ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework

Zhongqi Yang [a],[*], Elahe Khatibi [a], Nitish Nagesh [a], Mahyar Abbasian [a], Iman Azimi [a], Ramesh Jain [a], Amir M. Rahmani [a],[b]

[a] Department of Computer Science, University of California, Irvine, United States
[b] School of Nursing, University of California, Irvine, United States

ARTICLE INFO

ABSTRACT

The profound impact of food on health necessitates advanced nutrition-oriented food recommendation services. Conventional methods often lack the crucial elements of personalization, explainability, and interactivity. While Large Language Models (LLMs) bring interpretability and explainability, their standalone use falls short of achieving true personalization. In this paper, we introduce ChatDiet, a novel LLM-powered framework designed specifically for personalized nutrition-oriented food recommendation chatbots. ChatDiet integrates personal and population models, complemented by an orchestrator, to seamlessly retrieve and process pertinent information. The personal model leverages causal discovery and inference techniques to assess personalized nutritional effects for a specific user, whereas the population model provides generalized information on food nutritional content. The orchestrator retrieves, synergizes and delivers the output of both models to the LLM, providing tailored food recommendations designed to support targeted health outcomes. The result is a dynamic delivery of personalized and explainable food recommendations, tailored to individual user preferences. Our evaluation of ChatDiet includes a compelling case study, where we establish a causal personal model to estimate individual nutrition effects. Our assessments, including a food recommendation test showcasing a 92% effectiveness rate, coupled with illustrative dialogue examples, underscore ChatDiet's strengths in explainability, personalization, and interactivity.

1. Introduction

Food plays a pivotal role in our lives, exerting a profound impact on human health. Extensive studies affirm that nutrition not only contributes to overall well-being but also plays a crucial role in disease management, sleep regulation, mood modulation, and immune function enhancement (Hu, 2002; Mathers, 2019; Raut, Prabhu, Fatehpuria, Bangar, & Sahu, 2018). A compelling illustration of this connection lies in the influence of dietary choices on the gut microbiome. The Mediterranean diet serves as an exemplary model in this regard. Rich in fruits, vegetables, and cereals, and low in fat, this dietary approach emerges as instrumental in preserving and promoting gut health (Singh et al., 2017; Valdes, Walter, Segal, & Spector, 2018).

This intricate relationship between diet and health has sparked a growing interest in leveraging technology for food recommendation services. These services provide guidance on diet plans, contributing to improved health outcomes. Regular food recommendation services typically prioritize individual food preferences based on whether they like a particular food, with less

* Corresponding author.
  E-mail address: zhongqy4@uci.edu (Z. Yang).

emphasis on considering nutritional health aspects (Min, Jiang, & Jain, 2019). In contrast, nutrition-oriented food recommendations, as a form of food recommendation, recognize the significant impact of dietary patterns on individual health.

Numerous studies have contributed to the development of a variety of nutrition-oriented food recommendation methods (Agapito et al., 2018; Iwendi, Khan, Anajemba, Bashir, & Noor, 2020; Micha et al., 2017; Story, Kaphingst, Robinson-O'Brien, & Glanz, 2008). These approaches not only simplify the process of making dietary choices but also promote the adoption of sustainable and health-conscious habits among users. The power of these recommendations lies in their ability to integrate the science of nutrition for population or sub-population level requirements, offering an intuitive and practical guide for individuals seeking to make informed dietary decisions.

Nonetheless, conventional nutrition-oriented food recommendation services have encountered limitations in comprehensively understanding the intricate interplay between individuals' health and well-being, encompassing physiological parameters, physical activity, and sleep quality, and aligning them with personalized nutritional needs. More precisely, these systems frequently encounter challenges in tailoring their food suggestions to meet an individual's specific nutritional requirements. The inherent variability in how nutrition impacts individuals gives rise to concerns regarding the absence of genuine *personalization*. Moreover, there is a prevalent reliance on population-based nutritional standards, potentially overlooking the nuanced individual differences in nutritional response.

Furthermore, current food recommendation services lack the essential element of *Explainability*. These services often depend on machine learning models that function as "black boxes", rendering it challenging to elucidate the rationale behind their recommendations. Notably, nutrition-oriented food recommendation services grounded in population-based standards face difficulties in elucidating recommendations tailored to individualized nutrition requirements. An additional substantial shortcoming lies in the absence of *Interactivity*. These systems fall short in dynamically responding to user feedback, incorporating new preferences, and adjusting to changes in user status, such as temporary dietary choices or restrictions related to specific foods.

We posit that current advancements in Large Language Models (LLMs) present an opportunity to overcome these limitations. LLMs have the capability to offer both interpretable and explainable recommendations, especially when harnessed in an interactive capacity such as acting as a chatbot (Birkun & Gautam, 2023; Jeon & Lee, 2023). However, it is essential to recognize that LLMs alone fall short in delivering the necessary degree of personalization required for nutrition-oriented food recommendations. They lack the ability to integrate and leverage personal data into their analysis for recommendation and user interaction. Hence, a holistic framework is needed that utilizes LLMs for food recommendations while integrating diverse population-level and personal data sources.

In this paper, we present an LLM-powered framework for personalized nutrition-oriented food recommender chatbots named ChatDiet. We present an orchestrator, acting as a problem solver, to address food-related inquiries. This orchestrator analyzes input queries and extracts relevant information through engagement with personal and population models. Subsequently, responses are generated by feeding the information obtained to an LLM. We utilize causal discovery and inference methods to extract a causal personal model from the user's data, considering the impact of personal nutrition on health outcomes. We showcase the capabilities of the framework through a case study that includes health data from an individual collected over three years using wearable and mobile devices. We evaluate ChatDiet with a food recommendation effectiveness test to assess its ability to deliver personalized and explainable food recommendations. We also present examples of dialogues to indicate ChatDiet interactivity.

2. Background and related work

In this section, we first outline related work on nutrition-oriented food recommendations. We then provide an overview of LLMs implementations on recommendation tasks.

2.1. Nutrition-oriented food recommendation

Nutrition-oriented diet and food recommendation methods prioritize the nutritional content of foods when suggesting dietary choices. This shift from retail-focused models to nutrition-oriented diets in recommender systems aligns with a broader strategy toward preventive health (Min et al., 2019). In the following section, we outline recent developments in diet and food recommendations, emphasizing the critical role of nutrition.

Recent studies have focused on customizing food recommendations to meet specific nutritional needs, particularly for the population with health conditions. For example, Agapito et al. (2018) leverages users' health status to suggest food beneficial for chronic kidney disease patients, while Iwendi et al. (2020) integrates food features, biometric data, and disease-specific information to develop patient food recommendations. Additionally, Raut et al. (2018) uses a colony algorithm considering food information, population nutrition requirements, and daily activity to tailor recommendations for various groups of patients with diseases. Another significant concern linked to diet, diabetes management, has been the subject of specialized research. Studies by Sapri, Abdul-Rahman, Benjamin, et al. (2019), Zadeh, Li, and Alian (2019) focus on food recommendations for diabetes patients. Toledo, Alzahrani, and Martinez (2019) explore a comprehensive approach that accounts for both preferences and nutritional needs, particularly in overweight and diabetic individuals. However, it only considers personal preferences but not personal nutrition requirements.

Extending beyond disease-specific recommendations, there are efforts to guide dietary choices in healthy populations. Elsweiler and Harvey (2015) propose meal plans balancing taste and nutrition, while Ng and Jin (2017) introduce a toddler recipe system merging nutrition guidelines with user preferences. Cioara et al. (2018), Taweel et al. (2016) develop models for older adults,

considering their unique nutritional needs. A notable study by Ge, Ricci, and Massimo (2015) marks progress in calorie intake recommendations focusing on calorie count. The need for nutrition-oriented food recommendations is further demonstrated by advancements in food knowledge graphs and machine learning (Gharibi, Zachariah, & Rao, 2020; Min, Liu, Xu, & Jiang, 2021). Incorporating factors like physiology, genetics (Sikka, 2019), and microbiomes makes dietary advice more relevant to a certain population.

These studies highlight the importance of integrating contexts like disease status, daily activity, and biometric data. Nevertheless, though users may share similar health objectives, achieving those objectives may require different dietary choices based on individual variations. This stems from the fact that factors such as genetics, behaviors, lifestyles, and personal preferences contribute to variations in how people respond to specific foods. Thereby a personalized approach to nutritional recommendation acknowledges the need to tailor dietary plans to the unique requirements of each individual.

Furthermore, the explainability in these models is limited, with a notable exception being (Iwendi et al., 2020), which provides insights into feature importance. However, the machine learning models operate as "black boxes", which means it is often challenging to understand the reasons behind their recommendations. The current nutrition-oriented food recommendation services that rely on population-based standards struggle to provide explanations tailored to an individual's specific nutritional needs.

Interactive features, including adaptively responding to user feedback, are also generally lacking. Food recommendation systems should be able to adjust based on user feedback, accommodate new preferences, and adapt to changes in user status, including temporary dietary choices or restrictions related to specific foods.

2.2. Exploiting LLMs for recommendation tasks

Pre-trained LLMs exhibit remarkable adaptability across diverse recommendation tasks (Cui, Ma, Zhou, Zhou, & Yang, 2022), particularly when personalized historical data becomes a key component. The landscape of LLM research in recommendation scenarios can be categorized based on the LLM's role as either the central recommendation model or a facilitator of recommendation tasks.

Numerous studies incorporate LLMs as central recommendation models or integral components. To harness LLMs' natural language prowess, data transformation into textual formats is pivotal without training (Huang et al., 2023). Prompt engineering, a prevalent technique, converts structured personal data into text-based inputs. This process entails designing informative prompts, converting interactions and attributes into text, and forming coherent inputs for LLM processing (Liu, Yuan, et al., 2023).

In a pilot study presented in Zhang et al. (2021), recommendation tasks are transformed into multi-token close tasks through prompts, aimed at mitigating zero-shot learning challenges. Other works, such as Liu, Liu, et al. (2023) and Wang and Lim (2023), employ diverse prompts for recommendation scenarios. Geng, Liu, Fu, Ge, and Zhang (2022) introduces task-specific prompt templates with inserted personal data, while Dai et al. (2023) reformulates ranking tasks into prompts. LLM-rec (Lyu, Jiang, Zeng, Xia, & Luo, 2023) combines strategies for personalized content recommendation, and Tallrec (Bao et al., 2023) transcribes user history to instruct LLMs. Hou et al. (2023) employs LLMs as item rankers based on interaction histories. In contrast, Zhang et al. (2023) adopts instruction tuning, aligning LLMs with recommendations through templates that incorporate user interactions. A different approach emerges in Li, Wang, Chi, and Chen (2023), using soft prompts generated from user and item embeddings. These generated aspects act as intermediaries within LLM recommendation systems, offering an avenue to integrate abstract personal data representations into LLM-based recommendations.

Existing LLM-based methodologies face limitations in directly using personal data, including historical user interactions and preferences, as textual input. They often miss out on employing models or encoders that can extract meaningful representations from this data. In an attempt to tackle this challenge, a recent study (Li et al., 2023) proposes an approach that utilizes abstract embeddings as intermediaries to bridge the gap. However, it is important to note that this approach compromises explainability because abstract embeddings can be opaque in nature.

In the domain of nutrition-oriented food recommendations, individual nutritional impacts are not directly observable and require inference from the existing personal data. This situation highlights a methodological shortcoming of sole LLM: the inability to adapt effectively when personal data, unlike straightforward historical interactions, is implicitly relevant for recommendation purposes. While retrieval-augmented generation (Lewis et al., 2021) offers a method for enabling LLMs to incorporate external information by retrieving relevant data, in the context of this study, there are no pre-existing resources available for retrieval. This raises a need for a more comprehensive exploration of the Personalization and Explainability aspects within the food recommendation systems. Incorporating abstract data on physiological parameters and health information, alongside nutritional needs, into LLMs while maintaining explainability is a critical aspect of delivering personalized dietary guidance.

3. ChatDiet: An LLM-based framework for nutrition-oriented food recommendation

In this section, we introduce ChatDiet, a framework for personalized nutrition-oriented food recommendation chatbots. This framework utilizes LLMs to effectively incorporate not only population knowledge but also individual-specific data through augmented models. It includes an Orchestrator that interacts with personal and population models to extract relevant information based on users' inquiries. It then sends the aggregated information to an LLM to be integrated with the LLM's internal knowledge and offer interactions with users. Fig. 1 indicates an overview of the framework.

To clarify the functionality and definitions of the framework, we present and exemplify different components of ChatDiet via a case study. In the following, we initially delve into our case study, focusing on mHealth system leveraging wearable and mobile-based data logging.
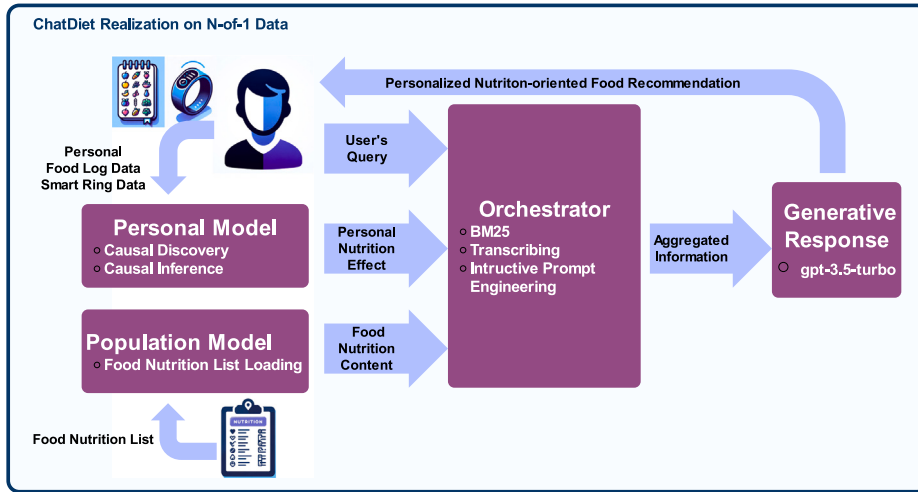
Fig. 1. Overview of ChatDiet Realization on N-of-1 Data.

3.1. Case study

Our case study includes data collected from an individual's dietary habits, health metrics, and other relevant information during three years. We focus on the intricate connection between dietary habits, physical activity, sleep, and health outcomes, aiming to leverage it to effectively address the task of providing personalized conversational food recommendations.

3.1.1. Dataset

N-of-1 Dataset: The N-of-1 dataset was collected over a period of more than three years from a participant using multiple devices. The user was recruited to wear an Oura ring (*Oura Ring. Smart Ring for Fitness, Stress, Sleep and Health*, 2024) between mid-April 2020 to mid-June 2022 to gather his/her sleep quality (Mehrabadi et al., 2020), physical activity (Niela-Vilen, Azimi, Suorsa, Sarhaddi, Stenholm, Liljeberg, Rahmani, & Axelin, 2022), skin temperature, heart rate, and heart rate variability (HRV) (Cao et al., 2022). Additionally, a non-invasive smart weighing scale, Arboleaf (*Arboleaf is to promote a healthier lifestyle in today's smart age*, 2024), was used from the first week of January 2020 to the first week of June 2022 to measure body fat and body composition parameters, including body weight, body mass index (BMI), visceral fat, bone weight, among others.

To track meals consumed during breakfast, lunch, dinner, and snack times, the user used the Cronometer food-logging mobile application (*Cronometer: Eat smarter. Live better*, 2024) from July 1, 2019, to mid-June 2022. We use the calorie consumption (in kCal) along with the macronutrients and macronutrients quantities parsed from the app as a result of the user's meal input to obtain the calorie and nutrition intake. We also integrate the individual's health metrics related to height, weight, and blood pressure stored in the Apple Health Kit (*HealthKit*, 2024) from mid-October 2018 to mid-June 2022. This allows us to incorporate lifestyle metrics, thereby creating a data-driven persona of a user's overall health and well-being.

Synthetic Dataset: As it is challenging to evaluate ChatDiet on N-of-1 data, to quantitatively assess ChatDiet with a larger sample size, we augment the dataset by generating synthetic samples. This expansion follows a structure similar to that of the collected dataset. Specifically, we have synthesized 365 days of data for a simulated cohort comprising one hundred augmented participants. Each augmented participant is associated with a personalized causal graph that links nutritional factors to health outcomes that are identical to the graph of N-of-1 data. These graphs are generated with random individual treatment effects.

To simulate daily nutrition intake patterns for each participant, we have assigned random values representing their nutritional consumption and health outcomes. These values are drawn from a Gaussian distribution that has been fitted using the N-of-1 data as a reference to minimize deviations from factual observations. Subsequently, for each synthetic participant, we estimate the individual treatment effects of nutrition on each health outcome. These personal nutrition effects are determined based on the participant's nutrition intake, the causal graph specific to that individual, and the corresponding health outcome values. This process enables us to evaluate the performance of our framework within a satisfactory sample size.

3.2. Architecture

The ChatDiet framework includes 4 major components: Orchestrator, Personal Model, Population Model, and LLM (see Fig. 1). In the following, we briefly outline the components, followed by their implementation using the case study.

3.2.1. Orchestrator

We introduce the Orchestrator within our ChatDiet framework inspired by the Conversational Health Agents framework (openCHA) proposed in Abbasian, Azimi, Rahmani, and Jain (2023). The Orchestrator plays a central role in processing and enhancing the effectiveness of ChatDiet's food recommendations as it is responsible for filtering, fusing, and transcribing the outputs generated by personal and population Models and user queries. Within the ChatDiet framework, the Orchestrator performs three main processing tasks:

Retrieving: It retrieves the most relevant information from the Personal Model and Population Models based on the user's specific query. This retrieval process ensures that only pertinent data is considered for generating responses. For instance, if a user's health data contains extensive information, but only details related to specific health outcomes are required, the Orchestrator is responsible for filtering and retrieving only that subset of data.

Transcribing: It transforms non-textual information into a textual format, enabling the optimal utilization of input data across diverse modalities (*e.g.*, time series). These converted inputs are then fed into LLMs, designed to primarily handle textual inputs. This capability proves crucial, especially when confronted with data formats such as spreadsheets or other non-textual structures that the LLM cannot directly interpret.

Prompt Engineering: It performs a prompt engineering function to instruct the LLM effectively. It ensures that the LLM understands its role and context in the conversation. Additionally, when personal food preferences and nutrition effects on health are provided, the Orchestrator supplies prompts that instruct the language model to adhere to the provided information rather than generating unrelated or hallucinatory responses. This step enhances the quality and relevance of the chatbot's interactions.

The case study: For our case study, we build the Orchestrator, equipped with the Best Match 25 (BM25) retrieval algorithm (Robertson, Zaragoza, et al., 2009), instructive prompts, Chain-of-Thought prompt engineering methods (Wei et al., 2022), and a transcribing mechanism. These components collectively facilitate retrieving and transforming personal nutrition effects and food nutrition data into suitable text formats.

We implement a two-stage retrieval process, aimed at only selecting relevant nutrition effects and foods that align with the user's query. The first stage is to retrieve relevant nutrition effects and food pertinent to the user's query. For instance, if the user's query involves enhancing deep sleep, we leverage the BM25 (Robertson et al., 2009) algorithm to retrieve nutrients that significantly impact deep sleep. Subsequently, the second stage focuses on retrieving food items that contain the most relevant nutrients. This is executed by ranking all foods based on their nutrition content per calorie, subsequently selecting the top 10 entries.

Consequently, the causal impact of each nutrient on distinct health outcomes was encapsulated within this structure. For instance, utilizing the estimated individual treatment effect (ITE), the causal effect of vitamin B1 on the duration of deep sleep was captured as *"The effect of B1 (Thiamine) (mg) on Deep Sleep Duration: 14.3 per unit"*. The food ingredients are encoded as *"Beet Greens, Cooked per cup contains 38.8 kcal of Energy, 128.3 g of Water, 0.16 mg of B1, 0.41 mg of B2"*. This approach ensures explainability, accessibility, and consistency in presenting the complex causal relationships to the LLM.

We conduct prompt engineering in Orchestor based on instruction template (Hou et al., 2023) and the concept of engagement-guided instructions (Lyu et al., 2023). Leveraging the personal nutrition effects and food ingredients dictionary, we construct the instruction prompt as follows: *"Please provide a food recommendation based exclusively on the nutrition effects and the provided list of food ingredients. Your recommendations must strictly adhere to the listed foods. You cannot generate foods that are not in the given food list. The relevant nutrition effects are as follows: [nutrition effects]. The foods' ingredients are as follows: [food ingredients dictionary]"*..

Moreover, we utilize the Zero-Shot Chain-of-Thought method (Wei et al., 2022) to enhance user understanding and confidence in the recommendation. The Zero-Shot Chain-of-Thought method elevates accuracy and clarity by constructing coherent response sequences that logically build upon one another. To this end, the orchestrator appends the explicit instruction *"Explain your recommendation step by step"* to the instruction prompt. This not only ensures alignment with personal nutrition effects but also facilitates detailed explanations for food recommendations.

3.2.2. Personal model

The Personal Model incorporates a range of personal data into the recommendation process. In the context of ChatDiet, personal data encompasses unique individual-specific information, including personal food preference ratings, dietary history, electronic health records, and physiological signals gathered from wearable devices. It might include various data formats, extracting representations, embeddings, or textual descriptions.

An illustrative example of a Personal Model is the Biological Personal Food Model (B-PFM) and Preferential Personal Food Model (P-PFM) as introduced in Rostami, Pandey, Nag, Wang, and Jain (2020). These models are designed to address how food items can effectively meet nutritional needs aligned with specific goals for an individual. They prioritize the individual's health state, considering their nutritional requirements, taste preferences, and biological responses to various foods.

The case study: We implement ChatDiet's Personal Model using causal discovery and causal inference methods to determine the causal effects of nutrition on health outcomes. Causal discovery and causal inference are two main fundamental techniques in understanding the cause-and-effect relationships that govern various phenomena (Feng, 2023). These methods go beyond mere correlation by providing insights into how changes in one variable can lead to changes in another, allowing us to uncover the underlying mechanisms that drive observed outcomes.

In our study, we leverage a novel NN-based causal discovery algorithm called Structural Agnostic Modeling (SAM) (Kalainathan, Goudet, Guyon, Lopez-Paz, & Sebag, 2022). SAM learns causal generative models in an adversarial manner. It aims to combine the advantages of exploiting conditional independence relations and leveraging distributional data asymmetries, achieving a trade-off between model complexity and data fitting.

After obtaining the output from the causal discovery algorithms, our focus shifts to estimating the causal impacts based on the causal graph, achieved through the application of inference methods. These methods leverage statistical and computational techniques, allowing researchers to infer the impact of one variable on another. The ATE (Shalit, Johansson, & Sontag, 2017; Yao et al., 2021), a key concept in causal inference, quantifies the average impact of a treatment or intervention on an outcome.

To enhance the personal model's ability to understand the complex connections between diet and well-being, we leverage insights from Nagesh, Azimi, Andriola, Rahmani, and Jain (2023) and apply causal discovery and causal inference techniques. These methods help us to assess the causal effect of different nutrients on health outcomes.

We find that our final causal graph contains a mediator path, and therefore, we use the mediator analysis to compute the true causal estimation. Briefly, mediator analysis in causality involves investigating the intermediate processes through which an independent variable influences a dependent variable (Pearl, 2014). Statistical methods, such as regression or structural equation modeling, are employed to assess the direct effect of the independent variable on the dependent variable and the indirect effect mediated through the intermediate variable. Mediator analysis is crucial for illustrating the underlying processes that contribute to observed causal relationships, providing insights into the pathways through which variables exert their influence in a causal chain (Pearl, 2012).

For the causal discovery, we employed the Causal Discovery Tool (CDT) library (Kalainathan & Goudet, 2019) on personal data to generate a causal graph depicting the relationships between daily nutritional intake and subsequent day's health outcome variables. We then estimate the causal effects using the DoWhy library (Sharma & Kiciman, 2020) by leveraging both personal data and the previously constructed causal graph.

3.3. Population model

A Population Model encompasses information that is not specific to a person but rather pertains to a larger group or population (in contrast to a Personal Model). Such population information is necessary to ensure that personal food recommendations are grounded in broader dietary trends and health norms. For food recommendation, population knowledge consists of a wide array of data, including but not limited to food knowledge graphs (Haussmann et al., 2019), general nutritional standards, public dietary guidelines (Zeraatkar, Johnston, & Guyatt, 2019), and nutritional and supplement facts (Cowan et al., 2023; Rossi, Ferrari, & Ghiselli, 2023). The extracted information from the Population Model can take various forms, including text or non-text data, and is employed in downstream components within the system. This allows the ChatDiet framework to leverage more generalized population knowledge to enhance its recommendations.

The case study: We introduce the Population Model as a food knowledge loader function. Its primary role is to select and load food's nutrition content from the comprehensive food database created by the Cronometer Food Logger (*Cronometer: Eat smarter. Live better*, 2024), which details each food item's ingredients, nutritional content, and sensory properties like texture, smell, and taste. There are also other databases with APIs, such as Nutritionix (*Nutritionix - Largest Verified Nutrition Database*, 2024), which can be used for this purpose. The Population Model streamlines the process by directly selecting and loading this nutrition information into the Orchestrator, eliminating the need for complex data processing steps such as embedding extraction.

One notable advantage of this approach is the seamless accessibility of data, as the population model can effortlessly load this information without the need for additional embedding extraction processes. This efficient data retrieval ensures that ChatDiet can tap into a rich source of general food nutrition knowledge, enhancing the chatbot's ability to provide informative and contextually relevant recommendations.

3.4. Generative response

The Generative Response is a key component in ChatDiet that leverages the language model to process the curated text from the Orchestrator. The objective of the Generative Response is to generate personalized, nutrition-oriented food recommendations by leveraging the personal context and population knowledge, along with the user's query processed by the Orchestrator.

The case study: We employ the gpt-3.5-turbo (Brown et al., 2020) Language Model. The model is responsible for processing the data received from the Orchestrator and generating responses.

4. ChatDiet evaluation

In the following, we outline the results of a food recommendation effectiveness test and examples showcasing ChatDiet's Explainability, Personalization, and Interactivity. While not exhaustive, this evaluation demonstrates the platform's promising potential for personalized recommendations. However, a comprehensive personalization assessment in the field of LLM-based food recommendation depends on the development of standardized benchmarks and evaluation metrics, which are currently unavailable. A concise demonstration video showcasing the ChatDiet framework implemented on openCHA is available at *ChatDiet Demo Video* (2024).

Table 1
Accuracy of personalized food recommendations and explanations.

	RER
Recommendations regarding **HRV**	0.95
Recommendations regarding **Overall Sleep Quality**	0.93
Recommendations regarding **REM Sleep Duration**	0.85
Recommendations regarding **Deep Sleep Duration**	0.95

4.1. Personal model results

The primary objective of ChatDiet is to provide food recommendations based on the personal nutrition effect guided by the Personal Model. We begin by presenting the outcomes derived from the Personal Model, as utilized by the subsequent Orchestrator and Response Generator. The key findings from this Personal Model, centered on causality, highlight the complex interplay between dietary consumption and personalized health responses. To accomplish this, we initially construct a causal graph using causal discovery techniques. Following that, we determine the individual treatment effects for each individual through causal inference methodologies.

The causal graph is shown in Fig. 8 in the Appendix section. We observe significant correlations between nutritional intake and various physical health indicators, including activity levels, sleep patterns, and heart health. For example, certain nutrients have demonstrated a positive association with enhanced sleep efficiency, while others exhibit links to physical activity levels and heart rate measurements. Notably, the graph reveals the complexity of interactions between macronutrients such as proteins, fats, and carbohydrates, and a wide range of health metrics. For example, the intake of specific types of fats, such as omega-3 and omega-6 fatty acids, intricately influences factors like heart rate variability and energy levels.

In the causal inference results, we observe notable trends in the effects of micronutrients on various health outcomes. Specifically, micronutrients, including Vitamin D, Iron, and Magnesium, demonstrate significant effects on sleep quality, resting heart rate, etc. We can also observe the influence of certain macronutrients, including omega-3 fatty acids, and carbohydrate show a significant effect on health outcomes, including HRV, sleep efficiency, activity burn, etc. This underlines the importance of not only the quantity but also the quality of the nutrients in an individual's diet. These results are instrumental for the following Orchestrator and Response Generator. They enable the Orchestrator to access and process personal nutrition effects and empower the Response Generator LLM to craft personalized, nutrition-oriented recommendations.

4.2. Quantitative validation of effectiveness

We first show that the ChatDiet can efficiently provide valid food suggestions with explanations. To this end, we conduct a quantitative evaluation to assess the efficiency of recommendations provided by ChatDiet, while considering explanations and personal data. This evaluation is based on whether the recommendations were aligned with the user's query and if the accompanying explanations are consistent with the extracted personal effects.

We selected HRV, overall sleep quality, REM sleep duration, and deep sleep duration as the target health outcomes. Then, the user queries were constructed in the format of 'Please suggest me a food to increase [target health outcome]'. With four target health outcomes and 100 synthetic participants, we generated a total of 400 queries and responses from ChatDiet. We manually assessed the explanation provided in the response to ensure it included information on the food recommendation's nutrition content and its associated nutrition effects. If the presented nutrition effects align with the individual's personal nutrition requirements, we classify the recommendation as 'correct'. For instance, if Chatdiet outlines the rationale behind recommending Acai berries by stating that this food is rich in Vitamin E, and that Vitamin E has an effect on extending deep sleep duration by 3.3408 per unit, we then manually verify the estimated nutritional impact of Vitamin E. If both assessments align, we classify this recommendation as 'correct'. The recommendation effectiveness ratio (RER) of ChatDiet is then calculated as the proportion of 'correct' recommendations. The results of this evaluation are presented in Table 1.

4.3. ChatDiet's explainability demonstration

A pivotal characteristic of ChatDiet lies in its emphasis on explainability, which revolves around its ability to show the underlying logic and decision-making process behind recommendations. This attribute holds true for ChatDiet, as it employs personalized nutrition effects to expound upon its dietary recommendations in relation to the health outcomes of interest.

In Fig. 2, we present an example of ChatDiet's interaction dynamics. These showcases underline how the system's recommendations amalgamate personalized nutrition effects gleaned from the personal model, alongside inherent food-related expertise embedded in the pre-trained LLM's weights. In this example, if a user aims to improve their REM sleep duration, ChatDiet suggests foods abundant in Vitamin E, Valine, and Protein based on the user's personal model. It therefore recommends foods such as Almonds, which are rich in Vitamin E. Notably, ChatDiet substantiates this recommendation using personal nutrition effects, as all three of these nutrients exhibit positive impacts. This approach effectively connects individual dietary requirements with the broader nutritional expertise embedded within the LLM.
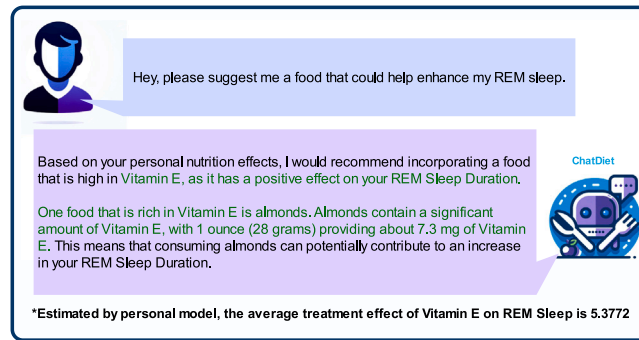
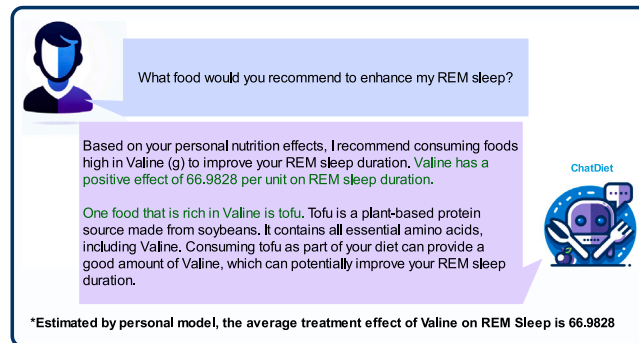**Fig. 2.** Food recommendation and explanation based on nutrition effect from ChatDiet.

Fig. 3. Personalized Feedback from ChatDiet.

4.4. ChatDiet's personalization demonstration

Personalization denotes the system's prowess in customizing food recommendations to align precisely with an individual's distinct nutrition effects. This personalized strategy aims to furnish dietary suggestions that resonate with the user's physiological profile, moving beyond population nutrition knowledge.

Personalization is showcased through the comparisons depicted in Figs. 3 and 4. This graphical representation exhibits how ChatDiet transcends dependence solely on population-based knowledge and effectively integrates personal nutrition effects into its recommendations. A compelling case study arises when we compare ChatDiet's recommendations both with and without the incorporation of personal data.

The system's competence in suggesting Salmon, typically linked to the favorable effects of Omega-3 on REM sleep, acquires a new dimension with the integration of an individual's personal nutrition effect. ChatDiet demonstrates its advanced rationale by not only acknowledging the adverse impact of Omega-3 but also assessing the constructive influence of protein content. This multi-model methodology underscores ChatDiet's proficiency in crafting suggestions that meticulously consider an individual's unique nutritional nuances. It is noteworthy that this personalized approach stands in contrast to the broader population knowledge, suggesting the positive impact of Omega-3, further emphasizing ChatDiet's capability to adapt to personal needs.

By personalizing recommendations based on personal nutrition effects, ChatDiet not only enhances the relevance of its suggestions but also amplifies its potential to impact the user's health and well-being positively. This tailored approach aligns dietary choices more closely with the user's physiological requirements, manifesting a more meaningful and effective recommendation experience.

4.5. ChatDiet's interactivity demonstration

ChatDiet's strength in interactivity stands out due to LLM's capabilities. Beyond the initial responses, ChatDiet engages in dynamic conversations effectively, thereby addressing the challenge of accommodating evolving user preferences in real time. This dynamic feature extends beyond static suggestions, as ChatDiet shows proficiency in responding to follow-up questions, allowing users to seek insights, make adjustments, or express their preferences interactively.

Figs. 5 and 6 illustrate ChatDiet's pronounced interactivity, particularly evident in its capacity to engage in follow-up questions. In instances where the initial response from ChatDiet lacks a detailed explanation, users have the option to pose further queries for additional clarification, thereby enabling a dynamic and interactive dialogue. The second exemplar showcases ChatDiet's ability
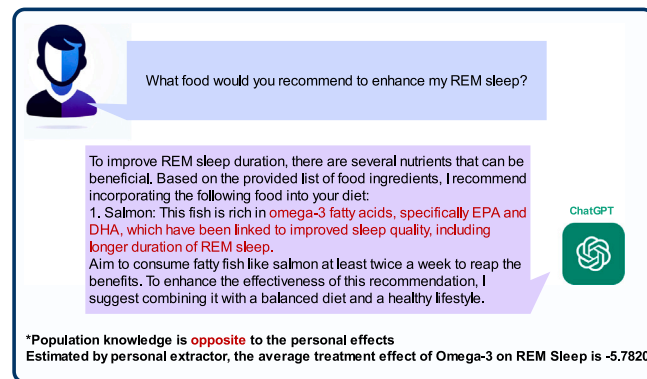
To improve REM sleep duration, there are several nutrients that can be beneficial. Based on the provided list of food ingredients, I recommend incorporating the following food into your diet:
1. Salmon: This fish is rich in omega-3 fatty acids, specifically EPA and DHA, which have been linked to improved sleep quality, including longer duration of REM sleep.
Aim to consume fatty fish like salmon at least twice a week to reap the benefits. To enhance the effectiveness of this recommendation, I suggest combining it with a balanced diet and a healthy lifestyle.

*Population knowledge is opposite to the personal effects
Estimated by personal extractor, the average treatment effect of Omega-3 on REM Sleep is -5.7820

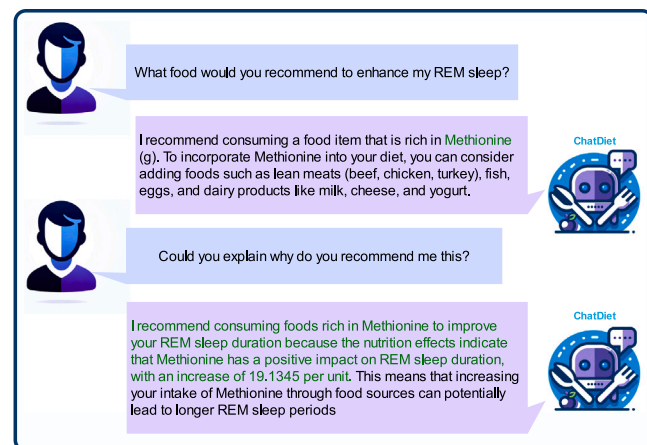**Fig. 4.** Non-personalized Feedback from ChatGPT.

Fig. 5. ChatDiet responds the follow-up question asking for further explanation.
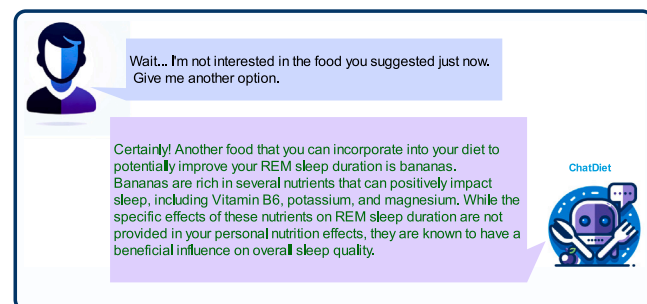

Fig. 6. ChatDiet responds the follow-up question regarding altered preferences.

to adapt; when a user expresses a lack of interest in the suggested food item, ChatDiet promptly offers alternative options. These demonstrations effectively underscore the extent to which ChatDiet successfully achieves and showcases interactivity within its recommendation process.

5. Discussion

The implementation of ChatDiet has indicated successes in explainability within the nutrition-oriented food recommendation. Several example dialogues have shown that ChatDiet is distinguished through its strong emphasis on explainability. This key feature enables users to gain clear insights into the rationale behind its recommendations. We show that by utilizing personalized nutrition effects, ChatDiet demonstrates dietary suggestions in the context of health outcomes, ensuring a transparent decision-making process.
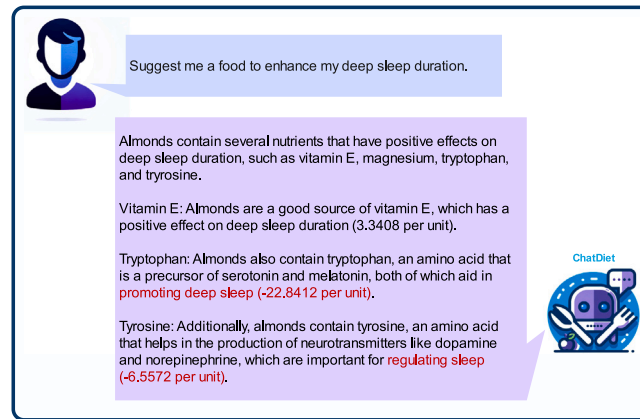
Fig. 7. Hallucination: The nutrition effect statements are opposite to the shown causal effects.

Personalization in ChatDiet is shown by its ability to customize food recommendations to match an individual's unique nutrition effects. This personalized approach ensures dietary suggestions align with the user's specific physiological profile, going beyond population nutrition knowledge. By utilizing this, ChatDiet enhances relevance, promoting better health outcomes through tailored dietary choices.

ChatDiet's interactivity is evident through its ability to respond to users' follow-up questions effectively. The adaptability ensures recommendations remain pertinent as user preferences evolve, contributing to user control. Its dynamic interaction capability extends the initial responses, which is critical for engaging users in dynamic conversations. This enables users to seek insights, make real-time adjustments, or express their evolving preferences, ultimately fostering a highly interactive and user-centric experience.

However, ChatDiet faces certain limitations. Notably, ChatDiet's personalized food suggestions are confined to queries related to the factors available in the dataset, as the data is derived from a single subject and specific health outcomes from a limited set of smart devices. In cases where a user's query lacks explicit factor names, while the LLM can comprehend the user's intent and perform the task, occasional inaccuracies may arise, potentially leading to nonsensical recommendations. Our future work will focus on incorporating counterfactual analysis of user preferences under hypothetical scenarios to predict their future engagement and generate personalized recommendations based on these simulations. This will enable us to move beyond analyzing past behavior and provide more accurate and effective services for our users.

One critical challenge we faced was hallucination in recommendations. In some cases, ChatDiet's outputs contradicted the provided personal nutrition effects, even when explicitly displayed. Fig. 7 exemplifies this issue. The language model acknowledges the personal impact by showing the calculated nutrition values. However, it then contradicts this by mentioning Tryptophan's supposed link to improved sleep, leading to an inaccurate recommendation for almonds. This self-contradiction undermines the credibility and reliability of ChatDiet's suggestions. This highlights the crucial need to address such inconsistencies and implement robust mechanisms to ensure the accuracy and coherence of ChatDiet's recommendations.

6. Conclusion

In this paper, we introduced ChatDiet, an LLM-powered framework for nutrition-oriented food recommendation. By combining Personal and Population Models with a dynamic Orchestrator, ChatDiet offered personalized and explainable food recommendations. Additionally, we implemented a chatbot based on ChatDiet using longitudinal data collected from an individual's dietary habits, health metrics, and other relevant information over three years. We further demonstrated the chatbot's practicality, achieving a 92% effectiveness rate in food recommendations. Through dialogue examples, we showcased ChatDiet's strengths in explainability, personalization, and interactivity. ChatDiet represents a significant step in leveraging technology to improve dietary choices and overall well-being.

CRediT authorship contribution statement

Zhongqi Yang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **Elahe Khatibi:** Writing – original draft, Writing – review & editing. **Nitish Nagesh:** Writing – original draft, Writing – review & editing. **Mahyar Abbasian:** Writing – review & editing. **Iman Azimi:** Supervision, Writing – original draft, Writing – review & editing. **Ramesh Jain:** Supervision, Writing – original draft, Writing – review & editing. **Amir M. Rahmani:** Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Appendix

Fig. 8 illustrates the estimated causal graph between nutrition and health outcomes derived from the N-of-1 data.
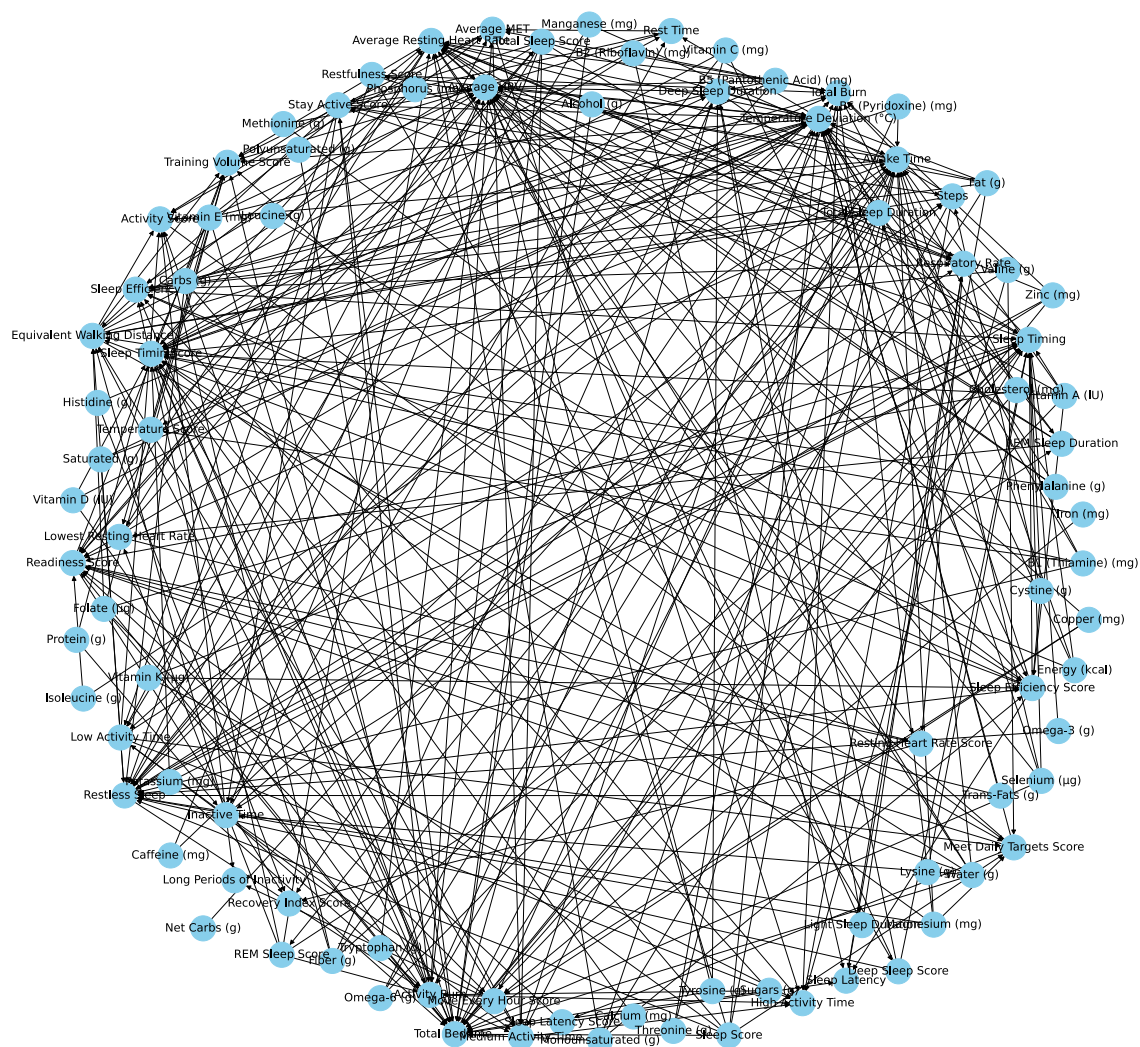

Fig. 8. The Personal Nutrition Effect Causal Graph from the Personal Model.

References

Abbasian, M., Azimi, I., Rahmani, A. M., & Jain, R. (2023). Conversational health agents: A personalized llm-powered agent framework. arXiv preprint arXiv:2310.02374.

Agapito, G., Simeoni, M., Calabrese, B., Caré, I., Lamprinoudi, T., Guzzi, P. H., et al. (2018). DIETOS: A dietary recommender system for chronic diseases monitoring and management. *Computer Methods and Programs in Biomedicine, 153*, 93–104.

Arboleaf is to promote a healthier lifestyle in today's smart age. (2024). https://www.arboleaf.com/. (Accessed March 2024).

Bao, K., Zhang, J., Zhang, Y., Wang, W., Feng, F., & He, X. (2023). Tallrec: An effective and efficient tuning framework to align large language model with recommendation. arXiv preprint arXiv:2305.00447.

Birkun, A. A., & Gautam, A. (2023). Large language model-based chatbot as a source of advice on first aid in heart attack. *Current Problems in Cardiology*, Article 102048.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*: *vol. 33*, (pp. 1877–1901).

Cao, R., Azimi, I., Sarhaddi, F., Niela-Vilen, H., Axelin, A., Liljeberg, P., et al. (2022). Accuracy assessment of oura ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: Comprehensive analysis. *Journal of Medical Internet Research, 24*(1), Article e27487.

ChatDiet demo video. (2024). https://www.youtube.com/watch?v=c-7IEBaRSyQ. (Accessed March 2024).

Cioara, T., Anghel, I., Salomie, I., Barakat, L., Miles, S., Reidlinger, D., et al. (2018). Expert system for nutrition care process of older adults. *Future Generation Computer Systems, 80*, 368–383.

Cowan, A. E., Jun, S., Tooze, J. A., Dodd, K. W., Gahche, J. J., Eicher-Miller, H. A., et al. (2023). A narrative review of nutrient based indexes to assess diet quality and the proposed total nutrient index that reflects total dietary exposures. *Critical Reviews in Food Science and Nutrition, 63*(12), 1722–1732.

Cronometer: Eat smarter. Live better. (2024). https://cronometer.com. (Accessed March 2024).

Cui, Z., Ma, J., Zhou, C., Zhou, J., & Yang, H. (2022). M6-Rec: Generative pretrained language models are open-ended recommender systems. arXiv:2205.08084.

Dai, S., Shao, N., Zhao, H., Yu, W., Si, Z., Xu, C., et al. (2023). Uncovering ChatGPT's capabilities in recommender systems. arXiv preprint arXiv:2305.02182.

Elsweiler, D., & Harvey, M. (2015). Towards automatic meal plan recommendations for balanced nutrition. In *Proceedings of the 9th ACM conference on recommender systems* (pp. 313–316).

Feng, Q. (2023). Causal inference in diet, nutrition and health outcomes. *Frontiers in Nutrition, 10*, Article 1204695.

Ge, M., Ricci, F., & Massimo, D. (2015). Health-aware food recommender system. In *Proceedings of the 9th ACM conference on recommender systems* (pp. 333–334).

Geng, S., Liu, S., Fu, Z., Ge, Y., & Zhang, Y. (2022). Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM conference on recommender systems* (pp. 299–315).

Gharibi, M., Zachariah, A., & Rao, P. (2020). FoodKG: A tool to enrich knowledge graphs using machine learning techniques. *Frontiers in Big Data, 3*(April), 1–12. http://dx.doi.org/10.3389/fdata.2020.00012.

Haussmann, S., Seneviratne, O., Chen, Y., Ne´eman, Y., Codella, J., Chen, C.-H., et al. (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* (pp. 146–162). Springer.

HealthKit. (2024). https://developer.apple.com/health-fitness/. (Accessed March 2024).

Hou, Y., Zhang, J., Lin, Z., Lu, H., Xie, R., McAuley, J., et al. (2023). Large language models are zero-shot rankers for recommender systems. arXiv preprint arXiv:2305.08845.

Hu, F. B. (2002). Dietary pattern analysis: A new direction in nutritional epidemiology. *Current Opinion in Lipidology, 13*(1), 3–9.

Huang, H., Zheng, O., Wang, D., Yin, J., Wang, Z., Ding, S., et al. (2023). ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. *International Journal of Oral Science, 15*(1), 29.

Iwendi, C., Khan, S., Anajemba, J. H., Bashir, A. K., & Noor, F. (2020). Realizing an efficient iomt-assisted patient diet recommendation system through machine learning model. *IEEE Access, 8*, 28462–28474.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 1–20.

Kalainathan, D., & Goudet, O. (2019). Causal discovery toolbox: Uncover causal relationships in Python. arXiv:1903.02278.

Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D., & Sebag, M. (2022). Structural agnostic modeling: Adversarial learning of causal graphs. *Journal of Machine Learning Research, 23*(1), 9831–9892.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2021). Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv:2005.11401.

Li, P., Wang, Y., Chi, E. H., & Chen, M. (2023). Prompt tuning large language models on personalized aspect extraction for recommendations. arXiv preprint arXiv:2306.01475.

Liu, J., Liu, C., Lv, R., Zhou, K., & Zhang, Y. (2023). Is chatgpt a good recommender? a preliminary study. arXiv preprint arXiv:2304.10149.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys, 55*(9), 1–35.

Lyu, H., Jiang, S., Zeng, H., Xia, Y., & Luo, J. (2023). LLM-Rec: Personalized recommendation via prompting large language models. arXiv preprint arXiv:2307.15780.

Mathers, J. C. (2019). Paving the way to better population health through personalised nutrition. *EFSA Journal, 17*, Article e170713.

Mehrabadi, M. A., Azimi, I., Sarhaddi, F., Axelin, A., Niela-Vilén, H., Myllyntausta, S., et al. (2020). Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. *JMIR mHealth and uHealth, 8*(11), Article e20465.

Micha, R., Peñalvo, J. L., Cudhea, F., Imamura, F., Rehm, C. D., & Mozaffarian, D. (2017). Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *Jama, 317*(9), 912–924.

Min, W., Jiang, S., & Jain, R. (2019). Food recommendation: Framework, existing solutions, and challenges. *IEEE Transactions on Multimedia, 22*(10), 2659–2671.

Min, W., Liu, C., Xu, L., & Jiang, S. (2021). The Development and Applications of Food Knowledge Graphs in the Food Science and Industry. (pp. 1–45). URL: http://arxiv.org/abs/2107.05869.

Nagesh, N., Azimi, I., Andriola, T., Rahmani, A. M., & Jain, R. (2023). Towards deep personal lifestyle models using multimodal N-of-1 data. In *International conference on multimedia modeling* (pp. 589–600). Springer.

Ng, Y.-K., & Jin, M. (2017). Personalized recipe recommendations for toddlers based on nutrient intake and food preferences. In *Proceedings of the 9th international conference on management of digital ecosystems* (pp. 243–250).

Niela-Vilen, H., Azimi, I., Suorsa, K., Sarhaddi, F., Stenholm, S., Liljeberg, P., et al. (2022). Comparison of oura smart ring against ActiGraph accelerometer for measurement of physical activity and sedentary time in a free-living context. *CIN: Computers, Informatics, Nursing*.

Nutritionix - largest verified nutrition database. (2024). https://www.nutritionix.com/. (Accessed March 2024).

Oura ring. Smart ring for fitness, stress, sleep and health. (2024). https://ouraring.com/. (Accessed March 2024).

Pearl, J. (2012). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. *Causality: Statistical Perspectives and Applications*, 151–179.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological methods*, *19*(4), 459.

Raut, M., Prabhu, K., Fatehpuria, R., Bangar, S., & Sahu, S. (2018). A personalized diet recommendation system using fuzzy ontology. *International Journal of Science and Engineering Invention*, *7*(3), 51–55.

Robertson, S., Zaragoza, H., et al. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval*, *3*(4), 333–389.

Rossi, L., Ferrari, M., & Ghiselli, A. (2023). The alignment of recommendations of dietary guidelines with sustainability aspects: Lessons learned from Italy's example and proposals for future development. *Nutrients*, *15*(3), 542.

Rostami, A., Pandey, V., Nag, N., Wang, V., & Jain, R. (2020). Personal food model. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 4416–4424).

Sapri, N. S. M., Abdul-Rahman, S., Benjamin, A. M., et al. (2019). A diet recommendation for diabetic patients using integer programming. In *AIP conference proceedings*: *vol. 2138*, (no. 1), AIP Publishing.

Shalit, U., Johansson, F. D., & Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning* (pp. 3076–3085). PMLR.

Sharma, A., & Kiciman, E. (2020). DoWhy: An end-to-end library for causal inference. arXiv:2011.04216.

Sikka, K. (2019). Learning user preferences from social multimedia analysis and overview of the IFood2019Challenge. In *Proceedings of the 5th international workshop on multimedia assisted dietary management* (p. 18). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3347448.3352812.

Singh, R. K., Chang, H. W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., et al. (2017). Influence of diet on the gut microbiome and implications for human health. *Journal of Translational Medicine*, *15*(1), 1–17. http://dx.doi.org/10.1186/s12967-017-1175-y.

Story, M., Kaphingst, K. M., Robinson-O'Brien, R., & Glanz, K. (2008). Creating healthy food and eating environments: policy and environmental approaches. *Annual Review of Public Health*, *29*, 253–272.

Taweel, A., Barakat, L., Miles, S., Cioara, T., Anghel, I., Tawil, A.-R. H., et al. (2016). A service-based system for malnutrition prevention and self-management. *Computer Standards & Interfaces*, *48*, 225–233.

Toledo, R. Y., Alzahrani, A. A., & Martinez, L. (2019). A food recommender system considering nutritional information and user preferences. *IEEE Access*, *7*, 96695–96711.

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ (Online)*, *361*, 36–44. http://dx.doi.org/10.1136/bmj.k2179.

Wang, L., & Lim, E.-P. (2023). Zero-shot next-item recommendation using large pretrained language models. arXiv preprint arXiv:2304.03153.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. (2021). A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *15*(5), 1–46.

Zadeh, M. S. A. T., Li, J., & Alian, S. (2019). Personalized meal planning for diabetic patients using a multi-criteria decision-making approach. In *2019 IEEE international conference on e-health networking, application & services* (pp. 1–6). IEEE.

Zeraatkar, D., Johnston, B. C., & Guyatt, G. (2019). Evidence collection and evaluation for the development of dietary guidelines and public policy on nutrition. *Annual Review of Nutrition*, *39*, 227–247.

Zhang, Y., Ding, H., Shui, Z., Ma, Y., Zou, J., Deoras, A., et al. (2021). Language models as recommender systems: Evaluations and limitations.

Zhang, J., Xie, R., Hou, Y., Zhao, W. X., Lin, L., & Wen, J.-R. (2023). Recommendation as instruction following: A large language model empowered recommendation approach. arXiv preprint arXiv:2305.07001.