

# HealthGenie: Empowering Users with Healthy Dietary Guidance through Knowledge Graph and Large Language Models

Fan Gao  
The University of Tokyo  
Tokyo, Japan  
fangao0802@g.ecc.u-tokyo.ac.jp

Zhongyi Zhou  
The University of Tokyo  
Tokyo, Japan  
zhou.zhongyi@tc.u-tokyo.ac.jp

Hang Jiang  
MIT Media Lab  
Cambridge, Massachusetts, USA  
hjian42@mit.edu

Xinjie Zhao  
The University of Tokyo  
Tokyo, Japan  
xinjie-zhao@g.ecc.u-tokyo.ac.jp

Rui Yang  
Duke-NUS Medical School  
Singapore, Singapore  
yang.rui@duke-nus.edu.sg

Chanjun Park  
Korea University  
Seoul, South Korea  
bcj1210@naver.com

Ding Xia  
The University of Tokyo  
Tokyo, Japan  
dingxia1995@gmail.com

Jinghui Lu  
Smarter Inc.  
Shanghai, China  
contact@smarter.ai

Irene Li  
The University of Tokyo  
Tokyo, Japan  
bkjl6178@g.ecc.u-tokyo.ac.jp

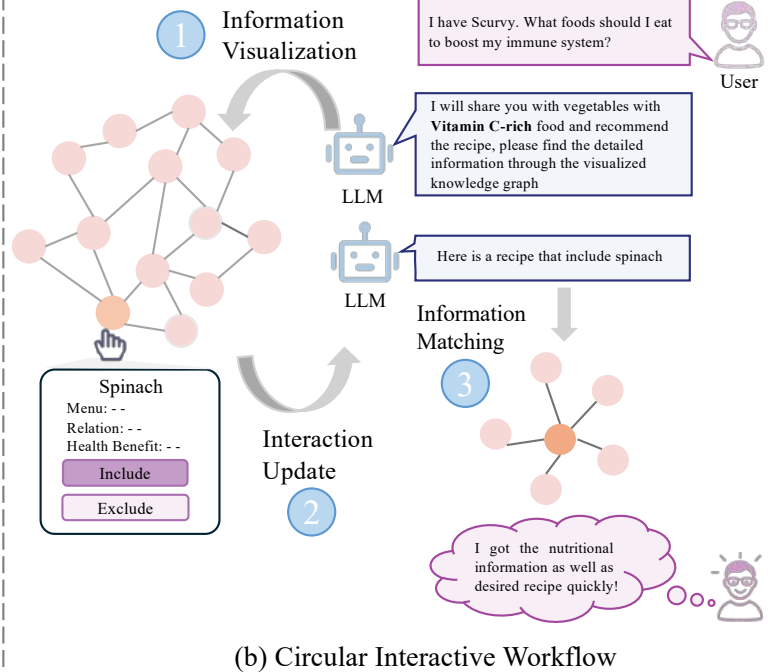
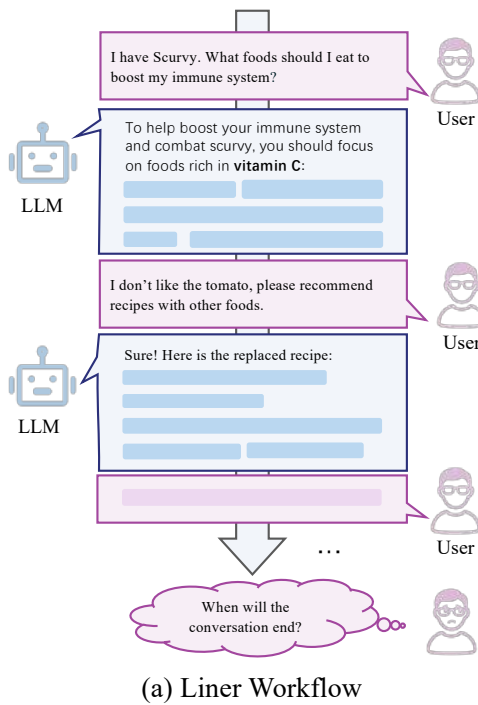


Figure 1: Unlike the traditional linear workflow (a) of LLMs, *HealthGenie* offers dietary guidance through interactive knowledge graph visualizations within a circular, interactive workflow (b), allowing users to quickly gain an overview of desired information.

## Abstract

Seeking dietary guidance often requires navigating complex professional knowledge while accommodating individual health conditions. Knowledge Graphs (KGs) offer structured and interpretable nutritional information, whereas Large Language Models (LLMs) naturally facilitate conversational recommendation delivery. In this paper, we present *HealthGenie*, an interactive system that combines the strengths of LLMs and KGs to provide personalized dietary recommendations along with hierarchical information visualization for a quick and intuitive overview. Upon receiving a user query, *HealthGenie* performs query refinement and retrieves relevant information from a pre-built KG. The system then visualizes and highlights pertinent information, organized by defined categories, while offering detailed, explainable recommendation rationales. Users can further tailor these recommendations by adjusting preferences interactively. Our evaluation, comprising a within-subject comparative experiment and an open-ended discussion, demonstrates that *HealthGenie* effectively supports users in obtaining personalized dietary guidance based on their health conditions while reducing interaction effort and cognitive load. These findings highlight the potential of LLM-KG integration in supporting decision-making through explainable and visualized information. We examine the system’s usefulness and effectiveness with an N=12 within-subject study and provide design considerations for future systems that integrate conversational LLM and KG.

## CCS Concepts

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

## Keywords

Knowledge Graphs, Large Language Models, Nutrition, Health, Interactive Systems, Human-Computer Interaction

### ACM Reference Format:

Fan Gao, Xinjie Zhao, Ding Xia, Zhongyi Zhou, Rui Yang, Jinghui Lu, Hang Jiang, Chanjun Park, and Irene Li. 2025. *HealthGenie: Empowering Users with Healthy Dietary Guidance through Knowledge Graph and Large Language Models*. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 16 pages. <https://doi.org/XXXXXXX.XXXXXXX>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, Woodstock, NY

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06

<https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

Making informed dietary choices plays a crucial role in managing personal health and preventing chronic diseases. To make dietary information accessible and interpretable, research has focused on structuring this complex knowledge in an organized form — Knowledge Graphs (KGs), which connect entities such as foods, nutrients, and health conditions through semantically meaningful relationships [1, 12, 19, 20]. At the same time, the emergence of Large Language Models (LLMs) has revolutionized information-seeking behavior by offering interactive, conversational interfaces [49, 50, 74]. The combination of LLMs with KGs enhances both the usability of structured dietary knowledge and enables more intuitive, intelligent exploration of complex nutritional relationships [23, 38].

While integrating KGs can greatly enhance the quality of LLMs’ outputs greatly, this benefit is underutilized with current text-based interfaces, which usually rely on linear text formats [7], such as extended paragraphs, and do not take advantage of the inherent structure of knowledge graphs by providing (i) visualized output [37, 39] and (ii) direct interaction [11, 27, 55]. First, existing LLM-based methods are capable of providing more detailed information through knowledge graphs [64, 65, 70]. However, these text-based outputs are often too verbose, preventing users from easily focusing on the information that interests them and increases the time and effort required to identify desired content. For example, comparing the purposes of suggested dishes in a textual context is challenging and calls for visualizing relationships and unique features of each dish to assist users in making decisions. Second, in terms of direct interaction, users often need to engage in multiple rounds of conversation to eventually obtain their preferred recipes because text-based exchanges require them to independently determine their needs and input lengthy descriptions, which is time-consuming and inefficient. Therefore, developing a new interactive interface that facilitates decision-making and offers an intuitive method is critical. Combining KGs as the interface may provide a viable solution for dietary recommendations [46, 69, 71]. Such a combination can foster the advantages of each—leveraging the natural language capabilities of LLMs while retaining the structural clarity and expressiveness of KGs. In Figure 1, we illustrate an example comparing the conventional linear text-based approach (a) with our circular, knowledge-graph-driven interactive workflow (b). The left side shows how users typically must go through multiple rounds of text-based queries for dietary recommendations, while the right side demonstrates how visualizing and directly interacting with the knowledge graph can help them more efficiently refine or exclude certain ingredients in real time.

In this paper, we propose *HealthGenie* — a novel interactive system that synergizes LLM-powered conversational interfaces with recipe-specific knowledge graph visualizations. *HealthGenie* is designed to empower everyday users by providing clear, evidence-based dietary recommendations. Drawing on insights from a formative study involving seven participants, our work addresses critical needs such as information transparency, reduced cognitive load, intuitive interaction, and personalized guidance. *HealthGenie* utilizes a circular workflow to enhance the LLM-KG-User interaction. Through this approach, the LLM identifies individual user

preferences based on their interactions with the KG, dynamically incorporating these insights into its recommendations, which are then visualized back into the KG. By grounding LLM responses in a curated nutritional knowledge graph and actively offering intuitive visual guidance, *HealthGenie* enhances both the credibility and usability of dietary recommendations, ultimately supporting more informed recipe choices.

To evaluate how effectively *HealthGenie* facilitates knowledge graph visualization and intuitive interaction with LLMs for personalized dietary recommendation tasks, we conducted a within-subject evaluation with 12 experienced LLM users. We demonstrate the functionality of *HealthGenie* in four scenarios and guide users to complete the related tasks with our system and baseline system in a counterbalanced order. Both quantitative and qualitative results demonstrate that our system effectively delivers well-organized recipe information through integrated textual and visual representations while actively supporting users' preference-based decision-making.

The main contributions of this work can be summarized as follows:

- A formative study (N=7) that summarizes practices, challenges, and expectations on dietary information display.
- *HealthGenie*, an interactive system featuring visualized KG response and intuitive interaction, empowering users in personalized dietary exploration.
- An empirical user study (N=12) on the usefulness and effectiveness of *HealthGenie* in providing recipe recommendations and a discussion of insights derived from the user study for future LLM-KG integrated interface design.

## 2 Background & Related Work

In this section, we review prior work on the role of Artificial Intelligence (AI) in healthcare support systems, the current state of LLM-based conversational interfaces, the design of knowledge graph visualization and interaction—particularly within the Human-Computer Interaction (HCI) research community—as well as the limitations and challenges of LLM-driven nutrition and dietary guidance.

### 2.1 AI-powered Healthcare Support Systems

Traditional healthcare support systems focus on designing and implementing interactive technologies that provide users with the information they need to make informed decisions and manage their well-being [75]. Early work shows that users commonly rely on search engines (e.g., Google and Yahoo) or social media platforms such as Quora and Reddit to obtain health information [25, 44]. To address the need for professional healthcare consultation, more advanced AI tools have been explored. For example, Joshi et al. [28] developed an interactive voice response system to support individuals living with AIDS by providing healthcare tips. Similarly, Zhang [75] focused on the design of AI-driven interfaces to assist users in accessing healthcare information and consultations.

In recent years, large language models (LLMs) with task-agnostic architectures and extensive pre-training have seen wide application in both healthcare and nutrition [62]. Studies have suggested the potential of LLMs for providing informed decision support

throughout clinical care, from diagnosis to treatment recommendations [30, 51, 52]. Moreover, researchers have explored utilizing LLMs' superior natural language understanding and generation ability to engage in open-ended conversations with access to healthcare information [40, 50]. Despite the success of LLMs, the generative nature makes it unavoidable to result in hallucination, especially in the healthcare domain. To tackle this, Sachdeva et al. [53] developed *Build Your Own expert Bot platform* to create LLM-powered chatbots with integrated expert verification. These systems aim to provide more and more accurate information for non-expert users.

### 2.2 Large Language Model-Based Conversational Interface

In response to the rapid advancement of LLMs, the Human-Computer Interaction (HCI) community has been actively exploring innovative conversational interface design to improve user interaction with LLMs across diverse applications [8, 14, 73]. These efforts can be broadly categorized into two key research directions: (1) enhancing user comprehension of LLM-generated content and (2) leveraging LLM's open-ended conversational capabilities to provide task-specific guidance.

A prominent research focus involves designing systems that help users better understand and interpret LLM outputs. For example, *Graphlogue* [27] converts text-based responses from LLM into graphical diagrams, facilitating information search and question-answering tasks. Similarly, *WatiGPT* [68] proposed to empower users with enhanced comprehension and augmented control over analysis conducted by LLMs. *FathomGPT* [31] supports interactive exploration of ocean data through intuitive visualizations. These systems aim to reduce cognitive load and improve transparency in human-LLM interactions.

Another key trend involves leveraging LLM's conversational flexibility to offer tailored assistance in specialized tasks. *Diary-Mate* [32] assists users with reflective journaling by providing LLM-generated writing prompts and feedback. *InkSync* [34] extends traditional chat-based conversational interfaces by integrating LLM-generated feedback during document editing, while *Compeer* [41] introduces a proactive conversational agent that provides adaptive peer support. These systems show how LLMs can be harnessed to specific user tasks effectively.

Building upon these advancements, *HealthGenie* aligns with interactive conversational systems that enhance user understanding of LLM-generated responses. Aimed at reducing conversational turns, *HealthGenie* offers insights on how LLMs can support personalized dietary or recipe recommendations.

### 2.3 Knowledge Graph Visualization and Interaction

Fundamentally, a knowledge graph is a data model that represents knowledge as a set of nodes (i.e., entities), edges (i.e., relations between entities), and properties (i.e., attributes) that can be associated with both nodes and edges [2]. Due to their robust structure for reasoning over data, knowledge graphs have attracted growing interest in the research community, particularly in exploring how they can be leveraged to enable more effective visualizations and interactive interfaces. Previous works have used KGs as common

knowledge ground method and design visual interfaces to directly view and manipulate structured data for both experts [36] and non-expert [6, 67, 76] users. For example, KGScope[22] supports interactive visual exploration of knowledge graphs by providing embedding-based guidance to help users derive insights and navigate the broader network. Their evaluation demonstrates that s can offer valuable information and facilitate comprehensive exploration. Similarly, Ashby et al.[4] leverage both KGs and LLMs to create personalized, context-aware conversational experiences.

In professional domains such as healthcare - where precision and comprehensive analysis are critical — KGs are especially valuable for offering hierarchical explanations of data. Santos et al. [54] presents an open-source platform based on clinical knowledge graphs and provides extendable node edit functions to facilitate user demands. While Xu et al. [69] and You et al. [71] have successfully integrated KGs and LLMs for health information seeking but tend to overlook the importance of incorporating user interaction feedback and provide personalized recommendations.

## 2.4 Large Language Models in Nutrition and Dietary Guidance

Recent research has also explored the diverse applications of LLMs in nutrition-related tasks. Studies have demonstrated their potential in areas ranging from dietary data analysis to generating personalized nutritional recommendations for specific health conditions [43, 56]. For example, some proposed applications envision LLMs as interactive tools that can answer nutrition-related questions in real-time, offering a readily accessible source of information for individuals seeking dietary advice [15, 29]. Moreover, *ChatDiet* [72] employs LLMs to create tailored meal plan that consider a user's specific dietary restrictions, preferences, and underlying health conditions.

Despite recognizing the potential of LLMs, several studies have raised concerns regarding the accuracy and reliability of the information these models provide [5, 45]. The need for rigorous validation of LLM-generated nutritional advice by human experts has been widely emphasized to ensure the safety and efficacy of these tools. For instance, in collaboration with registered dietitians, Szymanski et al. [60] developed The Food Product Nutrition Assistant, which generates detailed food product descriptions and personalized nutrition information. Another line of work explores the use of Retrieval-Augmented Generation (RAG) frameworks to improve LLMs' ability to interpret and apply established dietary guidelines [13].

By establishing a professional corpus as its knowledge base, *HealthGenie* retrieve relevant and recommend healthy dietary with keyword extraction. It is designed to help users quickly access desired recipes with minimal mental effort by prioritizing reliable, transparent, and visually clear information compared to traditional long texts that often display disorganized columns that compromise readability, reliability, and transparency [44].

## 3 Formative Study

We conducted a formative study to better understand users' needs for acquiring and exploring healthcare information and to identify

the challenges they face when interacting with current conversational LLM interfaces. The insights gained from this study inform the design goals of our system.

### 3.1 Setup

**Participants.** Seven participants (P1–P7) with diverse experiences in LLMs were recruited for the study. The group was comprised of two females and five males, aged between 25 and 34. Among them, three were casual users familiar with LLMs, and four were experienced users who employed LLMs daily to obtain information or make decisions. Additionally, concerning knowledge graphs, four participants had a strong understanding. In terms of health and diet management frequency:

- 4 participants (P1-P4) actively manage their health and diet;
- 1 participant (P5) occasionally manage;
- 2 participants (P6-P7) hardly manage health and dietary.

**Procedure.** To thoroughly understand users' challenges and demands in dietary recommendations, this formative study employs a two-stage procedure: a questionnaire and a co-design brainstorming session. First, participants completed a well-structured questionnaire about their experiences seeking nutrition information through search engines and LLMs, which took approximately 15 to 20 minutes. The aim was to identify the shortcomings of existing tools. Second, we presented our preliminary *HealthGenie* workflows and invited participants to co-design the system interface, encouraging them to share their opinions openly. In this phase, we sought to understand users' interaction requirements and expectations in real-world scenarios. All responses and discussions were recorded and transcribed for analysis.

### 3.2 Findings

Using the reflexive thematic analysis method [9], we analyzed the recordings of each participant's viewpoints gathered from the questionnaire and co-design brainstorming session. All participants reported having experience using search engines (e.g., Google, Bing) to search for health information. Among them, six participants also used social media and conversational LLMs, with two participants occasionally consulting professionals such as doctors and nutritionists. Only one participant mentioned referring to books or academic papers. Here, we summarize three common types of information needs that emerged from the formative study.

**Visualized and Rapid Information Delivery.** Participants expressed a strong preference for a system that not only presents information visually but also retrieves and displays results efficiently. Most participants preferred hybrid visualizations over text-only formats when seeking healthcare information. For example, P1 and P6 favored node graph diagrams due to their clarity in presenting concepts and logical relationships. P2 and P7 emphasized the usefulness of hierarchical visualizations, such as flashcards, for organizing complex content. While P5 preferred traditional text-only formats, P3 expressed concern that graphs alone might lack sufficient detail for full comprehension. Furthermore, several participants concurred on the necessity to minimize conversational turns and provide quick results. Participants P1, P2, P3, and P4 additionally desired features such as visual highlights to support rapid

information catching, with P1 noting, “Quick response is critical, as a user, I wish to obtain answers in a very short waiting time.”

**Trust and Transparency.** Given the critical importance of healthcare consulting, the reliability of information sources is often a significant concern, especially those generated by LLMs. Participants P4, P5, and P7 highlighted that one of the biggest challenges when searching for healthcare information is distinguishing reliable sources from unreliable ones. They emphasized that the prevalence of misleading or conflicting information could complicate information exploration and undermine confidence and trust in LLM-driven tools. Furthermore, P2 mentioned that “real-world data is hard to recognize,” meaning that in the era of big data, recognizing and verifying real-world evidence remains a significant gap, particularly in domains where accuracy is paramount, which is also about the difficulties in distinguishing it from other types of information in a data-saturated environment. The application of KGs to present detailed relational information was suggested as a means to enhance the sense of trust and transparency.

**Personalized guidance.** All participants anticipated an LLM-based interface designed for specific health information exploration, particularly one that offers inquiry generation and supports continuous interaction based on individual preferences. Specifically, P1, P3, P4, and P6 preferred a system capable of organizing conversations, predicting their intentions and interests, and suggesting inquiries to foster deeper discussions, thus going beyond merely answering questions. Conversely, P2, P5, and P7 emphasized the importance of maintaining the user’s initiative, suggesting that the system should allow users to type their own questions while also providing selected query recommendations. P4 further suggested that “the LLM agents should just deeply understand the user’s context to address personal matters effectively.”

### 3.3 Users’ Preference on the Interface

During the co-design brainstorming session, all participants expressed satisfaction with our design, which integrates a KG visualization alongside an LLM interaction interface. The KG effectively visualizes retrieval results, while the LLM-generated text responses outline key findings. P6 suggested adding a chat history feature to record users’ query contents, enhancing recall and usability. Meanwhile, P3 recommended providing direct and tailored instructions to help users quickly understand the functions of the interface. P5 proposed making interactions with the KG nodes more dynamic, such as allowing users to rearrange or group healthcare information of interest for better organization. These suggestions contribute to a more well-rounded design direction for *HealthGenie*.

### 3.4 Design Goals

Based on insights from the formative study and participant feedback on *HealthGenie*’s design, we establish the following **Design Goals (DG)** to guide the development of our healthcare-focused, KG-enhanced conversational LLM with dynamic interaction capabilities.

**DG1. Support Information Visualizations: Both Graphs and Hierarchical Outlines.** Non-expert users often struggle to comprehend nutritional information when it is presented in text format

due to a lack of domain knowledge. Consequently, they may find it difficult to determine the next steps and may spend a considerable amount of time grasping basic concepts. To alleviate this issue, our system should incorporate KGs to visualize the relationships between complex concepts and employ structured outlines to present information in smaller, more manageable chunks [10]. This approach enables users to navigate from broad overviews to detailed specifics as needed. Such a strategy aligns with the design principle of *progressive disclosure* [57], which is often recommended in healthcare interfaces to manage complexity without obscuring important details. By using graphs and flashcards, users can more easily form an accurate mental model of the information structure instead of being confronted with an overwhelming wall of unstructured text.

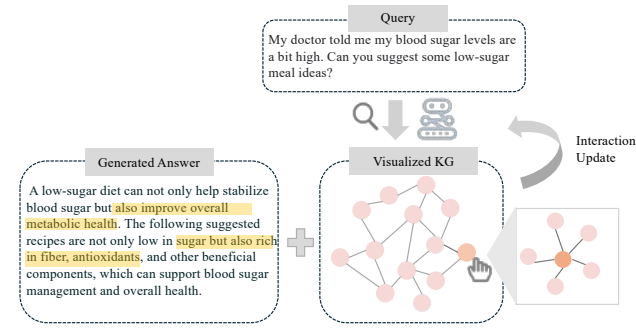
**DG2. Provide Detailed Explanations with Professional Reasons.** To enhance trust, LLM systems in healthcare must be designed with an emphasis on trust and transparency. A key principle is ensuring that the AI’s reasoning and explanations are accessible to users [35]. The system should deliver well-supported answers with logical and evidence-backed reasoning, enabling users to better understand the basis for its recommendations. Demonstrating explainability also addresses a common user concern: “Where is this advice coming from?” By structuring responses with clear justifications and contextual details, the system can alleviate doubts about accuracy and enhance overall reliability [48].

**DG3. Adapting Guidance to User Preferences.** The development goal of our system is to assist users in navigating tailored recipes and dietary plans. To achieve this, guidance on personalized preference is critical, and prioritizing user preferences ensures the retrieval of highly relevant and personalized information. By deeply understanding user preferences (e.g., dietary restrictions, flavor preferences, or health goals), our system can refine search results, predict intentions more accurately, and deliver highly customized suggestions that align with individual needs [3, 17, 42]. Furthermore, 5 out of 7 participants emphasized the importance of AI systems that dynamically adapt recommendations to sustain engaging interactions. Leveraging intention prediction modeling, our system can proactively suggest relevant content such as recipe variations, nutritional insights, or meal-plan adjustments to maintain user interest. This approach ensures conversations remain contextually relevant and exploratory, encouraging users to engage more deeply with their dietary choices and overall health.

## 4 Workflow and User Scenario

### 4.1 Workflow

Informed by the formative study and design goals, we design a *circular interaction workflow* among the LLMs, KGs, and the user. In this workflow, the user initiates a query to the LLM, which retrieves relevant information from the KG, and the corresponding subgraphs are visualized. The user can then directly manipulate the visualized KG, such as by including or excluding specific entity nodes. This, in turn, affects the behavior of the LLMs, enabling a more adaptive and controllable interaction (see Figure 2). This loop enables adaptive interactions by combining LLM reasoning with the visual, intuitive power of knowledge graphs. It allows users to explore AI



**Figure 2: Circular Interaction Workflow: Users query an LLM, which retrieves and visualizes relevant knowledge graph (sub-graphs); users can then manipulate the graph to iteratively refine outputs. Demonstrated via HealthGenie, this cyclical loop enables adaptive, non-linear exploration of dietary recommendations.**

suggestions while directly navigating and modifying structured information. The cyclical process supports dynamic, non-linear exploration, making it easier to access and refine details without a repetitive, linear approach. This *circular interaction* (i.e., query → visualization → manipulation → refined query) synergizes the intelligent reasoning capabilities of LLMs with the intuitive, visual interactivity of knowledge graphs. Users are empowered to explore AI-generated suggestions while directly examining and modifying the underlying structured information space. The cyclical nature of this workflow also supports adaptive and non-linear explorations. Instead of relying on repetitive linear queries, users can iteratively refine information based on immediate visual feedback.

Based on this workflow, we designed our system interface. As illustrated in Figure 3, *HealthGenie* combines a chat panel (A) for conversational queries and responses with a query-generation panel (B) that helps users refine or rephrase their requests. The central area (C) displays a dynamic knowledge graph, highlighting entities such as recipes, ingredients, and nutritional benefits. Users can also control the granularity of the displayed subgraph via an adjustable slider (D), choosing whether to see “show more” or “show less” detail. Finally, the interaction record panel (E) keeps track of user actions (like including or excluding certain ingredients), which can be reviewed or undone at any time.

## 4.2 Usage Scenario

Below, we showcase how *HealthGenie* assists Caroline, an office worker who has been advised to reduce her protein and salt intake.

**Initial Query and Response.** Caroline begins by asking the system: “I plan to reduce protein and salt intake, please recommend some related recipes.” As shown in Figure 3 (A), this query is processed by the LLM, which synthesizes an informative response explaining why low-protein and low-salt diets can alleviate the burden on the kidneys. Meanwhile, *HealthGenie* retrieves relevant items (recipes, ingredients, and nutritional facts) from the underlying knowledge graph (DG1).

**Knowledge Graph Visualization.** The interface then visualizes a subgraph of related ingredients and recipe nodes in the central panel (C), helping Caroline see potential meal options at a glance. She can expand or contract this network by adjusting the slider (D) to reveal more or fewer connected entities. A relevant explanation of the recommended recipe and ingredients will also be provided as part of the LLM’s output (A), offering insights into their health benefits to enhance nutritional understanding and support better decision-making (DG2).

**Refining Through Inclusion and Exclusion.** While reviewing the suggested recipes, Caroline discovers that some recommendations still contain ingredients she prefers to avoid. As shown in Figure 4, she clicks the “Black Pepper” node to exclude it from further suggestions. *HealthGenie* then prompts her to *Include* or *Exclude* this item. Once she excludes “Black Pepper,” the action immediately appears in the interaction record panel (E). Similarly, she might include “Crushed Tomato” if she particularly wants to incorporate tomato-based dishes. After confirming these choices by clicking *Apply*, the system updates the knowledge graph and regenerates refined recipe suggestions via the LLM, incorporating all of her stated preferences.

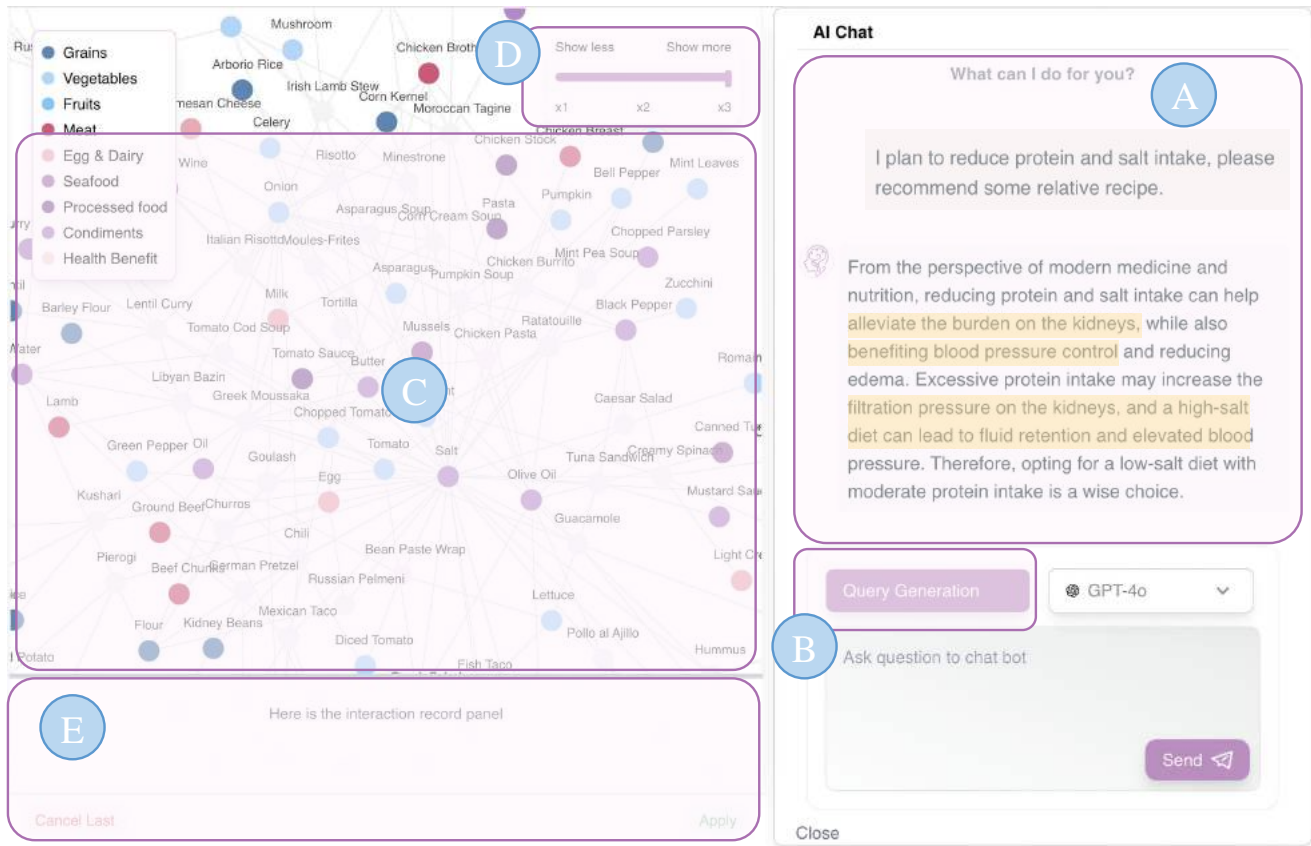
**Conversational Queries and Query Generation.** Besides explicit inclusion/exclusion actions, Caroline can also ask open-ended questions (e.g., “What are the nutritional values of these recipes?”) at any time, and *HealthGenie* will switch back to a conversational mode. It references both the current chat context and her interaction history to provide more detailed, personalized nutritional analyses. For more advanced or exploratory needs, Caroline may use the *Query Generation* button (B) to let the LLM automatically propose queries based on her current dietary goals and interaction log, sparing her from having to articulate every question herself (DG3).

**Outcome.** Through this circular interaction workflow—comprising initial queries, visualization, direct graph manipulation, and refined responses—Caroline arrives at a set of low-protein, low-salt recipes tailored to her preferences. She can dynamically expand or narrow down the recipe space, gaining a comprehensive understanding of each dish’s key ingredients and potential health benefits. This intuitive, iterative process enables Caroline to make confident, well-informed dietary decisions that align with her medical advice and personal taste.

## 5 Design and Implementation

In this section, we present the design and implementation of *HealthGenie*, a sophisticated health recipe recommendation system that features a responsive user interface with real-time chat capabilities and interactive visualization of recipe knowledge graphs. *HealthGenie* is designed as a web application, where the front end is implemented using React and Next.js while the back end is created with Python and Flask. Specifically, we equip *HealthGenie* with multiple LLM configurations to provide answers: GPT-4o mini-2024-07-18, DeepSeek-v3, Claude-3.5-Haiku, and LLaMA-3.2-90B. Additionally, *HealthGenie* supports both English and Chinese.





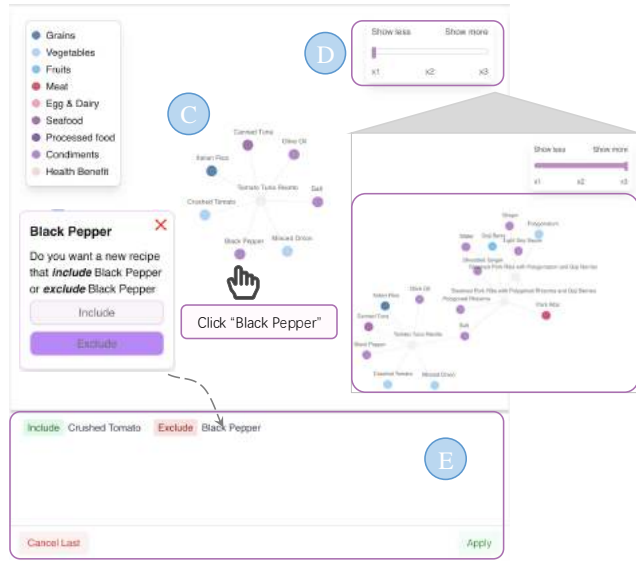
**Figure 3: The overview of HealthGenie interface, which integrates a visualized nutritional knowledge graph and a conversational dialogue system powered by an LLM. Users can initiate interactions by asking nutrition-related questions and requesting dietary guidance, then LLM retrieves relevant information and generates informative responses (A). Users can explore more relevant information using query generation (B). Simultaneously, the corresponding nutritional information is visualized within a dynamic knowledge graph (C), allowing users to explore with more and less information (D). *HealthGenie* provides interaction tracing visualization, supporting users to perceive and operate their deletion or addition intuitively (E).**

The design of *HealthGenie* consists of four key components: Knowledge Graph Construction, User Query Processing, Knowledge Matching and Adaptive Output, and Interactive Feedback and Iteration. The system begins by constructing a recipe knowledge graph, which integrates a structured recipe database with ontologies covering ingredients, health benefits, and meal categories. When a user submits a query, the system analyzes the input, along with any available interaction history, to estimate the user’s intent. Based on this intent, it retrieves a relevant subgraph and either generates a text-based response or simultaneously visualizes the subgraph while presenting personalized recipe recommendations. Users can interact with the visualized graph to refine their preferences, and these interactions are continuously fed back into the system to enhance future intent prediction and recommendation accuracy.

### 5.1 Knowledge Graph Construction

A meticulously developed knowledge graph underpins *HealthGenie*, ensuring that dietary recommendations remain transparent, explainable, and dynamically adaptable. We begin with a large-scale recipe repository containing approximately 12,500 unique recipes and 27,500 ingredient mentions. Each recipe, ingredient, nutrient, or dietary constraint is represented as a node, while edges capture semantic relations such as “contains,” “belongsToCuisine,” “recommendsFor,” or “substitutableBy.” In total, the KG now comprises over 100,000 nodes (up from the original 40,000) and at least 45 distinct relation types, enabling a comprehensive mapping of culinary and health-related concepts.

To enrich coverage, we apply a zero-shot LLM-based extraction approach that processes around 12 million tokens of free-text nutritional notes. This step identifies potential relationships beyond those explicitly defined in our ontology by leveraging the capacity of LLMs for synonym detection, context-aware inference, and



**Figure 4: Interaction with the Knowledge Graph:** The visualized Knowledge Graph in *HealthGenie* enables users to explore a broader range of recipes (D). Users can interact directly with the visualized nodes, such as hovering to view detailed relationships or clicking to include or exclude ingredients (C) in the next recommendation. Any selection to include or exclude ingredients is recorded in the interaction history panel (E), ensuring all user actions are tracked.

domain-specific clustering. For example, if the LLM detects a statement such as “lemon juice alleviates fishy odor,” it can propose a new relation, “neutralize Odor,” which we then verify and integrate into the KG if deemed valid. By incorporating such inferred edges, the KG more accurately reflects nuanced dietary interactions that users may find relevant.

We store the KG in a hybrid format. A CSV-based index encodes each triple (*subject, relation, object*), enabling large-scale ingestion, while an in-memory representation supports high-speed pathfinding and subgraph extraction during user interactions. Nodes carry both numerical attributes (e.g., *calories=290, protein=12g*) and categorical labels (e.g., *allergenClass=shellfish*). This dual-layer architecture balances scalability and efficiency, which is essential for real-time recommendations. Versioning and incremental updates further enhance reliability: if users flag new allergies or preferences, *HealthGenie* revises the relevant edges or attributes, preserving an auditable history. Through iterative refinement driven by actual usage, the KG remains aligned with evolving nutritional best practices and personalized constraints.

## 5.2 User Query Processing

To interpret users’ natural language requests and map them to relevant graph entities, *HealthGenie* employs a domain-specific processing pipeline. Its primary objective is to capture a user’s dietary goals, whether simple or complex, and translate them into symbolic constraints that can be matched against the KG. This design

ensures that we can handle ambiguous or evolving requirements while upholding critical nutritional principles.

When a user initiates a query, *HealthGenie* orchestrates multiple steps to parse and contextualize intent, unify constraints, and deliver a precise system response. By synthesizing LLM outputs with symbolic parsing, the pipeline can robustly handle potential inconsistencies in user statements and respond with thoroughly verified recommendations. Whether the user needs a brand-new recipe suggestion or wishes to override a previously defined constraint, the query processing framework ensures that each demand is accurately documented and fed into subsequent matching and feedback routines.

**5.2.1 Intention Prediction.** Before performing detailed look-ups in the knowledge graph, the system first identifies the user’s overarching intention using a specialized intent classification procedure. For instance, if a user requests “Find me a vegan lunch under 400 kcal,” *HealthGenie* determines whether this query represents a new recipe request, a modification of previously stated constraints, or a request for factual information. To achieve this, we employ a lightweight LLM-driven classification layer that distinguishes statements into *recipe search*, *constraint override*, *information request*, or *general clarification* categories.

Under the hood, *HealthGenie* utilizes a domain-specific lexical dictionary, enhanced with synonyms extracted through the LLM, to match phrases like “tasty” or “I like cheese” to relevant symbols and to identify subjective or contradictory preferences. For example, if “tasty” indicates a user’s implicit desire for flavor-related attributes, the system prompts for clarification (e.g., “Should we treat ‘tasty’ as sweet, savory, or high in umami?”). By systematically mapping such discrete user demands to symbolic constraints, *HealthGenie* encodes high-level goals into parameters like *calorieCap=400* or *isVegan=true*. This approach enables the pipeline to manage potentially ambiguous or contradictory user requirements—such as simultaneously specifying “vegan” yet mentioning a liking for cheese—by formally acknowledging each constraint before deciding on conflict resolution strategies in subsequent steps.

**5.2.2 Query Parsing.** Once the system clarifies the user’s primary goal, it proceeds to a structured query parsing stage. Here, textual requests are converted into symbolic constraints that the KG can handle. Specifically, the parser identifies key entities (e.g., “soy sauce,” “diabetes”) and relational modifiers (e.g., “low in sodium,” “avoid dairy”) and then applies boundary values such as calorie or nutrient thresholds. By encoding user instructions in a formally defined constraint set, the system maintains consistency even when inputs are partially specified or ambiguous.

**Keyword Extraction.** In this phase, we rely on a domain-specific lexical resource and a lightweight morphological analyzer to extract entities from user queries. When an exact match is missing (for instance, the KG lacks an entry for “broccolini”), *HealthGenie* consults an LLM-based synonym dictionary to propose a suitable substitute (e.g., “broccoli”). Numerical expressions like “under 300 calories” or “less than 10 grams of sugar” are recognized and linked to the relevant attribute ranges in the KG. This approach ensures that each extracted term or numeric threshold is properly mapped to a node or edge attribute in the knowledge graph.



*Language Processing.* After extracting tokens, a shallow syntactic analysis interprets how the keywords relate to one another. For instance, “I want more fiber” indicates a positive inclusion constraint, whereas “I dislike shrimp” implies an exclusion. Additional instructions, such as “cook them to retain nutrients,” are translated into method-level flags (e.g., `highRetainNutrients=true`) that guide subsequent matching. The resulting constraint set  $U$  supplies the system with a coherent, logically structured representation of user demands, ready for advanced filtering and recommendation steps in the knowledge graph pipeline.

### 5.3 Knowledge Matching and Adaptive Output

Having identified user constraints and recognized the relevant parts of the graph, *HealthGenie* proceeds to match requests against the KG, rank the results, and generate an appropriate textual or graphical response. Crucially, this output is not static. If the user later refines or reverses a constraint, the system must adapt the underlying logic in real time.

*5.3.1 Knowledge Matching.* The system first collects candidate nodes that appear likely to meet the user’s main criteria. For example, with (`vegan = true`, `calorieCap = 400`), it retrieves all dishes that contain no animal-derived nodes and have calories below 400. Edges such as “containsIngredient” or “derivesFrom” help confirm that each candidate truly aligns with the user’s constraints. Once the initial filter is complete, we refine the set by discarding nodes that violate any preference gleaned from earlier user interactions (e.g., a repeated dislike for tomatoes). In cases where multiple partial matches arise, such as dishes that meet nutritional needs but do not appear entirely vegan, *HealthGenie* explicitly labels them as “borderline” and can request user feedback. This mechanism ensures that subgraphs are computed not only by referencing static data but also by merging the user’s personal history to produce more tailored recommendations.

*5.3.2 Output Generation.* After identifying a matched subgraph, *HealthGenie* employs a multi-agent pipeline to synthesize a user-facing response. In particular, we integrate three key modules: (1) a *Graph Retrieval Agent* that uses BFS-based subgraph extraction and symbolic checks to gather nodes conforming to user constraints, (2) a *Relevance Scoring Agent* that ranks these nodes by nutritional and preference alignment, and (3) a *Language Generation Agent* that composes a concise textual summary, leveraging LLMs for coherent phrasing. Specifically, the pipeline processes the subgraph to identify important dish properties, such as “Grilled Tofu Wrap has 320 kcal, excludes dairy, and is moderately high in protein,” and incorporates them into a concise explanation.

The system produces two synchronized outputs. First, a short textual summary clarifies why certain dishes were chosen, thus revealing how constraints like `calorieCap=400` or `isVegan=true` are satisfied. Second, an interactive graph visualization illustrates the relationships among recommended recipes, ingredients, and relevant health properties. Hovering on a node reveals attribute panels (e.g., “low-sodium marinade”), and highlighting edges displays synergy or conflicts. This dual presentation, combining LLM-driven textual generation with targeted graph queries, helps users grasp

both the overarching rationale and the specific interactions of food items.

*5.3.3 Adaptive Output.* Even after the system offers an initial recommendation, users often refine their constraints (e.g., “remove soy sauce,” “add more greens”). In response, *HealthGenie* automatically revisits the relevant portion of the KG, revalidates previous selections, and applies newly discovered matches. This “adaptive output” cycle leverages adjacency-list lookups to filter nodes that conflict with updated constraints and to expand the search when new preferences suggest additional viable options. The system’s real-time recalculation ensures minimal friction. For instance, when a conflict arises, such as an incompatible allergen or an overly restrictive calorie limit, visual nodes fade or disappear, and alternative edges or nodes are suggested. This iterative, feedback-driven workflow balances the creative flexibility provided by LLM-based text generation with the symbolic precision of knowledge-graph reasoning, ensuring that each final recommendation remains logically traceable to the user’s evolving preferences.

### 5.4 Interactive Feedback and Iteration

Because dietary choices are inherently personal and may shift as users explore various options, our system devotes special attention to fostering an iterative, user-centered feedback loop. This approach enriches the decision-making process by letting people constantly re-sculpt the set of recommendations according to taste, health conditions, or newly discovered constraints.

*5.4.1 Interactive Feedback.* To accommodate continuous dialogue and graph-based exploration, *HealthGenie* adopts a bidirectional feedback model. If a user indicates hesitation through text, for example, “I’m not sure if seitan is really gluten-free,” the system can present clarifications or request confirmation, such as, “Seitan typically contains wheat gluten; do you still want to include it?” On the graphical side, hovering over or clicking on questionable items highlights exactly which properties might conflict with the user’s preferences. When the user chooses to remove or modify these items, the system logs that adjustment, updates the knowledge graph’s relevant edges or attributes, and reruns the matching routine. This interplay makes every user action interpretable and reversible. If the user regrets excluding a certain item, re-adding it is as simple as toggling a node back on the graph or removing the corresponding textual constraint.

*5.4.2 Recommendation Iteration.* Finally, after refining a preliminary recommendation, a user often wishes to broaden or narrow the search further—for example, exploring additional dishes that share some nutritional profile or removing entire categories of food if they prove unsatisfactory. This iterative “recommendation evolution” unfolds seamlessly. The system merges each newly specified preference into a symbolic profile store, ensuring subsequent queries or expansions reflect these updated conditions. Over time, *HealthGenie* learns a user’s evolving preferences, such as frequently rejecting high-carb foods or consistently opting for leafy greens, and proactively suggests more relevant alternatives. This cyclical process culminates in a thoroughly personalized exploration of the culinary space. Users can experiment with a diversity of dishes while ensuring that each recommendation remains grounded in

the knowledge graph and anchored in the constraints they have deliberately set.

## 6 User Study

We evaluated *HealthGenie*, deployed on a cloud server for participants to access, with 12 participants from various backgrounds. In this study, we focus on how effective and easy *HealthGenie* is to use in providing dietary recommendations under diverse health conditions and personal limitations. Specifically, we are interested in the following research questions.

- (1) **RQ1.** How do users perceive the visualized output and the explained texts from KG and LLM?
- (2) **RQ2.** Can the proposed circle interactive workflow effectively support user preferences on recipe exploration?
- (3) **RQ3.** How do users perceive the usefulness and the experience of interacting with *HealthGenie*?

### 6.1 Participants

We recruited 12 participants (P1–P12) via social media, including 4 females and 8 males. Their ages ranged from 23 to 32 years ( $M = 27.4$ ,  $SD = 3.02$ ). Two participants held a bachelor’s degree, four held a master’s degree, and six held a doctorate. All participants had heard of knowledge graphs, and most reported understanding the general concept of combining LLMs with KGs (see Table 1).

| ID  | Education Level | Freq. of LLM Usage | Familiarity with KG | Familiarity with LLM-KG Integration |
|-----|-----------------|--------------------|---------------------|-------------------------------------|
| P1  | PhD             | Always             | Familiar            | Familiar                            |
| P2  | PhD             | Always             | Unfamiliar          | Unfamiliar                          |
| P3  | Bachelor        | Often              | Know the concept    | Know general idea                   |
| P4  | Master          | Often              | Familiar            | Very Familiar                       |
| P5  | Master          | Always             | Familiar            | Very Familiar                       |
| P6  | Master          | Always             | Familiar            | Know general idea                   |
| P7  | Bachelor        | Often              | Know the concept    | Know general idea                   |
| P8  | PhD             | Often              | Familiar            | Very Familiar                       |
| P9  | PhD             | Always             | Familiar            | Very Familiar                       |
| P10 | PhD             | Often              | Know the concept    | Know general idea                   |
| P11 | Master          | Always             | Know the concept    | Know general idea                   |
| P12 | PhD             | Always             | Know the concept    | Know general idea                   |

**Table 1: Participant Information: We present participant ID, education level, LLM usage frequency and Familiarity with LLM-KG integration.**

### 6.2 Study Design

We carefully designed the study task to evaluate *HealthGenie*’s usability regarding personalized dietary recommendations. The study includes four tasks across two different goals in total.

**Goal A: Personalized Dietary Recommendation.** This task evaluates the system’s ability to provide tailored dietary and recipe recommendations through interactive dialogue and knowledge graph visualization. Participants are instructed to engage with the system to achieve two objectives:

- **Task 1** The participants are expected to identify a recipe that satisfies general nutritional requirements, adhere to dietary restrictions related to a health condition or dietary goal, and aligns with personal taste preferences through interaction.

**Scenario Instruction.** *Imagine you are under unhealthy conditions (e.g., indigestion, high blood sugar, poor sleep) or have some*

*dietary goals (e.g., low-protein, low-salt, low-fat). Your objective is to obtain a dietary recommendation and recipe that not only satisfies general nutritional needs but also adheres to specific dietary restrictions associated with your condition. Additionally, you have personal food preferences that should be taken into account. You are encouraged to interact with the system until you receive a recipe that fulfills all three criteria: nutritional adequacy, medical suitability, and personal preference.*

- **Task 2** The participants aim to obtain a recipe recommendation derived from a predefined set of ingredients while also ensuring the recipe aligns with their personal preferences.

**Scenario Instruction.** *You are looking for a recipe recommendation based on a given set of ingredients you already have at home. In this goal, your primary aim is to obtain a recipe that utilizes the provided ingredients while aligning with your personal taste and health conditions. You should engage in the conversation and interaction with the system to refine its suggestions until a satisfactory recipe is found.*

**Goal B: Recipe Modification and Completion.** This task examines the system’s ability to support recipe adaptation and ingredient replacement in response to user needs or incomplete information with the following goals:

- **Task 3** Participants are tasked with modifying a recipe tailored to specific groups, ensuring it meets their dietary requirements while incorporating preferred ingredients.

**Scenario Instruction.** *You follow a vegan diet and have been provided with a non-vegan recipe. Your task is to adapt the original recipe so that it fully aligns with a vegan lifestyle. Additionally, you have specific ingredient preferences or constraints (e.g., allergies or dislikes). You should work with the system to modify the recipe accordingly, ensuring it respects both their dietary restrictions and ingredient preferences.*

- **Task 4** The participants are asked to complete an incomplete recipe by identifying and integrating suitable replacements for missing ingredients.

**Scenario Instruction.** *The provided recipe is incomplete. Your goal is to identify and request appropriate ingredient substitutions or obtain a revised version of the recipe that fills in the missing components. You should explore recommendations and engage in dialogue and interaction until a complete and satisfactory recipe is constructed.*

### 6.3 Conditions

**Baseline Condition.** The baseline system consists of two components: a ChatGPT-based web application and a web interface with a knowledge graph identical to the one used in *HealthGenie*. For the conversational model, we employ GPT-4o. The knowledge graph interface offers basic retrieval functionality; for example, users can input an entity query and receive a subgraph related to the mentioned entity. In the baseline condition, participants are allowed to interact with both the ChatGPT application and the standalone knowledge graph interface to complete their tasks. They are free to decide when, whether, and how to use the provided tools.

**HealthGenie Condition.** In the *HealthGenie* condition, participants are limited to using only the *HealthGenie* system to complete

the assigned tasks. The conversational agent within *HealthGenie* is powered by the same GPT-4o model used in the baseline condition. Participants are required to engage with both the conversational agent and the interactive knowledge graph, which allows viewing and manipulating relevant subgraphs. As in the baseline condition, participants have the freedom to decide when and how to use the integrated system throughout the task.

## 6.4 Procedure

We conducted a within-subjects study to understand *HealthGenie*'s effectiveness and user experience compared to the baseline system. The study begins with an introduction to *HealthGenie* and a description of its core functionalities. Then, the participants were asked to work on Task A (personalized dietary recommendation) and Task B (recipe modification and completion). To evaluate system effectiveness in a counterbalanced manner, participants are divided into two groups. In the first group, participants complete the tasks using *HealthGenie* first, followed by the baseline condition. In the second group, the order is reversed: participants begin with the baseline condition, then proceed to *HealthGenie*. After completing each task with a given system, participants are asked to fill out a well-designed, task-specific survey to provide feedback. Each task within a given condition, along with the corresponding survey, takes approximately 3–5 minutes to complete. Upon finishing all tasks across the different systems, participants are then asked to complete a questionnaire. This post-task survey focuses on participants' overall preferences, satisfaction, and perceived effectiveness across *HealthGenie*, as well as on expressing their preferred features and challenges in using the system. Completion of this overall evaluation survey requires 35–45 minutes per participant.

## 6.5 Measurements

In the study, we measured users' experience both quantitatively and qualitatively using the following approaches.

**RQ1. Information Perception.** To investigate RQ1, we conducted a post-task evaluation study. First, users evaluated the quality of the sub-knowledge graph visualization and the generated text output across four dimensions: comprehensiveness, relevance, clarity, and organization. Next, users assessed the characteristics of the overall presented information in terms of accuracy, granularity, reliability, and interpretability. Finally, we asked users to rate the extent to which the visual information organization helped them accomplish each task.

**RQ2. User Preference Support.** To evaluate the effectiveness of *HealthGenie*'s personalized recommendation capabilities, we examined user satisfaction with personalized recommendations provided during the task. Participants were asked to rate how well the system's suggestions matched their individual needs, preferences, and situational context. In the follow-up questionnaire, participants rated the system on three key dimensions [3, 33]: intuitiveness, personalization, and accuracy. Example statements included: "I can intuitively use the system to complete tasks", "The system's recommendations accurately align with my specific condition", and "The system's suggestions closely match my tastes and preferences." This metric serves as a critical indicator of how well *HealthGenie* supports user preference choice.

**RQ3. User Experience.** To address this question, we measured the task workload using the NASA Task Load Index [18] regarding *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Performance*, and *Frustration*. Additionally, we adopted the technology acceptance model [63] to assess perceived usefulness, ease of use, ease of learning, and enjoyment of interacting with *HealthGenie*, as well as participants' intention to use it. Participants also rated the perceived helpfulness of *HealthGenie* regarding dietary recommendations and recipe revision.

## 7 Results and Analysis

In this section, we present both the quantitative and qualitative results corresponding to each research question (RQ). For each rating item, we first examine whether the order in which participants use the two interfaces influences the outcomes. To analyze RQ1 and RQ3, we employ a Wilcoxon signed-rank test [66] to compare the results between the two conditions. For RQ2, we conduct a series of mixed ANOVA tests [58], treating the order as a between-subjects variable and the interface type as a within-subjects variable.

### 7.1 RQ1. Information Perception

Figure 5 presents the characteristic results from the visualized KG and LLM outputs, highlighting the participants' responses to various evaluation dimensions. Overall, participants highly rated the output quality, especially in terms of *Organization*. Both the KG and LLM outputs received high ratings for *Organization* (KG:  $M = 4.66, SD = 0.49, W = 1.50, p = 1.0$  and LLM:  $M = 4.67, SD = 0.49, W = 2.5, p = 0.625$ ). This suggests that the presentation of information in a structured manner was highly effective across both systems.

Regarding data characteristics, *Interpretability* received the highest ratings ( $M = 4.58, SD = 0.51, W = 2.00, p = 1.00$ ), showing that participants found the outputs understandable and easy to follow. However, ratings for *Accuracy* ( $M = 4.25, SD = 0.75, W = 9.00, p = 0.1.00$ ) and *Granularity* ( $M = 4.25, SD = 0.62, W = 2.71, p = 0.042$ ) were slightly lower, with the latter indicating a statistically significant concern among participants regarding detail. This suggests that while the data were generally accurate and interpretable, there may be a preference for greater granularity in certain contexts.

Participants overwhelmingly agreed that the visual output substantially improved task efficiency. Notably, the data visualization received exceptionally high ratings for its effectiveness in facilitating task completion. Participants strongly agreed that "The visual output made revising the recipe easy" ( $M = 4.91, SD = 0.28$ ). Similarly, high ratings were given to statements that "The visual output made understanding nutrition-related information easy" and "The visual output helped select preferred recipes quickly" (both  $M = 4.88$ ), demonstrating the format's effectiveness for comprehension and decision-making. Qualitative feedback further supported these findings. P4 emphasized, "It is easy to understand through the graph," confirming the value of visual aids. P7 noted, "I think the visualization is very beneficial during my interaction since it vividly presents the main content of the response with graphs," underscoring how graphical representation enhanced engagement. Additionally, P12 appreciated the detailed analysis, stating, "The analysis of my

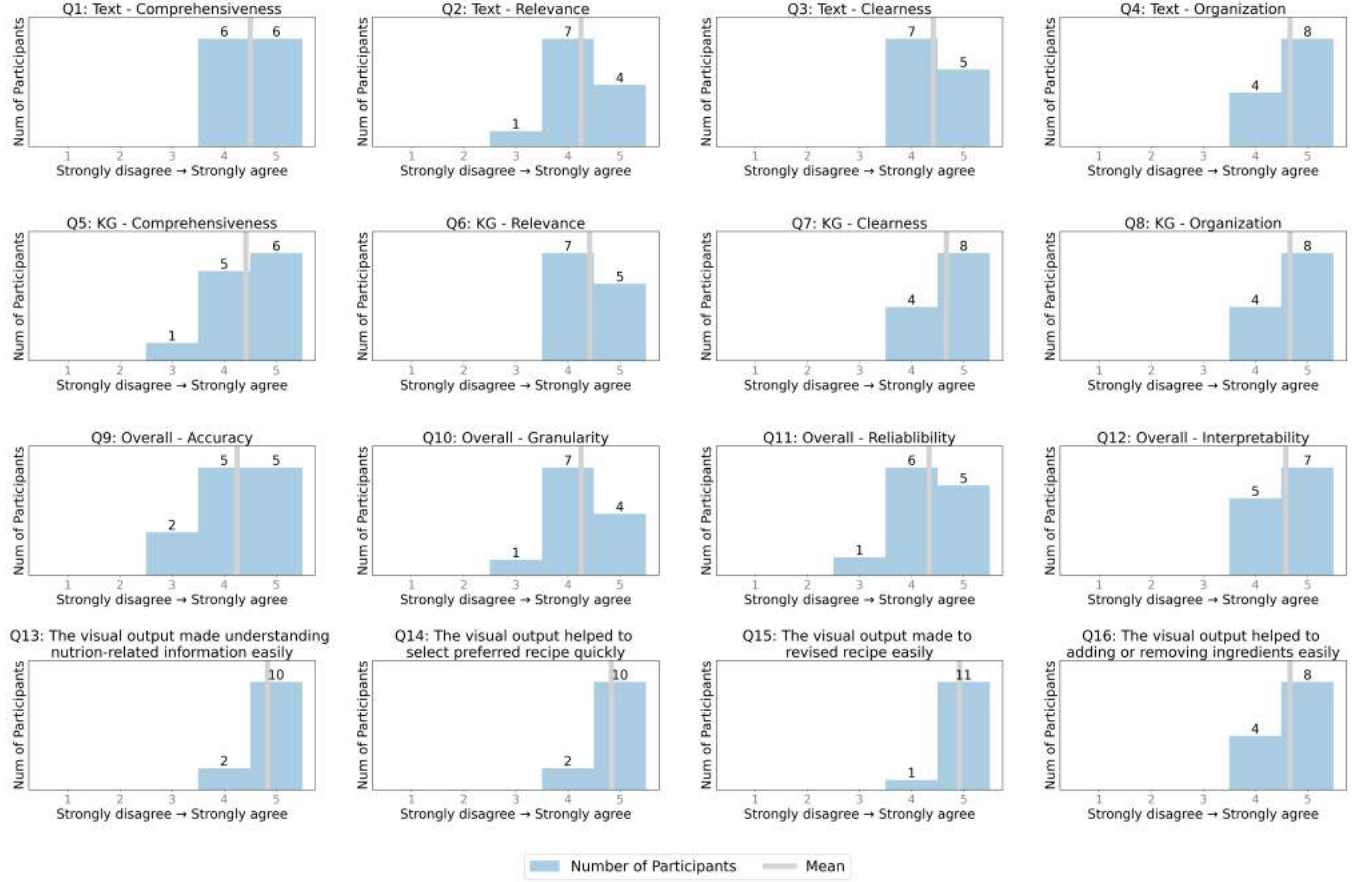


Figure 5: Participants' rating on output features of HealthGenie.

health condition and symptoms is remarkably thorough,” suggesting participants valued both clarity and depth in the visual presentation.

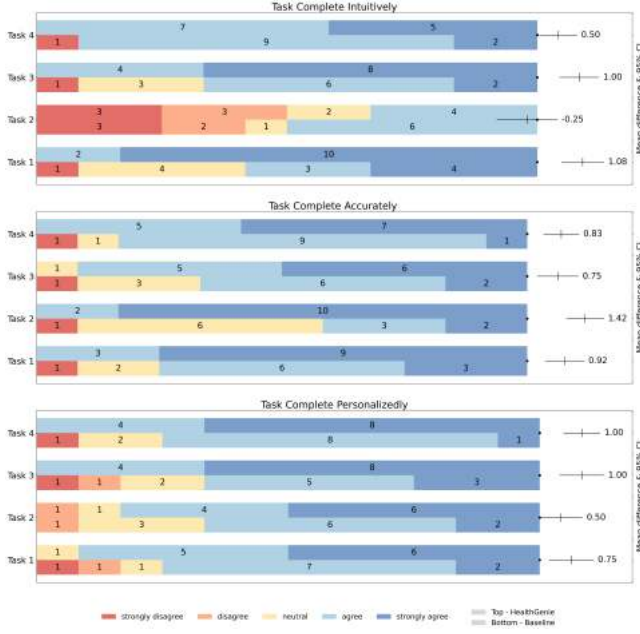
## 7.2 RQ2. User Preference Support

Figure 6 presents a comparative analysis of task completion ratings between *HealthGenie* and the baseline interface. *HealthGenie* performed better than the baseline system in most of the tasks across different perspectives (Group 1 - *HealthGenie*:  $M = 4.51, SD = 0.71$ , Baseline:  $M = 3.89, SD = 0.76$ ; Group 2 - *HealthGenie*:  $M = 4.33, SD = 0.98$ , Baseline:  $M = 3.38, SD = 1.29$ ). However, both systems demonstrated limited effectiveness in recipe recommendation tasks with constrained ingredients (*HealthGenie*:  $M = 2.58, SD = 1.24$ , Baseline:  $M = 2.833, SD = 1.33, MD_n = -0.25, F = 4.94, p = 0.038$ ), in which users experienced frustration due to repeated ingredient clarifications before obtaining viable recommendations. This limitation may stem from challenges in accurately retrieving relevant recipes from the extensive knowledge graph, particularly when processing constrained ingredient inputs.

Notably, *HealthGenie* showed superior performance in accurately completing user tasks (*Task Complete Accurately*), particularly for Task 2 (*HealthGenie*:  $M = 4.83, SD = 0.38$ ; Baseline:

$M = 3.14, SD = 1.98; MD_n = 0.75, p = 0.81$ ). This enhanced capability appears to derive from the system’s user-centered design, which consistently incorporates individual health conditions and dietary goals into its recommendation algorithm. By maintaining persistent focus on these critical factors throughout multi-turn interactions, *HealthGenie* ensures nutritional requirements are both met and remembered across conversational contexts.

In personalized recommendation tasks, *HealthGenie* consistently outperformed the baseline system, with particularly strong results in Tasks 3 and 4 (*HealthGenie*:  $M = 4.67$ , Baseline:  $M = 3.67$ ;  $MD_n = 1.00$ ). This performance advantage stems from *HealthGenie*’s cyclic workflow, which enables users to directly interact with and manipulate the recipe KG. This design not only fosters an engaging user experience but also allows the LLM to dynamically update and refine user preferences through iterative KG interactions, ensuring progressively better alignment with users’ evolving needs. Moreover, participants further reinforced these findings through positive feedback. For instance, P1 remarked: “I found it helps me to visualize the relationships of entities in a clear way; I would love to interact with the graph and explore interesting recipes.” Similarly, P2 highlighted the personalization of *HealthGenie*—“Personalizing



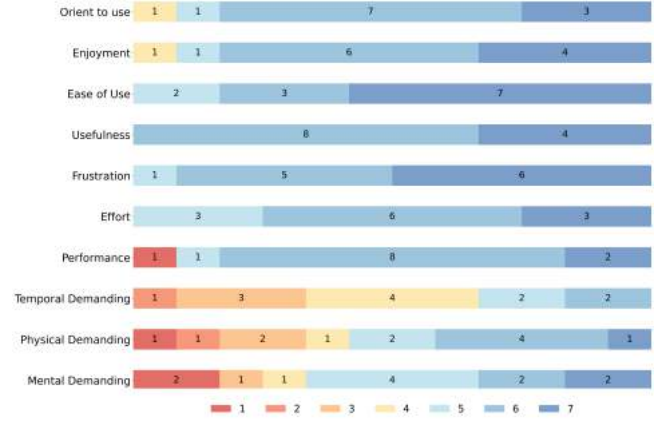
**Figure 6: Participants' responses regarding the task completion intuition, task completion accuracy and task completion satisfaction among four tasks, measured by the 5-point Likert scale questionnaire for both based and our system. Bars present the mean differences of our system compared to the Baseline. Dots indicated the 95% Confidence Interval.**

desired or unwanted vegetables using knowledge graphs was really helpful to improve recipe." P8 also added that "I think the KG graph interaction feature is really impressive because it allows me to customize the content I want in a very straight way." These comments underscore the effectiveness of *HealthGenie*'s three-party interaction workflow (LLM-KG-User), which not only personalizes recommendations but also actively encourages user exploration and system engagement.

### 7.3 RQ3. User Experience

Figure 7 presents an overview of users' perceptions regarding the workload and interaction experience with *HealthGenie*. Users generally perceived moderate but comparable levels of effort across all demand dimensions, with no statistically significant differences in mental workload ( $M = 4.58, SD = 2.02, W = 10.00, p = 1.00$ ), physical demands ( $M = 4.50, SD = 1.88, W = 5.0, p = 0.62$ ), or temporal demands ( $M = 4.08, SD = 1.24, W = 0, p = 1.00$ ). In contrast, participants reported high satisfaction with *HealthGenie*'s performance ( $M = 5.67, SD = 1.55, W = 2.00, p = 0.50$ ) and unanimously agreed that the system was easy to use for completing the assigned tasks.

Additionally, users consistently rated *HealthGenie* highly across key interaction experience dimensions. The system received good scores for Usefulness ( $M = 6.33, SD = 0.49, W = 1.50, p = 1.00$ ), indicating unanimous agreement about its value in meeting user needs. Participants also reported exceptional Ease of Use ( $M = 6.4, SD = 0.79, W = 2.00, p = 1.00$ ), confirming the interface's



**Figure 7: Participants' self-reported rating on usefulness and experience, measured by 7-point Likert scale.**

intuitive design. Enjoyment scores ( $M = 6.08, SD = 0.90, W = 3.00, p = 0.18$ ) were notably high, suggesting users found the experience genuinely engaging. Finally, strong ratings for Orientation to Use ( $M = 6.0, SD = 0.85, W = 5.00, p = 0.75$ ) demonstrated users' willingness to continue using *HealthGenie*, reflecting overall satisfaction with the system's usability and design.

### 7.4 Challenges

Despite the advanced usefulness and effectiveness of *HealthGenie*, from the open-ended experience sharing, we observed there exist three major challenges in using *HealthGenie*: (1) Insufficient data; (2) High latency; and (3) Limited KG structure.

*Insufficient data.* Some participants reported that one of *HealthGenie*'s issues is that it occasionally provides unrelated information. For example, one participant stated, "The range of recipes is somewhat limited and may lack knowledge of Eastern cuisines." This limitation is primarily due to our focus on supporting both Chinese and English systems. Initially, we prioritized Western recipes to ensure compatibility with both languages and translated the relevant entities into Chinese. However, translating Chinese recipes into English presents various challenges, such as multiple translations for the same ingredient. To maintain the quality and consistency of the recipes, we have primarily relied on Western culinary traditions.

*High latency.* Other challenges associated with the performance of *HealthGenie* are reaction speed and latency. P8 mentioned that "the latency is a little high," which we acknowledge as a valid concern. This issue arises because, in order to better detect user preferences and intentions, we have implemented multiple prompts within the system. This design allows us to gather more comprehensive and accurate feedback, helping us understand user behavior more clearly. However, to maintain consistency in the results and ensure that the system's output aligns with the user's expectations, we deliberately chose this approach. It is important to note that this introduces a trade-off between speed and performance, as the



system requires additional processing time to manage multiple prompts while still providing high-quality, relevant responses.

*Limited KG structure.* Participants mentioned that they would prefer the KG to be organized by types or to suggest more related content. One participant even suggested, “It would be even better if you could provide Amazon links to the products or recipe videos” (P5). In our pre-built KG, recipes are currently organized by similar ingredients and health benefits, which we designed to simplify the information structure and facilitate quick retrieval. This approach was intended to ensure that users could easily find recipes based on their dietary needs or specific health goals.

## 8 Discussion

In this work, we design and develop *HealthGenie*, an interactive system that provides personalized recipe recommendations and dietary guidance based on individual health conditions and goals, leveraging integrated KGs and LLMs. Our within-subjects study (N=12) compares *HealthGenie* with ChatGPT and a dummy KG retriever, thereby demonstrating its effectiveness in enhancing user support for meal planning decisions. Our user study reveals both the limitations and potential opportunities of *HealthGenie*, offering key design considerations for future interactive systems that integrate visualized KGs with intelligent LLMs.

### 8.1 Visualizing Knowledge

Visualization techniques are essential for enhancing the usability and interpretability of LLM-KG interfaces. In our work, we employ node-link diagrams [21, 27] to represent the structure and content of knowledge graphs, offering users an intuitive visual representation of entities and their relationships. However, effective KG visualization involves many additional considerations. Recent research has explored integrating KG and LLM reasoning [24, 26, 59, 61], where visualization can play a key role in presenting reasoning paths or evidence derived from the KG. This not only improves transparency but also helps users understand the system’s decision-making process.

On the other hand, *HealthGenie*’s visualization capabilities enable users to interactively explore the knowledge graph, fostering deeper insights. However, visualizing large-scale knowledge graphs remains challenging. KGs often contain millions of nodes and edges, making full-graph displays overwhelming and impractical. Future interface designs must address this complexity by incorporating techniques that simplify and focus visual representations, ensuring clarity without sacrificing information depth.

### 8.2 Enriching User Interaction: Proactive Strategies and Beyond

Effective LLM-KG interfaces should support diverse interaction paradigms that extend beyond basic question-answering. While our system currently enables interactive revision and exploration, user feedback highlights the need for richer engagement modes. A promising direction is proactive interaction, where the system anticipates user intent and delivers context-aware recommendations without explicit prompting. For example, in the health and

biomedical domains, an LLM-KG interface could (1) propose relevant literature or diagnostic pathways based on a clinician’s ongoing case analysis [71]. (2) suggest personalized learning resources by tracking a student’s progress against a structured knowledge graph of medical concepts [47]. Such proactive features transform the interface from a passive tool into an intelligent collaborator, reducing cognitive load and accelerating decision-making. However, implementing these capabilities requires careful design to balance automation with user control. Future work can explore adaptive triggering mechanisms.

Effective LLM-KG interfaces must evolve beyond basic question-answering to support more sophisticated interaction paradigms. While our current system enables interactive exploration and revision, user feedback reveals opportunities for richer engagement through proactive interaction—where the system anticipates needs and delivers context-aware recommendations. For example, in the health and biomedical domains, an LLM-KG interface could: (1) propose relevant literature or diagnostic pathways based on a clinician’s ongoing case analysis [71]; and (2) suggest personalized learning resources by tracking a student’s progress against a structured knowledge graph of medical concepts [47]. Such capabilities would transform the interface from a passive tool to an active collaborator, reducing cognitive load while accelerating decision-making.

However, implementing proactive features requires careful design to balance automation with user control. The system must provide timely, relevant assistance without becoming intrusive or overbearing. Future work should explore adaptive triggering mechanisms that respect user preferences while maintaining transparency about the system’s reasoning process. This includes developing robust intent prediction models and designing recommendation interfaces that preserve user agency—critical considerations for building trust in proactive LLM-KG systems.

### 8.3 Personalized Knowledge Graph Construction

While *HealthGenie* provides personalized recommendations through external knowledge sources, its effectiveness is inherently constrained by the static nature of pre-existing knowledge graphs, which cannot fully capture individual users’ unique needs and contexts. To overcome this limitation, we propose empowering users to construct and refine their own knowledge graphs by incorporating personal data sources such as documents, notes, and domain-specific materials [16, 77]. Future developments should focus on three key aspects: (1) intuitive interfaces for data ingestion and KG visualization, (2) LLM-assisted entity extraction and relationship identification from unstructured personal data, and (3) interactive tools for knowledge refinement and maintenance. This approach would enable users to not only build personalized knowledge structures but also actively curate and evolve them over time, creating dynamic representations that better reflect their individual knowledge domains and requirements.

## 9 Conclusion

We present *HealthGenie*, a novel interactive system that enhances dietary recommendations by dynamically integrating KGs with

LLMs. Our approach enables users to visually explore nutrient relationships, filter ingredient options by health constraints, and refine preferences through direct graph interactions, thereby eliminating the need for verbose text exchanges. Our user study demonstrates that *HealthGenie* provides more intuitive and organized recommendations compared to traditional text-based interfaces, with high user satisfaction and perceived usefulness. By bridging expert-level dietary knowledge with everyday decision-making through structured, interactive visualization, this work advances the design of human-AI interfaces for personalized health guidance.

## Acknowledgments

Acknowledgements go here. Delete enclosing begin/end markers if there are no acknowledgements.

## References

- [1] Bilal Abu-Salih, Muhammad Al-Qurishi, Mohammed Alweshah, Mohammad Al-Smadi, Reem Alfayez, and Heba Saadeh. 2023. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data* 10, 1 (2023), 81.
- [2] Oshin Agarwal, Heming Ge, Siamak Shakeri, and Rami Al-Rfou. 2021. Knowledge Graph Based Synthetic Corpus Generation for Knowledge-Enhanced Language Model Pre-training. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 3554–3565. <https://doi.org/10.18653/v1/2021.naacl-main.278>
- [3] Oscar Alvarado, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2022. A systematic review of interaction design strategies for group recommendation systems. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–51.
- [4] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. 2023. Personalized quest and dialogue generation in role-playing games: A knowledge graph and language model-based approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [5] Karin Bergling, Lin-Chun Wang, Oshini Shivakumar, Andrea Nandorine Ban, Linda W Moore, Nancy Ginsberg, Jeroen Kooman, Neill Duncan, Peter Kotanko, and Hanjie Zhang. 2025. From Bytes to Bites: Application of Large Language Models to Enhance Nutritional Recommendations in CKD. *Clinical Kidney Journal* (2025), sfaf082.
- [6] Xiwen Cai, Di Weng, Taotao Fu, Siwei Fu, Yongheng Wang, and Yingcai Wu. 2024. Linking text and visualizations via contextual knowledge graph. *IEEE Transactions on Visualization and Computer Graphics* (2024).
- [7] Kuo-En Chang, Yao-Ting Sung, and Ine-Dai Chen. 2002. The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education* 71, 1 (2002), 5–23.
- [8] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 103–119.
- [9] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298.
- [10] Rion Brattig Correia, Jordan C Rozum, Leonard Cross, Jack Felag, Michael Gallant, Ziqi Guo, Bruce W Herr, Aehong Min, Jon Sanchez-Valle, Deborah Stungis Rocha, et al. 2025. myAURA: a personalized health library for epilepsy management via knowledge graph sparsification and visualization. *Journal of the American Medical Informatics Association* (2025), ocaf012.
- [11] Steven Cox, Stanley C Ahalt, James Balhoff, Chris Bizon, Karamarie Fecho, Yaphet Kebede, Kenneth Morton, Alexander Tropsha, Patrick Wang, Hao Xu, et al. 2020. Visualization environment for federated knowledge graphs: development of an interactive biomedical query language and web application interface. *JMIR Medical Informatics* 8, 11 (2020), e17964.
- [12] Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Tianfan Fu, Chen Ling, et al. 2023. A survey on knowledge graphs for healthcare: Resources, application progress, and promise. In *ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH)*.
- [13] Mehak Preet Dhaliwal, Andong Hua, Laya Pulella, Ryan Burke, and Yao Qin. [n. d.]. NutriBench: A Dataset for Evaluating Large Language Models in Nutrition Estimation from Meal Descriptions. In *The Thirteenth International Conference on Learning Representations*.
- [14] Haoxiang Fan, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. Lesson-Planner: Assisting novice teachers to prepare pedagogy-driven lesson plans with large language models. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–20.
- [15] Peiqi Guo, Guancheng Liu, Xiaoling Xiang, and Ruopeng An. 2025. From AI to the Table: A Systematic Review of ChatGPT's Potential and Performance in Meal Planning and Dietary Recommendations. *Dietetics* 4, 1 (2025), 7.
- [16] Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. 2024. Lightrag: Simple and fast retrieval-augmented generation. (2024).
- [17] Fred X Han, Di Niu, Haolan Chen, Kunfeng Lai, Yancheng He, and Yu Xu. 2019. A deep generative approach to search extrapolation and recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1771–1779.
- [18] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [19] Steven Haussmann, Oshani Seneviratne, Yu Chen, Yarden Ne'eman, James Codella, Ching-Hua Chen, Deborah L McGuinness, and Mohammed J Zaki. 2019. FoodKG: a semantics-driven knowledge graph for food recommendation. In *The Semantic Web-ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II* 18. Springer, 146–162.
- [20] Xing He, Rui Zhang, Rubina Rizvi, Jake Vasilakes, Xi Yang, Yi Guo, Zhe He, Mattia Prosperi, Jinhai Huo, Jordan Alpert, et al. 2019. ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. *BMC medical informatics and decision making* 19 (2019), 1–18.
- [21] Marti Hearst and Melanie Tory. 2019. Would you like a chart with that? incorporating visualizations into conversational interfaces. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 1–5.
- [22] Chao-Wen Hsuan Yuan, Tzu-Wei Yu, Jia-Yu Pan, and Wen-Chieh Lin. 2024. KGScope: Interactive Visual Exploration of Knowledge Graphs With Embedding-Based Guidance. *IEEE Transactions on Visualization and Computer Graphics* 30, 12 (2024), 7702–7716. <https://doi.org/10.1109/TVCG.2024.3360690>
- [23] Qing Huang, Zhenyu Wan, Zhenchang Xing, Changjing Wang, Jieshan Chen, Xiwei Xu, and Qinghua Lu. 2023. Let's Chat to Find the APIs: Connecting Human, LLM and Knowledge Graph through AI Chain. In *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 471–483.
- [24] Yixin Ji, Kaixin Wu, Juntao Li, Wei Chen, Mingjie Zhong, Xu Jia, and Min Zhang. 2024. Retrieval and reasoning on KGs: Integrate knowledge graphs into large language models for complex question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 7598–7610.
- [25] Xiaoyun Jia, Yan Pang, and Liangni Sally Liu. 2021. Online health information seeking behavior: a systematic review. In *Healthcare*, Vol. 9. MDPI, 1740.
- [26] Jinhao Jiang, Kun Zhou, Wayne Xin Zhao, Yang Song, Chen Zhu, Hengshu Zhu, and Ji-Rong Wen. 2024. Kg-agent: An efficient autonomous agent framework for complex reasoning over knowledge graph. *arXiv preprint arXiv:2402.11163* (2024).
- [27] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring large language model responses with interactive diagrams. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–20.
- [28] Anirudha Joshi, Mandar Rane, Debjani Roy, Nagraj Emmadi, Padma Srinivasan, N Kumarasamy, Sanjay Pujari, Davidson Solomon, Rashmi Rodrigues, DG Saple, et al. 2014. Supporting treatment of people living with HIV/AIDS in resource limited settings with IVRs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1595–1604.
- [29] Efsthios Kaloudis, Victoria Kouti, Foteini-Maria Triantafyllou, Patroklos Ventouris, Rafail Pavlidis, and Vasiliki Bountziouka. 2025. AI-Powered Analysis of Weight Loss Reports from Reddit: Unlocking Social Media's Potential in Dietary Assessment. *Nutrients* 17, 5 (2025), 818.
- [30] Hsu-Ju Kao, Tsair-Wei Chien, Wen-Chung Wang, Willy Chou, and Julie Chi Chow. 2023. Assessing ChatGPT's capacity for clinical decision support in pediatrics: a comparative study with pediatricians using KIDMAP of Rasch analysis. *Medicine* 102, 25 (2023), e34068.
- [31] Nabin Khanal, Chun Meng Yu, Jui-Cheng Chiu, Anav Chaudhary, Ziyue Zhang, Kakani Katija, and Angus G Forbes. 2024. FathomGPT: A natural language interface for interactively exploring ocean science data. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.
- [32] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. 2024. Diary-Mate: Understanding User Perceptions and Experience in Human-AI Collaboration for Personal Journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [33] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504.
- [34] Philippe Laban, Jesse Vig, Marti Hearst, Caiming Xiong, and Chien-Sheng Wu. 2024. Beyond the chat: Executable and verifiable text-editing with llms. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and*

- Technology. 1–23.
- [35] Retno Larasati, Anna De Liddo, and Enrico Motta. 2021. AI healthcare system interface: explanation design for non-expert user trust. In *ACMUII-WS 2021: Joint Proceedings of the ACM UII 2021 Workshops*, Vol. 2903. CEUR Workshop Proceedings.
  - [36] Séverin Lemaignan, Mathieu Warnier, E Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2017. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* 247 (2017), 45–69.
  - [37] Harry Li, Gabriel Appleby, Camelia Daniela Brumar, Remco Chang, and Ashley Suh. 2023. Knowledge graphs in practice: Characterizing their users, challenges, and visualization opportunities. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 584–594.
  - [38] Harry Li, Gabriel Appleby, and Ashley Suh. 2024. A preliminary roadmap for llms as assistants in exploring, analyzing, and visualizing knowledge graphs. *arXiv preprint arXiv:2404.01425* (2024).
  - [39] Haotian Li, Yong Wang, Songheng Zhang, Yangqiu Song, and Huamin Qu. 2021. KG4Vis: A knowledge graph-based approach for visualization recommendation. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 195–205.
  - [40] Zhiqiang Liao, Jian Wang, Zhuozheng Shi, Lintao Lu, and Hitoshi Tabata. 2024. Revolutionary potential of ChatGPT in constructing intelligent clinical decision support systems. *Annals of Biomedical Engineering* 52, 2 (2024), 125–129.
  - [41] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. Compeer: A generative conversational agent for proactive peer support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
  - [42] Yiren Liu, Mengxia Yu, Meng Jiang, and Yun Huang. 2023. Creative Research Question Generation for Human-Computer Interaction Research.. In *UII Workshops*. 58–66.
  - [43] Marcello Maida, Ciro Celsa, Louis HS Lau, Dario Ligresti, Stefano Baraldo, Daryl Ramai, Gabriele Di Maria, Marco Cannemi, Antonio Facciorusso, and Calogero Cammà. 2024. The application of large language models in gastroenterology: a review of the literature. *Cancers* 16, 19 (2024), 3328.
  - [44] Ashlee Milton, Juan F Maestre, Abhishek Roy, Rebecca Umbach, and Stevie Chancellor. 2024. Seeking in Cycles: How Users Leverage Personal Information Ecosystems to Find Mental Health Information. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.
  - [45] Paweł Niszczoła and Iga Rybicka. 2023. The credibility of dietary advice formulated by ChatGPT: Robo-diets for people with food allergies. *Nutrition* 112 (2023), 112076.
  - [46] Allard Oelen and Sören Auer. 2024. Leveraging Large Language Models for Realizing Truly Intelligent User Interfaces. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.
  - [47] Shawn T O'Neil, Kevin Schaper, Glass Elsboukh, Justin T Reese, Sierra AT Moxon, Nomi L Harris, Monica C Munoz-Torres, Peter N Robinson, Melissa A Haendel, and Christopher J Mungall. 2024. Phenomics Assistant: An Interface for LLM-based Biomedical Knowledge Graph Exploration. *bioRxiv* (2024), 2024–01.
  - [48] Rob Procter, Peter Tolkie, and Mark Rouncefield. 2023. Holding AI to account: challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–34.
  - [49] Jianing Qiu, Kyle Lam, Guohao Li, Amish Acharya, Tien Yin Wong, Ara Darzi, Wu Yuan, and Eric J Topol. 2024. LLM-based agentic systems in medicine and healthcare. *Nature Machine Intelligence* 6, 12 (2024), 1418–1420.
  - [50] Niroop Channa Rajashekar, Yeo Eun Shin, Yuan Pu, Sunny Chung, Kisung You, Mauro Giuffrè, Colleen E Chan, Theo Saarinen, Allen Hsiao, Jasjeet Sekhon, et al. 2024. Human-algorithmic interaction using a large language model-augmented artificial intelligence clinical decision support system. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.
  - [51] Arya Rao, John Kim, Meghana Kamineni, Michael Pang, Winston Lie, Keith J Dreyer, and Marc D Succi. 2023. Evaluating GPT as an adjunct for radiologic decision making: GPT-4 versus GPT-3.5 in a breast imaging pilot. *Journal of the American College of Radiology* 20, 10 (2023), 990–997.
  - [52] Arya Rao, Michael Pang, John Kim, Meghana Kamineni, Winston Lie, Anoop K Prasad, Adam Landman, Keith Dreyer, and Marc D Succi. 2023. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. *Journal of Medical Internet Research* 25, 1 (2023).
  - [53] Bhuvan Sachdeva, Pragnya Ramjee, Geeta Fulari, Kaushik Murali, and Mohit Jain. 2024. Learnings from a Large-Scale Deployment of an LLM-Powered Expert-in-the-Loop Healthcare Chatbot. *arXiv preprint arXiv:2409.10354* (2024).
  - [54] Alberto Santos, Ana R Colaço, Annelaura B Nielsen, Lili Niu, Maximilian Strauss, Philipp E Geyer, Fabian Coscia, Nicolai J Wewer Albrechtsen, Filip Mundt, Lars Juhl Jensen, et al. 2022. A knowledge graph to interpret clinical proteomics data. *Nature biotechnology* 40, 5 (2022), 692–702.
  - [55] Priti Shah, Richard E Mayer, and Mary Hegarty. 1999. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of educational psychology* 91, 4 (1999), 690.
  - [56] Andrea Sosa-Holwerda, Oak-Hee Park, Kembra Albracht-Schulte, Surya Niraula, Leslie Thompson, and Wilna Oldewage-Theron. 2024. The role of artificial intelligence in nutrition research: a scoping review. *Nutrients* 16, 13 (2024), 2066.
  - [57] Aaron Springer and Steve Whittaker. 2019. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th international conference on intelligent user interfaces*. 107–120.
  - [58] Lars St, Svante Wold, et al. 1989. Analysis of variance (ANOVA). *Chemometrics and intelligent laboratory systems* 6, 4 (1989), 259–272.
  - [59] Jiashuo Sun, Chengjin Xu, Luminyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel M Ni, Heung-Yeung Shum, and Jian Guo. 2023. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. *arXiv preprint arXiv:2307.07697* (2023).
  - [60] Annalisa Szymanski, Brianna L Wimer, Oghenemaro Anuyah, Heather A Eicher-Miller, and Ronald A Metoyer. 2024. Integrating expertise in llms: crafting a customized nutrition assistant with refined template instructions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.
  - [61] Xingyu Tan, Xiaoyang Wang, Qing Liu, Xiwei Xu, Xin Yuan, and Wenjie Zhang. 2024. Paths-over-graph: Knowledge graph empowered large language model reasoning. *arXiv preprint arXiv:2410.14211* (2024).
  - [62] Sultan Turhan and Mustafa Berk Bacaksız. 2024. Recipe Recommendation Chatbot Based on Low FODMAP Dietary Knowledge Graph. In *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 6547–6555.
  - [63] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.
  - [64] Shijie Wang, Wenqi Fan, Yue Feng, Xinyu Ma, Shuaiqiang Wang, and Dawei Yin. 2025. Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation. *arXiv preprint arXiv:2501.02226* (2025).
  - [65] Laslo Welz and Carsten Lanquillon. 2024. Enhancing large language models through external domain knowledge. In *International Conference on Human-Computer Interaction*. Springer, 135–146.
  - [66] Robert F Woolson. 2005. Wilcoxon signed-rank test. *Encyclopedia of biostatistics* 8 (2005).
  - [67] Siyuan Xia, Nafisa Anzum, Semih Salihoglu, and Jian Zhao. 2021. KTabulator: Interactive ad hoc table creation using knowledge graphs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
  - [68] Liwenhan Xie, Chengbo Zheng, Haijun Xia, Huamin Qu, and Chen Zhu-Tian. 2024. Waitgpt: Monitoring and steering conversational llm agent in data analysis with on-the-fly code visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
  - [69] Jiawei Xu, Zhandos Sembay, Swathi Thaker, Pamela Payne-Foster, Jake Yue Chen, and Ying Ding. 2025. Interactive Visualization of Semantic Relationships in a Biomedical Project's Talent Knowledge Graph. *arXiv preprint arXiv:2501.09909* (2025).
  - [70] Yao Xu, Shizhu He, Jiabei Chen, Zihao Wang, Yangqiu Song, Hanghang Tong, Guang Liu, Kang Liu, and Jun Zhao. 2024. Generate-on-graph: Treat llm as both agent and kg in incomplete knowledge graph question answering. *arXiv preprint arXiv:2404.14741* (2024).
  - [71] Youfu Yan, Yu Hou, Yongkang Xiao, Rui Zhang, and Qianwen Wang. 2024. Knownet: Guided health information seeking from llms via knowledge graph integration. *IEEE Transactions on Visualization and Computer Graphics* (2024).
  - [72] Zhongqi Yang, Elahe Khatibi, Nitish Nagesh, Mahyar Abbasian, Iman Azimi, Ramesh Jain, and Amir M Rahmani. 2024. ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health* 32 (2024), 100465.
  - [73] Ryan Yen and Jian Zhao. 2024. Memolet: Reifying the Reuse of User-AI Conversational Memories. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
  - [74] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, et al. 2024. Multimodal healthcare AI: identifying and designing clinically relevant vision-language applications for radiology. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.
  - [75] Lanyun Zhang, Jiani Zhan, Jingyi Yang, and Verena Wai Wan Kwok. 2023. Understanding the AI-intervened User Interfaces of Online Health Consultation Platforms. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*. 1–8.
  - [76] Shengchen Zhang, Zixuan Wang, Chaoran Chen, Yi Dai, Lyumanshan Ye, and Xiaohua Sun. 2021. Patterns for representing knowledge graphs to communicate situational knowledge of service robots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
  - [77] Xinjie Zhao, Moritz Blum, Rui Yang, Boming Yang, Luis Márquez Carpiñero, Mónica Pina-Navarro, Tony Wang, Xin Li, Huitao Li, Yanran Fu, et al. 2024. AGENTiGraph: An Interactive Knowledge Graph Platform for LLM-based Chatbots Utilizing Private Data. *arXiv preprint arXiv:2410.11531* (2024).