hawkins, T. Stepleton, C. Biles, A. Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William S. Isaac, Sean Legassick, G. Irving, and Iason Gabriel. Ethical and social risks of harm from language models. *ArXiv*, abs/2112.04359, 2021.

[56] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. Madai. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20, 2020.

[57] Shaoxiong Ji, Tianlin Zhang, Kailai Yang, Sophia Ananiadou, and Erik Cambria. Rethinking large language models in mental health applications. *ArXiv*, abs/2311.11267, 2023.

[58] Oshani Seneviratne, Amar K Das, Shruthi Chari, Nkechinyere Agu, Sabbir Rashid, Ching-Hua Chen, Jamie McCusker, James A Hendler, and Deborah L McGuinness. Enabling trust in clinical decision support recommendations through semantics. CEUR Workshop Proceedings (CEUR-WS. org), 2019.

[59] Deborah L. McGuinness Sola S. Shirai, Oshani Seneviratne. Causal inference. *Workshop on The Personal Health Knowledge Graph, Knowledge Graph Conference*, 2020.

[60] Oshani Seneviratne, Jonathan Harris, Ching-Hua Chen, and Deborah L McGuinness. Personal health knowledge graph for clinically relevant diet recommendations. *arXiv preprint arXiv:2110.10131*, 2021.

[61] Ching-Hua Chen, Daniel Gruen, Jonathan Harris, James Hendler, Deborah L McGuinness, Marco Monti, Nidhi Rastogi, Oshani Seneviratne, and Mohammed J Zaki. Semantic technologies for clinically relevant personal health applications. In *Personal Health Informatics: Patient Participation in Precision Health*, pages 199–220. Springer, 2022.

[62] Shruthi Chari, Prasant Acharya, Daniel M Gruen, Olivia Zhang, Elif K Eyigoz, Mohamed Ghalwash, Oshani Seneviratne, Fernando Suarez Saiz, Pablo Meyer, Prithwish Chakraborty, et al. Informing clinical assessment by contextualizing post-hoc explanations of risk prediction models in type-2 diabetes. *Artificial Intelligence in Medicine*, 137:102498, 2023.

[63] Jade S Franklin, Karan Bhanot, Mohamed Ghalwash, Kristin P Bennett, Jamie McCusker, and Deborah L McGuinness. An ontology for fairness metrics. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 265–275, 2022.

[64] Jade S Franklin, Hannah Powers, John S Erickson, Jamie McCusker, Deborah L McGuinness, and Kristin P Bennett. An ontology for reasoning about fairness in regression and machine learning. In *Iberoamerican Knowledge Graphs and Semantic Web Conference*, pages 243–261. Springer, 2023.

[65] Priyanshul Govil, Vamshi Krishna Bonagiri, Manas Gaur, Ponnurangam Kumaraguru, and Sanorita Dey. Cobias: Contextual reliability in bias assessment. *arXiv preprint arXiv:2402.14889*, 2024.

[66] Richard J Chen, Judy J Wang, Drew FK Williamson, Tiffany Y Chen, Jana Lipkova, Ming Y Lu, Sharifa Sahai, and Faisal Mahmood. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature biomedical engineering*, 7(6):719–742, 2023.

[67] I. Straw and Chris Callison-Burch. Artificial intelligence in mental health and the biases of language based models. *PLoS ONE*, 15, 2020.

[68] Z. Obermeyer, Brian W. Powers, C. Vogeli, and S. Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366:447 – 453, 2019.

[69] Eli M. Cahan, T. Hernandez-Boussard, Sonoo Thadaney-Israni, and D. Rubin. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digital Medicine*, 2, 2019.

[70] S. Reijneveld and K. Stronks. The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data. *International journal of epidemiology*, 28 6:1134–40, 1999.

[71] Cai Long, Deepak Subburam, Kayle Lowe, André Dos Santos, Jessica Zhang, Sang Hwang, Neil Saduka, Yoav Horev, Tao Su, David WJ Côté, et al. Chatent: Augmented large language model for expert knowledge retrieval in otolaryngology–head and neck surgery. *Otolaryngology–Head and Neck Surgery*, 2024.

[72] Guangyu Wang, Guoxing Yang, Zongxin Du, Longjun Fan, and Xiaohu Li. Clinicalgpt: large language models finetuned with diverse medical data and comprehensive evaluation. *arXiv preprint arXiv:2306.09968*, 2023.

[73] Di Jin, Zhijing Jin, Joey Tianyi Zhou, and Peter Szolovits. Is bert really robust? a strong baseline for natural language attack on text classification and entailment. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8018–8025, 2020.

[74] Deepa Tilwani, Yash Saxena, Ali Mohammadi, Edward Raff, Amit Sheth, Srinivasan Parthasarathy, and Manas Gaur. Reasons: A benchmark for retrieval and automated citations of scientific sentences using public and proprietary llms. *arXiv preprint arXiv:2405.02228*, 2024.

[75] Lichao Sun, Yue Huang, Haoran Wang, Siyuan Wu, Qihui Zhang, Chujie Gao, Yixin Huang, Wenhan Lyu, Yixuan Zhang, Xiner Li, et al. Trustllm: Trustworthiness in large language models. *arXiv preprint arXiv:2401.05561*, 2024.

[76] R. Snyderman. Personalized health care: From theory to practice. *Biotechnology Journal*, 7, 2012.

[77] A. Sheth, Utkarshani Jaimini, K. Thirunarayan, and Tanvi Banerjee. Augmented personalized health: How smart data with iots and ai is about to change healthcare. *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)*, pages 1–6, 2017.

[78] Najma Taimoor and Semeen Rehman. Reliable and resilient ai and iot-based personalised healthcare services: A survey. *IEEE Access*, PP:1–1, 2022.

[79] Justin Cosentino, Anastasiya Belyaeva, Xin Liu, Nicholas A Furlotte, Zhun Yang, Chace Lee, Erik Schenck, Yojan Patel, Jian Cui, Logan Douglas Schneider, et al. Towards a personal health large language model. *arXiv preprint arXiv:2406.06474*, 2024.

[80] Mike A Merrill, Akshay Paruchuri, Naghmeh Rezaei, Geza Kovacs, Javier Perez, Yun Liu, Erik Schenck, Nova Hammerquist, Jake Sunshine, Shyam Tailor, et al. Transforming wearable data into health insights using large language model agents. *arXiv preprint arXiv:2406.06464*, 2024.

[81] Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866*, 2024.

[82] Mahyar Abbasian, Iman Azimi, Amir M Rahmani, and Ramesh Jain. Conversational health agents: A personalized llm-powered agent framework. *arXiv preprint arXiv:2310.02374*, 2023.

[83] Ajan Subramanian, Zhongqi Yang, Iman Azimi, and Amir M Rahmani. Graph-augmented llms for personalized health insights: A case study in sleep analysis. *arXiv preprint arXiv:2406.16252*, 2024.

[84] Manas Gaur, Keyur Faldu, and Amit Sheth. Semantics of the blackbox: Can knowledge graphs help make deep learning systems more interpretable and explainable? *IEEE Internet Computing*, 25(1):51–59, 2021.

[85] Nirmal Kumar Sivaraman, Manas Gaur, Shivansh Baijal, Sakthi Balan Muthiah, and Amit Sheth. Exo-sir: an epidemiological model to analyze the impact of exogenous spread of infection. *International Journal of Data Science and Analytics*, pages 1–16, 2022.

[86] Deborah L McGuinness, Li Ding, Tim Lebo, James P McCusker, Abdul R Shaikh, Glen D Morgan, Gordon Willis, Richard P Moser, Zaria Tatalovich, Bradford W Hesse, et al. Towards semantically enabled next generation community health information portals: The popscigrid pilot. In *2012 45th Hawaii International Conference on System Sciences*, pages 2752–2760. IEEE, 2012.

[87] Amelie Gyrard, Manas Gaur, Saeedeh Shekarpour, Krishnaprasad Thirunarayan, and Amit Sheth. Personalized health knowledge graph. In *CEUR workshop proceedings*, volume 2317. NIH Public Access, 2018.

[88] Utkarshani Jaimini and Amit Sheth. Causalkg: Causal knowledge graph explainability using interventional and counterfactual reasoning. *IEEE Internet Computing*, 26(1):43–50, 2022.

[89] Manas Gaur, Ankit Desai, Keyur Faldu, and Amit Sheth. Explainable ai using knowledge graphs. In *ACM CoDS-COMAD Conference*, 2020.

[90] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

[91] Tianyu Han, Lisa C Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K Bressem. Medalpaca–an open-source collection of medical conversational ai models and training data. *arXiv preprint arXiv:2304.08247*, 2023.

[92] Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. Huatuo: Tuning llama model with chinese medical knowledge. *arXiv preprint arXiv:2304.06975*, 2023.

[93] Zeming Chen, Alejandro Hernández Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, et al. Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*, 2023.

[94] Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. Biogpt: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6):bbac409, 2022.

[95] Sonish Sivarajkumar, Mark Kelley, Alyssa Samolyk-Mazzanti, Shyam Visweswaran, and Yanshan Wang. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. *JMIR Medical Informatics*, 12:e55318, 2024.

[96] Yao Ge, Yuting Guo, Sudeshna Das, Mohammed Ali Al-Garadi, and Abeed Sarker. Few-shot learning for medical text: A review of advances, trends, and opportunities. *Journal of Biomedical Informatics*, page 104458, 2023.

[97] Jing Miao, Charat Thongprayoon, Supawadee Suppadungsuk, Pajaree Krisanapan, Yeshwanter Radhakrishnan, and Wisit Cheungpasitporn. Chain of thought utilization in large language models and application in nephrology. *Medicina*, 60(1):148, 2024.

[98] Atmika Gorti, Manas Gaur, and Aman Chadha. Unboxing occupational bias: Grounded debiasing llms with us labor data. *arXiv preprint arXiv:2408.11247*, 2024.

[99] Hao Peng, Haoran Li, Yangqiu Song, V. Zheng, and Jianxin Li. Differentially private federated knowledge graphs embedding. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021.

[100] Hao Peng, Haoran Li, Yangqiu Song, V. Zheng, and Jianxin Li. Federated knowledge graphs embedding. *ArXiv*, abs/2105.07615, 2021.

[101] Jianwei Qian, Xiang-Yang Li, Chunhong Zhang, and Linlin Chen. De-anonymizing social networks and inferring private attributes using knowledge graphs. *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, 2016.

[102] Shruthi Chari, Daniel M Gruen, Oshani Seneviratne, and Deborah L McGuinness. Foundations of explainable knowledge-enabled systems. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, pages 23–48. IOS Press, 2020.

[103] A. Markus, J. Kors, and P. Rijnbeek. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of biomedical informatics*, page 103655, 2020.

[104] Linfeng Li, Peng Wang, Jun Yan, Yao Wang, Simin Li, Jinpeng Jiang, Zhe Sun, Buzhou Tang, Tsung-Hui Chang, Shenghui Wang, et al. Real-world data medical knowledge graph: construction and applications. *Artificial intelligence in medicine*, 103:101817, 2020.

[105] Enayat Rajabi and Somayeh Kafaie. Knowledge graphs and explainable ai in healthcare. *Information*, 13(10):459, 2022.

[106] P. Ping, K. Watson, Jiawei Han, and A. Bui. Individualized knowledge graph: A viable informatics path to precision medicine. *Circulation Research*, 120:1078–1080, 2017.

[107] Sola S. Shirai, O. Seneviratne, and D. McGuinness. Applying personal knowledge graphs to health. *ArXiv*, abs/2104.07587, 2021.

[108] Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Chen Ling, Joyce Ho, et al. A survey on knowledge graphs for healthcare: Resources, applications, and promises. *arXiv preprint arXiv:2306.04802*, 2023.

[109] Yu-Neng Chuang, Kwei-Herng Lai, Ruixiang Tang, Mengnan Du, Chia-Yuan Chang, Na Zou, and Xia Hu. Fair-rgnn: Mitigating relational bias on knowledge graphs. *ACM Transactions on Knowledge Discovery from Data*, 2024.

[110] Harry Shomer, Wei Jin, Wentao Wang, and Jiliang Tang. Toward degree bias in embedding-based knowledge graph completion. *Proceedings of the ACM Web Conference 2023*, 2023.

[111] Mario Arduini, Lorenzo Noci, Federico Pirovano, Ce Zhang, Y. Shrestha, and B. Paudel. Adversarial learning for debiasing knowledge graph embeddings. *ArXiv*, abs/2006.16309, 2020.

[112] Fan Gong, Meng Wang, Haofen Wang, Sen Wang, and Mengyue Liu. Smr: medical knowledge graph embedding for safe medicine recommendation. *Big Data Research*, 23:100174, 2021.

[113] Yu Wang, Nedim Lipka, Ryan A Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. Knowledge graph prompting for multi-document question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19206–19214, 2024.

[114] Yijun Tian, Huan Song, Zichen Wang, Haozhu Wang, Ziqing Hu, Fang Wang, Nitesh V Chawla, and Panpan Xu. Graph neural prompting with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19080–19088, 2024.

[115] M. Lean, W. Leslie, A. Barnes, N. Brosnahan, G. Thom, L. McCombie, C. Peters, S. Zhyzhneuskaya, A. Al-Mrabeh, K. Hollingsworth, Angela M Rodrigues, L. Rehackova, A. Adamson, F. Sniehotta, J. Mathers, H. Ross, Y. McIlvenna, R. Stefanetti, M. Trenell, P. Welsh, S. Kean, I. Ford, A. McConnachie, N. Sattar, and Roy Taylor. Primary care-led weight management for remission of type 2 diabetes (direct): an open-label, cluster-randomised trial. *The Lancet*, 391:541–551, 2017.

[116] J. Gagliardino, J. Chantelot, C. Domenger, Ambady Ramachandran, G. Kaddaha, J. Mbanya, M. Shestakova, and J. Chan. Impact of diabetes education and self-management on the quality of care for people with type 1 diabetes mellitus in the middle east (the international

diabetes mellitus practices study, idmps). *Diabetes research and clinical practice*, 147:29–36, 2019.

[117] J. Nagelkerk, K. Reick, and Leona Meengs. Perceived barriers and effective strategies to diabetes self-management. *Journal of advanced nursing*, 54 2:151–8, 2006.

[118] Renate Jansink, J. Braspenning, T. van der Weijden, L. Niessen, G. Elwyn, and R. Grol. Nurse-led motivational interviewing to change the lifestyle of patients with type 2 diabetes (mild-project): protocol for a cluster, randomized, controlled trial on implementing lifestyle recommendations. *BMC Health Services Research*, 9:19 – 19, 2009.

[119] Yanai Elazar, Nora Kassner, Shauli Ravfogel, Amir Feder, Abhilasha Ravichander, Marius Mosbach, Yonatan Belinkov, Hinrich Schütze, and Yoav Goldberg. Measuring causal effects of data statistics on language model'sfactual'predictions. *arXiv preprint arXiv:2207.14251*, 2022.

[120] Jeff Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. Large Language Models and Knowledge Graphs: Opportunities and Challenges. *Transactions on Graph Data and Knowledge*, June 2023.

[121] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[122] Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin A Raffel. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems*, 35:1950–1965, 2022.

[123] Giuseppe Futia and Antonio Vetrò. On the integration of knowledge graphs into deep learning models for a more comprehensible ai—three challenges for future research. *Information*, 11(2):122, 2020.

[124] Wenfei Fan, Jianzhong Li, Xin Wang, and Yinghui Wu. Query preserving graph compression. In *Proceedings of the 2012 ACM SIGMOD international conference on management of data*, pages 157–168, 2012.

[125] Isabelle Stanton and Gabriel Kliot. Streaming graph partitioning for large distributed graphs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1222–1230, 2012.

[126] Peixiang Zhao and Jiawei Han. On graph query optimization in large networks. *Proceedings of the VLDB Endowment*, 3(1-2):340–351, 2010.

[127] Xubin Ren, Jiabin Tang, Dawei Yin, Nitesh Chawla, and Chao Huang. A survey of large language models for graphs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6616–6626, 2024.

[128] Pengcheng Jiang, Cao Xiao, Adam Richard Cross, and Jimeng Sun. Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. In *The Twelfth International Conference on Learning Representations*, 2023.

[129] Ran Xu, Hejie Cui, Yue Yu, Xuan Kan, Wenqi Shi, Yuchen Zhuang, May Dongmei Wang, Wei Jin, Joyce Ho, and Carl Yang. Knowledge-infused prompting: Assessing and advancing clinical text data generation with large language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 15496–15523, 2024.

[130] Marvin Hofer, Daniel Obraczka, Alieh Saeedi, Hanna Köpcke, and Erhard Rahm. Construction of knowledge graphs: Current state and challenges. *Information*, 15(8):509, 2024.