Integrating Expertise in LLMs: Crafting a Customized Nutrition Assistant with Refined Template Instructions

Annalisa Szymanski
Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
aszyman2@nd.edu

Brianna L. Wimer
Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
bwimer@nd.edu

Oghenemaro Anuyah
Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
oanuyah@nd.edu

Heather A. Eicher-Miller
Nutrition Science
Purdue University
West Lafayette, IN, USA
heicherm@purdue.edu

Ronald A. Metoyer
Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
rmetoyer@nd.edu

ABSTRACT

Large Language Models (LLMs) have the potential to contribute to the fields of nutrition and dietetics in generating food product explanations that facilitate informed food selections. However, the extent to which these models offer effective and accurate information remains unverified. In collaboration with registered dietitians (RDs), we evaluate the strengths and weaknesses of LLMs in providing accurate and personalized nutrition information. Through a mixed-methods approach, RDs validated GPT-4 outputs at various levels of prompt specificity, which led to the development of design guidelines used to prompt LLMs for nutrition information. We tested these guidelines by creating a GPT prototype, *The Food Product Nutrition Assistant*, tailored for food product explanations. This prototype was refined and evaluated in focus groups with RDs. We find that the implementation of these dietitian-reviewed template instructions enhance the generation of detailed food product descriptions and tailored nutrition information.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Large Language Models, Artificial Intelligence, Food Recommendations

ACM Reference Format:
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 CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3613904.3641924

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3641924

1 INTRODUCTION

Consumers have access to a wide range of food information to aid them in making food decisions. This includes nutrition labels [18, 44, 57], macro-nutrient profiles [15, 35, 63], ingredient lists [13, 36, 54], allergen information [8], and specific health claims [6, 60]. Applying these details to guide food selection is essential as it empowers individuals to make informed choices that promote health and well-being [40, 43, 54, 55]. Registered dietitians (RDs) specialize in customizing diets to meet individual nutrient needs that align with the unique goals of consumers [39]. This personalized dietary advice not only promotes more informed decision-making, but also enhances overall nutrition education and understanding among those receiving their services [39, 69]. However, in economically disadvantaged areas, access to dietitians may be limited, leaving communities with less access to information on healthful dietary choices [67]. Large Language Models (LLMs) offer a potential solution by generating comprehensive explanations of food products and providing a context for the healthful or less healthful aspects of food. This could be particularly helpful in regions where the translation of nutrition information may be scarce or unavailable [31].

Similarly, food recommendation systems aim to assist consumers in making better choices by offering tailored suggestions based on an individual's goals [64]. However, these systems commonly lack transparency and fail to provide users with the underlying rationale for the recommended products [50]. We propose utilizing LLMs to craft detailed explanations for the translation of nutrition information of food products, taking into consideration personalized data that dietitians typically factor into their services. LLMs can be built to consider both the product and the dietary goals of the user to generate food explanations that assess whether a product is a suitable or unsuitable choice for the user, accompanied by detailed reasoning. In this study, we define "food explanations" as detailed descriptions of the nutritional and health qualities of a food product that are tailored to an individual's dietary and health needs.

LLMs hold promise for providing dietary information, yet it is imperative to address the inherent risks and potential inaccuracies these models might present. Although LLMs have the capacity to handle extensive datasets and generate explanations, their reliability is ultimately contingent on the quality of the data from which they were trained. There is a concern that the models might perpetuate existing biases, misunderstandings, and information that is not evidence-based, particularly when inaccurate dietary information may be present [12, 21, 22, 46, 56]. Furthermore, using LLM-generated advice can have limitations. For instance, without considering individual dietary behaviors or habits, LLMs might provide information that lacks context within the overall diet. Neglecting cultural or regional food preferences may also require further investigation. In socioeconomically challenged regions, where access to tailored dietary information may be limited, inaccurate information could exacerbate existing nutritional challenges and reinforce the perception that healthful food is not economical, making it even more difficult for individuals to make informed decisions [7, 11, 19]. Thus, while LLMs hold promise as tools to increase access to dietary information, rigorous validations and the transparency of limitations are essential to ensure the safety and efficacy of the food explanations they provide.

Our primary objective is to engineer LLMs in collaboration with RDs to produce explanations on food products that meet dietitian standards and mitigate reliability and accuracy concerns. We use GPT-4 due to its widespread adoption as a primary information source and advanced natural language processing capabilities [33].

To achieve our goal, a mixed-methods study is first conducted with RDs to investigate the strengths and weaknesses of GPT-4 in providing food product explanations. The RDs validated the outputs generated at three levels of prompt specificity and offered insights on the effectiveness of the model in conveying nutritional and health information about the food items. Their feedback was reviewed to determine design guidelines to be followed when using LLMs for generating nutrition information.

Using our design guidelines, we developed a customized GPT prototype that integrates crafted template instructions. This customized GPT prototype was evaluated and refined in two collaborative focus groups with dietitians to determine whether our validation findings could be mitigated. The focus groups involved an iterative process of updating the prototype instructions to tailor the output. The customized and refined prototype is named *The Food Product Nutrition Assistant*. We detail the development and potential of the prototype, highlighting its specific design and functionality. By integrating RD expertise into GPT-4's framework, we aim to enhance the model's utility in providing nutritionally sound and personalized food product explanations. Our paper makes the following contributions:

- We develop an empirical understanding of the capability of LLMs in generating food product explanations, particularly focusing on the perspectives and evaluations of dietitians.
- We formulate a set of design guidelines based on feedback from registered dietitians that may inform future researchers on how to enhance the performance and face validity of GPT-4 in providing food explanations.
- We develop a customized GPT prototype using template instructions based on our design guidelines and assess its effectiveness in a collaborative focus group comprised of registered dietitians.

Our work takes a step toward improving the design and outputs of LLMs through the use of template instructions and the development of a customized GPT prototype. This focus is important, considering the increasing reliance on digital tools for health and nutritional guidance. We aim to make a meaningful impact on improving consumer education on dietary choices. This customized GPT prototype has the potential to become a valuable asset in guiding consumers, especially in areas and among groups where access to professional dietary advice is limited.

2 RELATED WORK

Our examination of existing literature focuses on the need for food explanations to improve food literacy, the role of dietitians in health coaching, and the status of current capabilities and limitations of Artificial Intelligence (AI), particularly LLMs, in the domain of nutrition and dietetics. This review will aid in framing our study within the broader context of ongoing research and identifying gaps that our investigation seeks to address.

2.1 Food Literacy

To promote food literacy, defined broadly as proficiency in food related skills and knowledge [14], there have been recent technological advances that focus on a holistic approach to diet, encouraging informed food choices and promoting healthier eating habits [10, 51]. Perry et al. consolidates various aspects of food-related knowledge, skills, confidence, ecological factors, and decision-making processes as food literacy attributes [52]. Their study emphasizes the importance of not only knowing about healthy foods but also possessing the practical skills and confidence to make informed food choices, considering both personal and broader environmental contexts [52]. It was suggested that effective communication of food and nutrition information using commonly understood language empowers consumers to make choices that align with their health goals [52], which emphasizes the need for quality food product explanations.

Similarly, Dillahunt et al. discuss the importance of designing technology to support nutrition and health knowledge in relation to consumer agency [16]. Understanding of food labels, facilitating budget optimization, and promoting self-efficacy for healthy food choices were noted as important components to building higher-agency behaviors for overall health and well-being, especially in food desert areas [16]. Other findings have discussed that there are challenges in achieving nutrition knowledge because of the amount of information to consider, such as interpreting nutrition labels, managing portion sizes, aligning food to health needs, and cost, all of which could be addressed in a well designed explanation of the nutrition composition and relevance of a specific food or food product [24]. In addition, Bomfim et al. also highlights the need for awareness, knowledge, and skills among consumers when designing food-related technology [9]. Features that promote interpretation of nutrition content, understanding of ingredient listings, knowledge of food guidelines, and awareness of healthy food alternatives are important to incorporate in technological design to support food literacy [9]. These features of Bomfim's comprehensive review are very relevant to food explanations and must be incorporated when exploring the use of LLMs [9].

2.2 RDs in Health Coaching

Personalized nutrition advice from registered dietitians has been shown to have positive effects on nutritional intake, nutritional status, and clinical outcomes, especially when a holistic approach is utilized in providing nutritional care [2]. It has been found that dietitian customized regimens that take into consideration an individual's tendency for behavioral change, motivation for food selection, and lifestyle influences, will ultimately empower individuals and place them in command of their dietary needs [1]. In addition, care by RDs encompasses the linkages in food insecurity, food systems, dietary consumption, health promotion, and chronic disease prevention and treatment which has been shown to increase agency to make better health decisions [3, 27, 61]. Similarly, Mitchell et al. found that text-based virtual health coaching via chatbots was advantageous in promoting options to foster client autonomy [42], bringing evidence that using LLMs for health related purposes could be beneficial. Using generative explanations from LLMs can assist with health coaching by providing a comprehensive understanding of a food product and giving the user the agency to make informed decisions that can be tailored to individualized dietary needs.

2.3 Leveraging Explainable Suggestions to Enhance Nutrition Understanding

The abundance of information in healthcare and nutrition at the point of purchase is vast, leading to the development of food recommendation systems. These systems guide users towards healthier eating choices, yet they often lack transparency in their decision-making processes [64]. The need for Explainable AI (XAI) is emphasized to improve trust in these systems by providing clearer explanations for food recommendations [50]. Current platforms, while suggesting foods based on criteria such as eating habits or dietary preferences, typically do not offer detailed justifications for their choices [20, 43, 62, 66]. LLMs can bridge this gap by generating user-focused evaluations, enhancing user trust and understanding of AI-driven recommendations [50].

Moreover, integrating comprehensive explanations makes AI systems more accessible and understandable to users. Dragoni et al. demonstrate the effectiveness of using natural language generation for this purpose, aiding users in adhering to dietary guidelines through clear explanations [17]. Future collaborations with dietitians could lead to the creation of AI systems that are not only transparent but also ensure trustworthy and beneficial recommendations for users.

2.4 Challenges and Prospects of AI Integration in Nutrition and Dietary Guidance

While there is mounting evidence that AI can offer medical information, aid in patient support, and provide dietary information with varying degrees of accuracy and consistency, no substantial validation has been documented for AI-generated nutritional information pertaining to individual food products [5, 12, 21, 25, 26, 30, 34, 46, 56, 58]. Furthermore, although AI is gaining popularity among the general public, its incorporation into clinical practice or public health promotion remains limited because of a range of challenges, primarily the absence of trustworthy and well-documented information to strengthen AI's knowledge base [37]. Before utilizing AI for effective food translation to promote better health, it is essential to ensure that the system accurately represents the nutritional content of the numerous individual food items included in an overall purchasing context. While it is challenging to control public use, it remains crucial to confirm the face validity and appropriateness of AI models before considering their practicality for dependable healthcare or prevention applications. A compelling need exists for AI and HCI communities to foster meaningful partnerships within the broader healthcare sector, particularly with healthcare and nutrition experts [45, 59].

2.5 Large Language Models in Nutrition and Dietary Guidance

Chatelan et al. investigated ChatGPT's ability to provide nutritional guidance, such as diet plans and recipes, for individuals with various health conditions such as Type 2 Diabetes. The study found that while ChatGPT's advice was generally understandable and aligned with the American Diabetes Association information, it often lacked accuracy, included inappropriate food choices, and had inconsistent and poorly referenced responses which pose challenges in clinical reasoning [12]. Further studies have revealed ChatGPT's limitations in identifying misleading responses, particularly in crafting diets for people with food allergies and in generating meal plans aimed at weight loss [21]. Other AI-created meal plans often displayed potential risks due to imbalanced nutritional content and monotonous diets that lack variety [46]. Such inadequacies have raised concerns about the model's nutritional adequacy and safety. Garcia et al. also noted the limitations in using ChatGPT for nutrition knowledge, particularly in personalized meal planning and dietary advice [23]. Despite acknowledging its potential, the studies collectively suggest a critical need for in-depth validation by experts to address the weaknesses and to ensure the model's reliability and safety in nutritional guidance.

2.6 Prompt Engineering and AI Fine-Tuning

Through a prompt [38] or a series of instructions, LLMs can have a noticeable influence on subsequent interactions and the resulting output it generates. By offering precise directives and guidelines for an LLM's engagement, a prompt establishes the conversational context, stipulates the significance of specific information, and defines the preferred format and content of the desired output [68]. It has been further shown that the caliber of the outputs produced by a conversational LLM is directly related to the quality of the prompts furnished by the user [68]. A significant method for refining the quality of the prompt instruction process involves the application of supervised fine-tuning (SFT) with the aim of optimizing the performance of prominent commercial LLMs, such as ChatGPT. Fine-tuned prompts serve as instrumental tools for configuring interactions between the user and the LLM to enhance the proficiency of LLMs in addressing a diverse range of information [28, 49]. Therefore, it is imperative that a meticulously crafted prompt be employed to guide the excellence of the task performance of LLMs.

3 STUDY 1: VALIDATION OF LLM OUTPUTS

To better understand the strengths and weaknesses of LLMs in generating food explanations, we conducted a validation study with registered dietitians. In this section, we present details on prompt engineering and evaluation criteria for generating food explanations using GPT-4. Additionally, we offer insights into our study protocols, a review of participant details, and an outline of our data analysis methods. The results are then discussed in the context of key findings. Based on these findings, we have formulated a series of design guidelines for future reference.

3.1 Study Overview

We conducted a comprehensive mixed-methods analysis involving twelve registered dietitians to assess the capabilities and limitations of LLMs in generating food product explanations. We utilized GPT-4 to generate the outputs that analyzed food products based on a specified set of criteria.

We designed prompts encompassing three distinct levels of input specificity, with the goal of enabling the registered dietitians to offer feedback on the responses generated with varying degrees of prompt input details. Figure 1A shows the three different levels of prompt specificity. In Level 1, the prompt was limited to basic details of only the product name. Level 2 was expanded to incorporate more comprehensive data that included the Nutrition Facts label and ingredients list provided by the product manufacturer. At Level 3, in addition to the nutrition label and ingredients, we integrated individual dietary needs and goals of a potential user that may differentiate any resultant variations in the model's output. Figure 1B shows an example of the prompt with Level 3 specificity. This approach helped us understand how varying amounts of detail influenced the LLM's generation of the product explanations.

Each dietitian independently was tasked with reviewing and providing feedback on five different products (see Appendix A Table 2) across three levels of specificity in a single virtual interview session. Following each evaluation, the participant used a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) to rate the output based on four different measures: coherence, conciseness, quality, and accuracy. These measures have been similarity used in other studies when assessing generative language models [53]. Additionally, participants were asked to give verbal feedback after examining each level of specificity related to a product. Their expert feedback served as a critical component in understanding the face validity and practical applicability of LLMs in the realm of nutrition and dietetics. Our research questions for this study are as follows:

(1) **RQ1:** What are the advantages and limitations associated with the use of LLMs for producing explanations for various food products?
(2) **RQ2:** In the context of utilizing LLMs, how does the specificity of information given in the input prompt influence the effectiveness of the generated explanations, as evaluated by registered dietitians?
(3) **RQ3:** What design guidelines should be prioritized when applying LLMs for translation of dietary information for consumers?

3.2 Generating Food Product Explanations through LLMs

3.2.1 Model Selection and Rationalization. For our study, we selected GPT-4 due to its vast knowledge base and problem-solving abilities [33]. ChatGPT, a platform built upon the GPT-4 model, was one of the largest publicly known language models at the time of our study. The substantial size of this model is often associated with its enhanced capability to understand and generate contextually relevant text, making it particularly suitable for our study within the domain of nutrition and dietetics. We utilized single-prompt conversations with ChatGPT performed in English using a new chat for each prompt, and making each observation independent.

3.2.2 Evaluation Criteria and Prompt Engineering. Prompts were crafted in collaboration with a nutrition scientist who is a co-author on this paper. We limited the criteria of the explanation to analyze the product based on five different predetermined factors as shown in Figure 1B: Health Benefits, Allergens, Dietary Restrictions, Shelf Life v. Cost, and Healthier Product Substitutes. These factors were added to provide a structure to the output when assessing the healthful or less healthful aspects of the food product.

To ascertain the adaptability and precision of the generated product explanations, we examined the output crafted by GPT-4 based on three different variation levels of detail given to the input as shown in Figure 1A. The first level of specificity (Level 1) used only the product name and size. Recognizing that consumers often rely on nutrition labels when evaluating food products, the second level of specificity (Level 2) was aimed to understand the model's ability to accurately analyze and convey the detailed nutritional and ingredient details of a product within its responses. Because dietary goals play a pivotal role in product selection for many individuals, we hypothesized that the model might factor in different aspects of the provided information when customizing explanations for distinct dietary requirements that could be provided by an individual. Thus, the Level 3 specificity introduced a hypothetical user's dietary needs and goals as shown in Figure 1B.

We chose manufactured food products to generate explanations from GPT-4 due to their diverse nutrient compositions. A key factor in selecting these items was the consistent availability of the Nutrition Facts labels and ingredients list, which are often lacking for fresh fruits and vegetables. Unlike raw produce, which generally have a consistent nutrient profile, the manufactured products of a certain food item display a wide range of nutritional variations. This diversity, such as the varying sodium content in different cereal brands, adds complexity to consumer decision-making. These products represent each of the five MyPlate food groups [47] and were selected from Walmart's website for their varied nutrient compositions (see Appendix A Table 2).

3.3 Interview Procedures

We conducted semi-structured interviews that centered on the LLM generated product evaluations. Each registered dietitian was exposed to fifteen total product evaluations across the five different food products, progressing through each of the three levels of specificity for each product. Initially, they were introduced to the basic prompt output, then to the 2nd Level of specificity, and next the

A

Specificity Level	Information Included
Level 1	Product Name and Size only
Level 2	Product Name and Size + **Nutrition Facts Label and Ingredient Information**
Level 3	Product Name and Size + Nutrition Facts Label and Ingredient Information + **Individual's Dietary Goals**

B

Prompt Example: Level 3 Specificity

Analyze (Product} in {Size}. Include the nutrition label and ingredient information in your evaluation, and evaluate the following aspects to determine whether this product is a good or bad food choice for the user portfolio below:

1. Health Benefits: Identify the nutritional components and explain how they contribute to or detract from a healthy diet.
2. Allergens: List any known allergens in the product and discuss potential risks.
3. Dietary Restrictions: Assess how this product aligns with common dietary needs or restrictions, such as vegan, gluten-free, low-sodium, low-sugar, or low saturated fat diets.
4. Shelf Life vs. Cost: Compare the shelf life of the product with its cost and discuss how this relationship may influence the purchasing decision.
5. Healthier Product Substitutes: Suggest specific product alternatives from known brands that may be considered healthier or more suitable for specific dietary needs or preferences.

{Nutrition Facts Label}

{Ingredients)

Individual's Dietary Needs and Goals:

1. Reduce added sugars and refined carbohydrates to manage blood sugar levels
2. Increase intake of fiber-rich foods, such as whole grains, vegetables, and fruits
3. Avoid lactose-containing products due to lactose intolerance
4. Incorporate lean proteins like fish, chicken, and plant-based proteins
5. Maintain a balanced diet to lose weight and reduce the risk of heart disease

Please provide your analysis in a single paragraph, and keep the response under 250 words. Please provide a reference and source for your evaluation.

Figure 1: GPT Prompt Input details. A) Levels of prompt specificity. B) GPT Prompt Input Example: Level 3 Specificity.

3rd Level of specificity detailed with hypothetical dietary needs and goals. Each dietitian was informed about each of the levels of specificity to provide clarity on the considerations taken into account when generating the LLM explanations.

To ensure an unbiased evaluation and control for any order effects in the study, we employed a counterbalancing method across participants. While the specificity levels were fixed for each product, the sequence in which these products were presented varied for each individual. This systematic rotation of the products not only minimized potential biases but also provided a more comprehensive understanding of the explanations irrespective of the sequence in which they were assessed.

Upon being presented with each output, dietitians were accorded 2-3 minutes to analyze each explanation using the five primary criteria provided. Subsequently, they rated these evaluations on a 5-point Likert scale, gauging coherence, conciseness, accuracy, and content quality. Each dietitian received the standard Nutrition Facts label and ingredients list for the products to validate the output, which was the same as utilized in the prompt Level 2 and 3 specificity.

After independently reviewing and rating each product, participants were again presented with all three explanations corresponding to that product. They were then prompted with open-ended questions, probing their overall impressions of the model outputs, any potential concerns, discernible risks or challenges, perceived strengths, and suggested areas for improvement. Each virtual session lasted approximately 90 minutes with participants sharing their screens throughout the entirety of the study.

3.4 Registered Dietitian Recruitment

We recruited twelve RDs to do the evaluations. Participants were recruited through a network of university campuses and community dietitian groups, reached via email recruitment. All participants were compensated with a $100 gift card at the end of the study.

Upon agreeing to participate, the dietitians were first presented with a pre-study questionnaire and a consent form. We gathered insights into their educational background, specialty, workplace setting, tenure in the profession, and familiarity with AI. The participating RDs represented a spectrum of five specialties, namely clinical, community & public health, food service, research and education, and sports and wellness nutrition. Their professional engagements spanned diverse environments, such as hospitals, academic institutions, schools, and the broader food industry. Their experience in the field was varied, with durations ranging from 1-30 years. Only two of the twelve professionals had previously integrated AI into their nutrition-centric work or research endeavors. Furthermore, when questioned about their acquaintance with AI and its applications in food and nutrition analysis, the consensus was a marked lack of familiarity.

3.5 Data Analysis Methods

3.5.1 Likert Scale Analysis. The central aim of our analysis was to explore the variance in mean scores for four measures (coherence, conciseness, quality, and accuracy) at different levels of specificity to address our second research question (RQ2). Specifically, we investigated whether RDs assigned higher ratings to food product evaluations when they were supplemented with additional contextual details such as Nutrition Facts labels, ingredients lists, or dietary goals, as opposed to evaluations based solely on the product name and size. This approach was intended to assess the impact of enhanced context in the generation of food product explanations.

To achieve this, we first aggregated the scores from all five evaluated products at each level of prompt specificity. We then calculated the mean Likert score, standard deviation, and standard error for each respective measure. These statistical measures were obtained from the Likert scale assessments, which ranged from 1 (strongly disagree) to 5 (strongly agree). The scale provided insights into the RDs' evaluation of coherence, conciseness, quality, and accuracy in the explanations generated by the model for each product.

3.5.2 Qualitative Analysis. To analyze the qualitative data we followed standard open-coding procedures. Two researchers engaged in the coding process. Each independently read three interviews to familiarize themselves with the data. They then created initial codes and coded the three interviews based on their findings. They then reconvened in meetings to discuss and reconcile these initial codes. Subsequently, the rest of the interviews were coded using these codes. New codes were added as they emerged from the interviews. The codes were then organized and regrouped to develop themes. These themes were consolidated and evolved into key findings, which are detailed in the results section.

3.6 Findings

3.6.1 Quantitative Results. In this section, we report the results for coherence, conciseness, accuracy, and quality across different levels of specificity as detailed in Table 1.

The findings show that for coherence, quality, conciseness, and accuracy, there is a trend for the mean outcomes at Level 2 to be more favorable compared to Level 1, yet means overall had a somewhat similar range of 3.7-4.5. The trend may suggest that the inclusion of the Nutrition Facts label and ingredients list in the prompt (Level 2) has a tendency to show an overall slight positive effect on the explanations. However, the trend varied at Level 3. For coherence and quality, the mean slightly dropped from Level 2, indicating that the addition of individual dietary goals to the Nutrition Facts label and ingredients does not necessarily contribute positively to these aspects. This could indicate the complexity of the relationship between nutrition information and the GPT's ability to relate nutrition to individual dietary needs. Conversely, for conciseness and accuracy, the mean scores for Level 3 were slightly higher than Level 2, again suggesting the need for a careful balance in information provision.

3.6.2 Qualitative Results. Our findings provide valuable insights relevant to our first research question (RQ1), which examines the advantages and drawbacks of LLMs in generating food explanations. We have organized these themes into five key findings (KF) and give examples of each theme with statements from registered dietitians, denoted as RD1, RD2, and so on (RD#) in Appendix B Table 3). This section presents qualitative insights derived from experts to guide the discussion on the utilization of LLMs in providing information about food products salient to individuals at the point of food purchasing. We report the following key findings:

Table 1: The mean, standard deviation (SD), and standard error (SE) for all four metrics: Coherence (a), Quality (b), Conciseness (c), and Accuracy (d). Data is aggregated by the level of specificity.

	Coherence (a)			Quality (b)		
Specificity	Mean	SD	SE	Mean	SD	SE
Level 1	3.95	1.03	0.13	3.73	1.16	0.15
Level 2	4.45	0.75	0.10	4.27	0.94	0.12
Level 3	4.35	0.82	0.11	4.22	0.96	0.12

	Conciseness (c)			Accuracy (d)		
Specificity	Mean	SD	SE	Mean	SD	SE
Level 1	3.92	1.08	0.14	3.77	1.24	0.16
Level 2	4.30	0.77	0.10	4.23	1.01	0.13
Level 3	4.32	0.79	0.10	4.40	0.91	0.12

KF1: The outputs generated by the more detailed specificity prompts are preferred by dietitians.

Specificity Level of Input. The feedback from the registered dietitians suggests a trend of preference toward Level 2 and Level 3 specificity over Level 1 because the outputs provided more comprehensive nutrition information and individualized tailoring. Participants noted that the outputs at these levels included detailed references to macro and micronutrient amounts and Daily Value percentages (%DV). Conversely, participants observed that with the less instructive prompt (Level 1), the explanations were more generalized, lacked reference to specific nutrients within the food product, and had ambiguous text regarding whether certain nutrients and ingredients were actually present. For instance, the output initially indicated a product might not be gluten-free, but later confirmed it was gluten-free after the Nutrition Facts label and ingredients list was included in the input. Also, the dietitians noted that the addition of the dietary goals (Level 3) did lead to more individualized tailoring, which was preferred. However, this did not necessarily mean that there were no errors or misleading statements in the outputs.

KF2: The outputs do not align with the standards upheld by registered dietitians.

Reliability. During the interviews, the RDs found instances of irrelevant information within the output explanations. These discrepancies involved misinterpretation regarding suitability for specific diets or about potential allergens present. In one instance involving the Bubba Burger at Specificity Level 1, the output included the possibility of dairy allergens. The dietitians noted that discussing issues about dairy allergens in relation to meat appeared out of context and lacked clear relevance. While not necessarily incorrect—the Bubba Burger may contain dairy allergens when cheese is added—the model's approach to discussing allergens seemed indiscriminate, as if it aimed to cover all possible allergens without the targeted focus and reasoning a dietitian would typically apply. Moreover, the language used in these outputs often included speculative terms, such as "may" or "can", which introduced uncertainty when discussing potential allergens or compatibility with dietary goals. This approach can be misleading, especially when precise information is crucial for dietary choices.

Sources of Information. Dietitians observed that the shortcomings in output quality seem to originate from the reference materials utilized by the model during product evaluations. It was noted that the output did not consistently refer to authoritative sources, such as the Dietary Guidelines for Americans, the American Heart Association Diet and Lifestyle Recommendations, the American Diabetes Association guidelines, and MyPlate, which are standard references for professionals in offering food product suggestions to their clients. Most dietitians prefer to confirm the information they share from good sources before informing clients, whereas the output often fails to provide a reference.

KF3: There is a prevalence of falsehoods which undermines the clarity and coherence of the explanations and could lead consumers to form incorrect conclusions.

Imprecise "Buzz Words". The collective view among the 12 interviewed registered dietitians reveals that outputs provide product descriptions that fall short of the precision and comprehensive accuracy usually expected from professionals in the nutrition sector. The outputs often incorporate "buzz words" as highlighted by one of the interviewed dietitians. For example, for the Campbell's Soup Level 2 specificity, the output was said to be "a healthy dose" of fiber and potassium without fully substantiating why.

Prevalent Errors. With the inclusion of the Nutrition Facts label incorporated in the prompt, the model can not always give an adequate interpretation of its meaning. In some cases, the output would state that the product may not be suitable for a low-sodium diet, yet the dietitians noted that the sodium amount fit the federal guidelines to be a "low sodium" product. It was concluded by the dietitians that the reference of "low" or "high" when describing a product did not align with dietary guidelines.

Misleading statements. The registered dietitians found that there was a tendency to provide misleading remarks and inappropriately use specific terminologies when describing some products. A repeated concern was that the model was suggesting that gluten-free, lactose-free, or organic products may be healthier alternatives. RDs clarified that organic products are not inherently healthier and are likely to be less economical when purchasing manufactured products. It was emphasized by RDs that gluten-free products are only beneficial for individuals with celiac disease and it is incorrect to imply that they may be considered "healthier". The RDs noted that a rudimentary food explanation could have an overall negative rather than a positive effect on the consumer through misinterpretation. In addition, the model referred to the vitamin C and fiber content of Dole Fruit Cups (Level 1) to be contributing to immune function and digestive health. However, a registered dietitian noted that the product had such a small amount of fiber that she would not have mentioned it in counseling a client.

Fictitious Substitutions. The explanations sometimes also suggest food substitutes that are not marketed and have never existed. In their validations, the RDs found that the substitutes for some products, such as soy or almond based cottage cheese, were not available or sold in retail.

KF4: The model fails to maximize the comprehensiveness and educational value of its output.

Lack of Educational Context. A notable observation by the RDs was the lack of comprehensiveness in the explanations around the Daily Value percentage and nutrient amounts when presented in the outputs. Due to the concern of health literacy for the reader, dietitians suggest that there should be additional information provided on interpreting essential nutrients for meeting nutritional needs, supporting overall health, and adhering to dietary restrictions. The explanations do not clarify that the percent daily value is based on a specific calorie intake regimen that is not universally applicable across all genders and age groups. This oversight can potentially impede the model's effectiveness in conveying nutritional information.

Reading Level of Consumer and Health Literacy Support. The dietitians also observed that the vocabulary used may surpass the comprehension levels of consumers with lower reading proficiency, which could pose challenges in effective communication and user engagement. One dietitian described the language as sometimes too technical and that words such as "detrimental" or "detracting" may be confusing or misleading to the user without fully educating them.

KF5: The outputs outline potential alignments with consumers' immediate dietary needs, yet it does not provide information on broader contextual factors or integrate insights into an individual's comprehensive diet.

Customization and Consumer Health Benefit. While inputting dietary goals in the prompt did result in personalized information in the output, the consensus among nutrition experts was that the output could have been further customized to address how the described food item aligns with specific dietary requirements. RDs explained the lack of context as to how the food fits within the daily diet, how it might be paired with other foods (ie. cottage cheese with fruit), and how caloric value could influence satiation. In addition, the output should consider how products that are commonly used together (ie. cereal with milk) may change the nutritional value and impact individual dietary needs (ie. such as lactose intolerance).

Sometimes the explanation tries to educate the consumer on the product, but does so in a manner that undermines or exaggerates the health benefits. For example, an output suggested that eating Bubba Burgers could lead to muscle growth without explaining which nutrient helps to build muscle, how much of the nutrient is necessary for muscle growth, or whether the dietary intake of other foods was also necessary. Given that some consumers may rely on the explanations as a source of advice, the RDs agreed that the explanations did not offer the necessary insights to help consumers gain a comprehensive understanding of which foods should be restricted, why avoidance might be essential, and how an increased intake of other foods could be more beneficial. Consequently, while the models may offer some useful information to certain consumers, its lack of comprehensiveness may restrict its utility for others.

3.7 Design Guidelines for Prompt Instruction

Our research points out key challenges in using GPT-4 for nutrition information. Building on the Key Findings, we see that there is a need to develop a structured set of prompt instructions to build customized GPT prototypes to try to mitigate the challenges. These instructions should consider the strengths and weaknesses of GPT to function as a blueprint for generating nuanced and personalized dietary explanations. This approach has been shown in prior research to mitigate errors and potential biases [32, 41]. Utilizing the insights gained from the formative interviews, we discuss a set of design guidelines (DG) to consider when building prompt instructions for food related information.

3.7.1 DG1 - Provide Product Label Information. Based on the empirical evidence from both quantitative and qualitative analyses, it is apparent that the addition of the Nutrition Facts label and ingredients list improved the model's output in terms of its references to nutritional information. Therefore, as evidenced in KF1, we recommend including both the product's nutrition label and the list of ingredients in the prompt instruction. This observation aligns with existing literature on prompt engineering techniques [68]. Notably, the LLM outputs at Levels 2 and 3 showed an improvement in describing the nutritional content as conveyed through the nutrition label and ingredient details, primarily through the utilization of quantifiable metrics, such as the percent daily value and milligram or gram quantities.

3.7.2 DG2 - Include Dietary Guidelines and MyPlate Sources. As seen in KF2, the dietitians noted that the LLM did not pull from validated sources such as the Dietary Guidelines for Americans when explaining food products. We recommend to instruct GPT to be fed with information from credible and authoritative sources, ensuring its advice is grounded in the most current and accurate dietary guidelines. Primary sources for integration should include the Dietary Guidelines for Americans and the MyPlate guidelines.

3.7.3 DG3 - Include Example Sentences. Observations from KF2 and KF3 indicate inconsistencies and misleading information in the outputs, along with a disparity in the quality and accuracy of sentence formation compared to professional dietitian standards. To address these issues, we propose the integration of example sentences as a template for guiding the model's output. This approach could involve structured sentences with placeholders for the GPT model to complete, or the provision of complete sentence templates for inclusion in the output. This method aims to enhance clarity, precision, and reliability in the information presented. Specifically we focus on the following clarifications:

- Allergen Wording: While the model generally identified allergens accurately, it sometimes failed to elaborate on why a product might not align with certain dietary restrictions due to the presence of those allergens (KF3). The purpose of the proposed example sentences would be to clarify why a product, due to its specific ingredient composition, is incompatible with specific allergen-free dietary goals. RDs also noted the importance of offering alternative options for those with dietary restrictions. We recommend that the prompt be crafted to not only identify unsuitable ingredients but also to suggest alternatives to assist individuals in finding suitable products that align with their dietary needs.
- Relate Product Alternatives to Nutritional Value and Cost: Dietitians have raised concerns regarding the cost implications of suggesting alternative products, particularly when these alternatives are more expensive without offering additional suitability (KF3). To address this issue, we propose the inclusion of a template sentence that guides users in understanding the trade-offs between cost and nutritional value. We can provide example sentences that suggest that to modify your intake of a nutrient you may want to consider a specific alternative product with the cost of the product applied. This aspect is particularly crucial in facilitating informed decisions when consumers are evaluating the advantages and disadvantages of purchasing a food product, especially when faced with a dilemma between cost and health benefits.

3.7.4 DG4 - Limit Words and Terminology. As seen in KF3, the use of buzzwords or misleading terms appeared in the outputs. Concerns from registered dietitians are that this could influence consumers to make dietary choices that are unsuitable. To correct this, we recommend the instructions limit the use of generic terms. Also, the dietitians suggested that the prompt be instructed to focus on specific nutrient amounts rather than Daily Value percentage numbers. This approach aims to make the model's outputs not only informative but also tailored to the unique dietary requirements and preferences of individual users.

3.7.5 DG5 - Instruct for Comprehensiveness and Balanced Diets. KF4 demonstrates that there were also shortcomings in output quality due to a lack of comprehensive contextual information relating to how the product fits within established dietary guidelines. Based on RD recommendations, the output should state the product suitability for a user's diet in relation to the product's nutrients. We recommend a prompt design that incorporates a holistic view of each food product in relation to total daily dietary goals as outlined in the MyPlate guidelines and Dietary Guidelines for Americans. This would entail not just a focus on nutrients in relation to meeting daily requirements, but also how the product fits into the broader context of a balanced diet.

3.7.6 DG6 - Specify Functionality and Scope. As mentioned in KF4, dietitians noted that the outputs often lack effective communications to provide explanations that were comprehensive enough to be understood by all audiences. From this, we recommend that the GPT be instructed to emulate a virtual dietitian through the in-depth analyses of food products that consider their nutritional content, and also offering advice tailored to individual dietary needs and preferences. The GPT should be directed to abstain from giving medical advice or diagnostics, and instead encourage users to consult healthcare professionals. In addition, because dietitians were concerned about the reading proficiency of users, we suggest instructing GPT to use a professional and informative tone and ensure the content is suitable for a 5th-grade reading level.

3.7.7 DG7 - Specify Guidelines on Nutrient Content. We observed inaccuracies (KF3) and a notable gap in nutrition education and health literacy (KF4), especially regarding understanding the Daily Value percentage and the quantities of nutrients present in various products as they may relate to a consumer's energy and specific health needs. To address this, we advise that the prompt establish clear parameters for labeling nutritional content with standardized terms. Dietitians recommended that outputs should categorize nutrient quantities into distinct groups of "low" and ""high" based on Daily Value percentages from Nutrition Facts label guidelines (KF3). The prompt instructions can facilitate this process by outlining the specific ranges that define each of the nutrient categories. For example, a sodium content of 4% daily value would be considered low by falling within the defined range in the template. This aims to improve consumer understanding and application of nutritional information.

3.7.8 DG8 - Include Comprehensive Nutritional Guidance and Food Pairing Suggestions. Our observations indicate that the model currently lacks the ability to furnish information considering broader contextual elements, as observed in KF5. Incorporating this aspect could potentially facilitate more accurate education for the user and promote a holistic understanding of daily dietary patterns. As advised by the dietitians, the instructions can encourage food pairings to create a nutritionally rich meal or can guide users in balancing their daily nutritional intake with combinations of various products. To enhance this aspect, we recommend incorporating suggestions from dietitians into our instructions that emphasize the importance of food pairings and balanced daily nutrition. Our instruction would guide the explanations to include how the product fits into a daily balanced diet.

4 STUDY 2: CREATING AND REFINING A GPT PROTOTYPE

We next outline our collaborative approach to design and evaluate a customized GPT prototype that uses a structured set of template instructions based on our design recommendations. We first describe our process of creating the prototype. Next, we detail the refinement and evaluation stages, involving follow-up focus groups with RDs. The RD focus groups assess the model's outputs and contribute to refining the template instructions. We present changes to the template agreed upon by dietitians in our focus groups.

4.1 GPT Prototype Creation

Based on the design guidelines outlined in Section 3.7, our research team developed a customized GPT prototype that integrates an initial set of template instructions and knowledge sources. This approach is intended to guide the GPT in generating responses that align more closely with dietitian standards.

November 2023 marked a significant development in the GPT realm with OpenAI DevDay's introduction of customizable GPTs [48]. These customized models enable users to integrate specific instructions and additional knowledge to offer a tailored AI experience. The configuration component of the GPT allows the developer to modify instructions and attach files to store information for the ChatGPT to pull from for output creation (referred to as the *knowledge* section). For the purpose of this study, we use the configuration component of the customized GPT to edit the instructions. Our template instructions are applied in conjunction with the customized GPT interface to create a prototype. We show an example of the customized prototype in Figure 2. In addition, a full set of initial template instructions can be found in the Appendix C Figure 4.

First, we provide the GPT with basic instruction information to include the role and tone to be used to drive the language of the output. As referenced in DG6, we instruct the GPT to function as a virtual registered dietitian and define its role to provide explanations about food products based on nutritional content and suitability for various dietary needs. This was done by first accessing the customized GPT builder and then manually configuring the GPT to enter specific instructions.

As mentioned in DG4, we advise limiting the use of phrases such as "healthy-dose" or "muscle-growth" to avoid over-simplification and ensure a more nuanced and accurate portrayal of nutritional advice. We also instructed the GPT to focus on specific nutrient amounts rather than percentages as a means to make the information more easily understood.

In addition, the custom GPT is designed to integrate file contents that the model can access through the custom GPT knowledge section (Section C in Figure 2). This allows for more subject matter expertise and general knowledge about nutrition and dietetics so that the GPT may provide more customized information. As outlined in DG2, we have incorporated the Dietary Guidelines for Americans and the MyPlate PDF files as foundational knowledge sources. We have also instructed the GPT to explain how the product contributes to the daily goals in the Dietary Guidelines for Americans and fits within the MyPlate recommendations (DG5). As also suggested in DG1, the inclusion of the Nutrition Facts label and ingredients list for all products tested are added to the knowledge section to increase the accuracy of specific nutrient amounts.

As instructed in DG5, the first sentence of the output should clearly state whether or not the product is suitable for the user depending on the dietary goals considered. We instructed the GPT to do this through the template instruction: *"This product is [good choice/ok choice/bad choice] for your diet because of [provide one primary reason why related to the nutrients]"*. In addition, we provided an example for the GPT to follow: *"Product is an ok choice for your diet because of its high fiber content, but it has a notable amount of added sugars."*

In our instructions, we ask the GPT to consider five factors (Nutritional Analysis, Dietary Restrictions, Allergens, Shelf Life v. Cost, and Healthier Product Substitutes). This allowed the dietitians to consider the same information as in the previous study. The instructions further guide the GPT in explaining the food product by providing example sentences using brackets where the GPT can fill in information. These sentences are used to mimic the language used by a dietitian and contain placeholders where the GPT can access the information, while also serving to connect the nutritional aspects of food items with individual dietary needs. For instance, an example sentence is: *"This product's [macro/micronutrient] is [amount] per serving which is considered [low (DV% <5%), high (DV% 20%+)] and [suitable/unsuitable] for [individual's dietary needs]."* Additionally, the examples extend to explanations on food pairings (DG8), dietary restrictions (DG7), allergens (DG3), and shelf life v. cost (DG3) considerations (see Appendix C Figure 4).

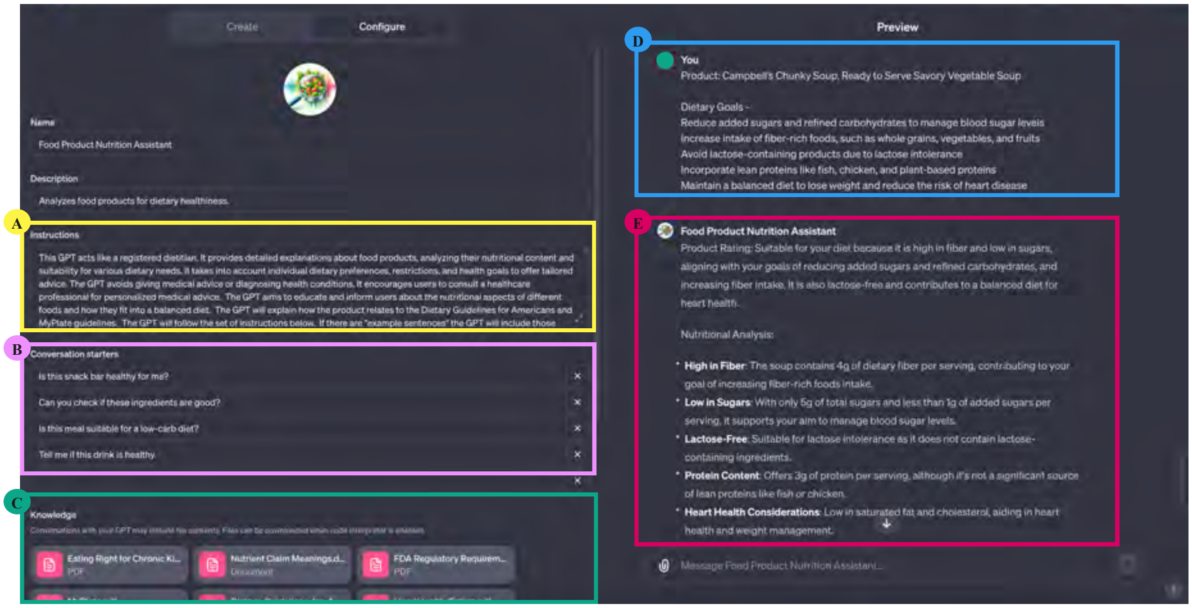

Figure 2: The refined "Food Product Nutrition Assistant" Customized GPT Prototype with features highlighted. A) Instruction input area, edited with the assistance of dietitians. B) Potential conversation starters to ask the GPT. C) Knowledge integration section for file uploads. D) User prompt for entering specific food product name and dietary goals. E) GPT's output with detailed food product explanations.

4.2 Focus Group Study Overview

In our focus groups, RDs played a key role in iteratively refining our initial template instructions within our prototype. This involved using the customized GPT prototype to provide real-time updates to the template instructions and generating corresponding outputs.

At the beginning of each session, the focus group was informed about the objective, which was to evaluate and refine the instructions in the GPT prototype with the aim to produce improved responses for food explanations. A screen was shared with the dietitians that presented the GPT prototype with initial template instructions and the resulting GPT response for a selected product and set of hypothetical dietary goals. The RDs were then asked to review the output. A discussion was then facilitated with the dietitians on further instructions or knowledge to be incorporated into the template. After the dietitians consolidated opinions and reached a consensus on the feedback, the template instructions in the prototype were updated. The revised output was then reviewed again by the dietitians. This process was repeated iteratively until a final, agreed-upon version was established.

Within the focus group, three different products were considered, two of which were used for the previous study in Section 3.2, and one new product. During the initial evaluation of the template, the same dietary goals were used as in Study 1. Towards the end of the focus group, the dietitians were asked to provide other dietary goals to see if the output could be tailored to accommodate.

4.3 Participants

A total of 6 RDs were recruited from our previous study. We organized a first and a second 90-minute focus group comprised of four and two RDs, respectively. Each focus group was convened in a single session and each participant was compensated with $100.

4.4 Prototype Evaluation and Refinement

In this section we highlight the iterations made by dietitians in the prototype. This involved both updating the template instructions as well as the knowledge sources. The final version of the template instructions used in the creation of the refined prototype is shown in Appendix C Figure 5.

4.4.1 Overall Format Structure and Organization. The consensus among dietitians is that there is an overall improvement in the format structure and organization of the output based on template instructions to act like a virtual dietitian. During the first focus group, the dietitians preferred the overall structure to not be in paragraph form. The dietitians requested that the GPT change the output structure to bullet points to organize the nutrition information.

In addition, Focus Group 1 raised the concern that the positive and negative aspects of the food were not clearly separated in the response. There was the concern that the GPT created a nutritional analysis which they found confusing to understand in relation to a healthy diet. To accommodate this, instructions were updated to the template to separate out the "good" and "bad" aspects of the food product under different headings. During this change, it was confirmed by RDs that the model was able to correctly group these aspects. Focus Group 2 agreed with the separation and changed the language of "good" and "bad", which they found to be subjective and vague to "Health Benefits" and "Considerations". Also, RDs in both focus groups confirmed that they preferred the readability of the output due to the 5th grade reading level limitation. A 2nd grade reading level had been tested but it was concluded that it was too minimal in information.

4.4.2 *Initial Product Rating.*

All dietitians agreed that there should be an overall suitability statement about the product and approved of the instructions in the example sentence as the first sentence. Through the reiterations and testing of each product, the model is able to follow the instructions to produce an initial product rating by filling in the information requested to form the sentence. However, initially the dietitians noted that the reasoning for the product needed to be improved. Through the reiteration, we found that providing an example sentence was the best way to get their preferred output. The example sentence for the template was crafted in collaboration and agreed upon by both focus groups. The rating of "good", "ok", and "bad" was changed to a product rating of "suitable" or "unsuitable" with reasoning included.

4.4.3 *Nutrition Explanation and Guidelines.*

The dietitians were at consensus in both focus groups that presentation of DV% be eliminated. Rather, product explanations were to include only nutrient amounts with high or low designations because dietitians in their practices find it easier to understand as it eliminates the relationship to a 2000 calorie intake. To accomplish this, dietitians worked together to update the template sentences when describing the nutritional aspects of the product. After this was complete, we saw that the nutritional analysis was able to follow a uniform structure when describing each nutrient by listing the amount, stating whether it is high/low, and providing an explanation as to why it is suitable/unsuitable to the individual's needs.

4.4.4 *Additional Template Sentence Instructions.*

Through evaluation, it was confirmed that the example sentences did help to modify the output to make the language consistent. Dietitians agreed with the template sentences that we had developed but there were a few additional added or edited. Pertaining to explaining the nutrition content, the dietitians worked together to add in template sentence instructions related to how to describe sugar intake by doing a conversion to the sugar amount in teaspoons. Dietitians also added additional example sentences to include allergens only if they pertained to the user's dietary goals. In addition, dietitians updated the template to also give brand name equivalents in the product alternatives to eliminate fictitious substitutions. An attempt was made to include a sentence in the template to identify whether products are covered by SNAP or WIC benefits, however it was determined that this remains a limitation of GPT to do correctly.

4.4.5 *Updates with Additional Knowledge.*

With regard to the knowledge included in the GPT instructions, the dietitians suggested that the template should always include guidelines related to common diseases such as diabetes, liver disease, kidney disease, and heart disease, and how products may or may not be suitable relative to those conditions.

4.5 The "Food Product Nutrition Assistant" Prototype

Based on our focus groups, we have refined a customized GPT prototype with specific prompt instructions and knowledge sources. This model, named "The Food Product Nutrition Assistant,"[1] can be provided with a specific food item and dietary objectives to generate a refined explanation about the product in alignment with the user's goals.

We show an example of our prototype output for the Campbell's Soup Product in Figure 3 in relation to the original output in Study 1. The output now provides specific nutrient information listed in milligrams/grams as opposed to DV% as guided by the RDs. After iterations, the format now describes nutrients in high/low categories with a statement on how they relate to a user's set of dietary goals. The new prototype incorporates a product rating, follows a bullet point structure, and organizes information under specific headings, all as reiterated and directed by the dietitians. Moreover, it now references Dietary Guidelines for Americans and MyPlate guidelines, while highlighting store brand substitutes and excluding references to organic substitutes as healthier options.

5 DISCUSSION

In the field of nutritional education and information, the use of ChatGPT, particularly in its GPT-4 version, presents both opportunities and challenges. Our study specifically examines the effectiveness of GPT-4 in providing explanations of food products tailored to individual dietary goals. Our efforts have explored the validation of GPT outputs and the design of a customized GPT prototype in collaboration with dietitians to standardize food product explanations for future use. We discuss the role of experts in output refinement and the use of customized GPTs for producing food product explanations.

5.1 The Importance of Experts in GPT Development

A key implication of our research is the importance of expert validation and collaboration in the design and implementation of LLMs, such as GPT-4. The involvement of nutrition experts was essential in validating outputs and refining the instructions used in the prototype. Their input ensured that the outputs were relevant and applicable, and contained minimal risks if put in front of users. This collaborative approach is important because it allows for the incorporation of specialized knowledge and content into the GPT model.

5.1.1 *Expert Validation and Insights.*

The standard GPT-4 model, as acknowledged in existing literature, exhibits shortcomings and errors that raise concerns about its reliability and accuracy [21, 29]. In Study 1, we aim to validate these findings with RDs within the specific context of nutrition and the generation of food product explanations. The results of our research in Study 1 confirm previous findings that the outputs often contain nutritional misinformation, lack regulatory health guidelines, omit holistic dietary information, and fail to provide relevant nutrition education. These findings are particularly concerning when considering the potential consequences of presenting such outputs to vulnerable populations, as it could lead to misinformation and nutritional risks. When we engaged dietitians in the validation process of the outputs, they

[1] The instructions to replicate this custom GPT can be found at https://github.com/ND-HCI/The-Food-Product-Nutrition-Assistant.git.

Original (Level 1 Specificity) Output:

Campbell's Chunky Savory Vegetable Soup, available in an 18.8 oz can, provides a mix of vegetables that offer essential vitamins and minerals but can be high in sodium, making it less suitable for a low-sodium diet. The typical canned soup may contain allergens such as wheat and soy, posing risks to those with relevant allergies. It may not align with vegan or gluten-free diets due to potential animal products or gluten-containing ingredients. The shelf life is typically longer for canned products, which, coupled with its relatively low cost, makes it an appealing option for budget-conscious shoppers. However, those prioritizing health may find alternatives like Amy's Organic Low Sodium Vegetable Soup or Progresso's Reduced Sodium Garden Vegetable Soup to be more in line with specific dietary needs or preferences. Always check the specific product label for the most accurate information on ingredients and nutrition.

Refined Prototype Output:

Product Rating: Suitable for your diet because it is low in added sugars and lactose-free, but consider its high sodium content.

Nutritional Analysis:

- High in sodium with 770mg per serving, which exceeds the low sodium threshold.
- Contains 3g of dietary fiber per serving, aligning with your goal to increase fiber intake.
- Low in sugars, with less than 1g of added sugars per serving, suitable for managing blood sugar levels.
- Offers 3g of protein per serving, which can contribute to your goal of incorporating lean proteins.

Health Benefits:

- Good source of dietary fiber from vegetables, which is beneficial for digestive health and blood sugar control.
- Contains vegetables like potatoes, carrots, and green beans, contributing to a varied intake of vitamins and minerals.
- Low in saturated fat and cholesterol, supporting heart health and weight management goals.

Considerations:

- High sodium content may not be ideal for heart health and weight management.
- Contains wheat and soy, which may be relevant for those with allergies to these ingredients.

Dietary Guidelines and MyPlate:

- Aligns with MyPlate guidelines by including a variety of vegetables.
- However, its high sodium content does not fully support the Dietary Guidelines for Americans, which recommend choosing foods with less sodium.

Dietary Restrictions:

- Lactose-free, fitting your need to avoid lactose-containing products.
- Contains wheat and soy, to consider if you have allergies to these ingredients.

Healthier Product Substitutes:

- For a lower sodium option, consider brands like Amy's Light in Sodium or Progresso's Reduced Sodium range.
- A homemade vegetable soup using fresh or frozen vegetables can be a healthier alternative with controlled sodium levels.
- Store brands (like Walmart or Target brand) may offer similar soups with reduced sodium content.

Figure 3: On left: The original GPT output from Study 1 (Level 1 Specificity). On right: The refined GPT prototype output from Study 2. Both outputs are for the Campbell's Soup food product.

were instrumental in pinpointing these errors and enabling us to develop specific design guidelines for prompt instructions aimed at mitigating these issues.

5.1.2 Interactive Feedback and Joint Problem Solving. One of the most effective strategies we employed was real-time iterative development with RDs. This iterative method has been shown previously to work when developing prompts as it provides users with an idea of how varied instructions can improve the model [41]. After each adjustment to the GPT prototype based on feedback, the changes to the output were immediately reviewed by the RDs and further refined. This more rapid prototyping allowed for the dietitians to understand how the changes to the instruction would affect the model's output. In addition, it allowed for a dynamic development process, quickly incorporating their expert insights and then observing the results.

RDs played a pivotal role in error identification and rectification of the model's language that otherwise may have remained undetected. They had direct involvement in providing example sentences and refining them to guide the model in presenting nutrition information in a manner that would meet dietary standards. This collaboration led to significant improvements in the model's ability to provide reliable dietary guidelines, understand the nuances of nutrient levels, as well as provide implications of various dietary choices. Also of importance, the RDs assisted in validating and curating sources to be used by GPT to ensure that the model drew upon the most credible and up-to-date nutritional information, which was vital for maintaining the integrity and improving the accuracy of the model's outputs.

5.2 Customized GPTs for Food Product Explanations

Our research has shown that customizing GPT models with instructions refined by dietitians significantly improves the alignment of the model's outputs to be within dietitian standards. As seen in other work on prompt engineering, instructing prompts with more context produces more efficient and informative outputs that avoid hallucinations [4, 41, 65]. In this section, we discuss some of the improvements with the refined prototype output compared to Study 1.

5.2.1 Personalization of Nutrition and Dietary Information.

As we found in KF1 in Study 1, the dietitians preferred outputs that include nutrition information provided in the prompt. Our study has shown that GPT may not have access to specific up-to-date nutrition information on the product, such as gram/milligram amounts or daily value percentages, without the input of the Nutrition Facts label. In the focus group study, we saw that the addition of nutrition information and ingredients list improves the output to meet dietitian standards by providing the information either in the prompt or the knowledge section of the customized GPT.

Our refined instructions have also been shown to influence the behavior of GPT when the prompt lacks nutritional information. For instance, in response to a query such as *"Is Kellogg's Raisin Bran healthy for me?"*, when the Nutrition Facts label was absent, the ChatGPT prototype requested additional nutritional information and dietary goals to provide a more customized output. The response stems from the explicit directives given to the GPT model to evaluate both the product's nutritional qualities and the user's dietary objectives and has led to explanations with more relevant content, contextual awareness, and greater satisfaction among dietitians. These clarification requests are beneficial to include when building customized GPTs for nutritional information as it aims to have the most information necessary to provide an explanation to the user.

In our focus groups, dietitians highlighted that the GPT was able to correctly align nutrition information to users' goals through the headings of "Health Benefits" and "Considerations." Dietitians noted that this was done more consistently than in the previous outputs due to the use of example sentences to guide the language.

5.2.2 Knowledge Integration.

We learned through our validation in KF2 in Study 1 that the reliability of the GPT outputs increases when available knowledge sources, such as MyPlate guidelines and dietary guidelines, are provided and instructed to be used in the prompt. The resulting outputs produced are not only better aligned with dietitian standards but are also contextualized within the framework of a daily diet. It also brought about an increased consistency among outputs that is practical and more evidence based, which raised trust among the dietitians in our study. As an example, when dietitians recommended the uploading of a document of nutrient claims describing guidelines for the labeling of products as "fat-free", "low-sugar", etc. into the prompt, it was shown to be highly beneficial in the classification of food products. In addition, to test the knowledge section of the customized GPT, a renal dietitian in one focus group suggested that we upload guidelines on fluid intake to determine the model's ability to incorporate this into the product explanation for a renal patient. We found that when we prompted the GPT prototype with specific dietary goals of a renal patient, the model was able to reference that knowledge in the output when crafting the personalized response. This shows that the ability to store knowledge in a customized GPT has the potential to make a difference in output quality and tailor the output to individual dietary needs.

5.2.3 Mitigating Errors in Irrelevant Information, Buzzwords, and Falsehoods .

The large number of errors initially observed in our study (KF3) is mitigated upon the incorporation of example sentences into the prompt by bringing an enhanced specificity of nutritional analysis in GPT outputs. These example sentences appear successful in enabling the outputs to better quantify nutrients and explain their alignment with dietary goals. When template instructions provide a structured framework for the outputs in the prompt, there is a specific clarity of intent to provide context and reduce the likelihood of misinterpretation. The less ambiguous the prompt, the less inaccurate or nonsensical the output, with the example sentences acting as training signals for correctness. The example sentences also guide the model on language and style by explicitly listing buzzwords to avoid and instructing the model to mimic a dietitian's approach. Furthermore, by using example sentences to direct the model to suggest healthier product alternatives, we mitigate the risk of misleading food substitutions by limiting the suggestion of organic foods and offering users a broader range of generic choices to facilitate informed dietary decisions. While this may not guarantee error-free outputs, dietitians observed an improved comprehensive and educational understanding by the model, which had been shown as problems in KF4 and KF5. The example sentences have shown to bring a reduced likelihood of generating implausible responses. Continuous testing is required to fully assess error mitigation. However, the use of example sentences was noted to significantly enhance the comprehensibility of outputs and bring better responses with fewer errors.

When exploring the personalization of the output to dietary goals, we observed that the customized model excelled in tailoring information for individuals and achieved this due to clear instructions and example sentences provided. Dietitians particularly appreciated how the outputs distinctly highlighted allergens and ingredients pertinent to users' dietary requirements. Moreover, when catering to specific needs such as gluten-free or plant-based diets, the model adeptly suggested healthier product alternatives. This enhanced relevance in the outputs can be attributed to the incorporation of detailed instructions into the model, which successfully make it more adept at addressing diverse dietary preferences and needs.

5.2.4 Consistency and Format.

When testing different product outputs with dietitians, the prototype consistently follows the newly refined format and style instructions that have been implemented into the prompt. This has brought an improved consistency of outputs, such that only the content of the output will vary based on the product information and its relevance to a user's specific dietary goals, whereas the organization of the information will remain the same as instructed. The contrast is apparent when the prototype outputs are compared to those in Study 1; the Level 1 – 3 outputs were in a narrative style which dietitians found to be unclear and difficult to interpret (as a concern raised in KF4) relative to the prototype outputs, which employ structured formats with clear headings and bullet points. The dietitians agree that the new prototype structure enhances the clarity of the information, and from their experience they believe it will aid in the overall comprehension of the food product explanation.

6 LIMITATIONS AND FUTURE WORK

Our study, while instrumental in investigating the feasibility of employing LLMs in food product recommendation domains, exhibits several limitations that warrant acknowledgment. The scope of our analysis was confined to a restricted set of food products, and the applicability of the findings might be limited considering the diverse range of food products available in the market. More work is needed to validate our findings with culturally diverse food products. Moreover, our analysis predominantly focused on the dietary needs and preferences of mock individuals, potentially not considering all the variable dietary requirements and goals that can influence the output. The complex dynamics of various dietary plans and goals necessitate a more extensive analysis to understand fully how LLMs can adapt and tailor information to cater to a broader and more diverse array of dietary needs and preferences.

In addition, we recognize the limitation of replicability of our findings due to potential changes in OpenAI's models. Our study's reliance on GPT-4, a dynamic and evolving model, introduces challenges in reproducing our specific results over time. For the first study, our results were generated in August 2023. For the second study, our results were generated in November 2023. Updates and modifications to the model by OpenAI may result in variations in the outcomes if the study were to be replicated.

Our investigation is a first step in validating food product explanations generated by LLMs and has unveiled numerous avenues for prospective research. Future work should consider updating evaluated templates as new dietitian resources and knowledge become available. Additionally, the use of GPT in real user environments should be tested to assess its practical utility and user experience. As our work primarily explores food product explanations, further research focusing on creating visualizations or infographics using GPTs could enhance the interpretability and engagement of the information provided. In addition, more work is needed to update LLMs to produce real-time food cost information regarding whether products are SNAP/WIC eligible. This will provide users with information to make food choices that are both nutritious and affordable. All of these approaches could significantly improve the way nutritional information is communicated by making it more accessible and impactful to a wider audience.

7 CONCLUSION

This study expanded upon the use of LLMs for creating food explanations, aimed at enhancing consumer decision-making in food selection. Collaborating with twelve registered dietitians, we evaluated the face validity of LLM outputs, particularly focusing on GPT-4's performance in providing nutritionally accurate and personalized information. Through assessing responses to varying levels of input specificity for five common food products, we gained insights into how the quality and specificity of information provided to the model influences its response accuracy.

Our research emphasizes the need for further advancements in LLM design, with the goal of making them dependable tools for consumers. We observed consistent patterns of strengths and weaknesses in the generated outputs, leading to the formulation of design guidelines for template instructions which were then implemented in a customized GPT prototype and evaluated in focus group discussions with dietitians. These guidelines are aimed at addressing the identified challenges and enhancing the LLM's utility in the nutrition and dietetics fields.

The refinement of *The Food Product Nutrition Assistant*, a GPT prototype based on these guidelines, exemplifies a practical application of our findings. This study not only highlights the potential of expert-informed LLM customization but also marks a significant step toward more effective utilization of LLMs in providing accurate and reliable food-related information to ultimately guide consumers to make safer and more informed food choices.

ACKNOWLEDGMENTS

This work was supported by the Agriculture and Food Research Initiative grant no. 2021-67022-33447/project accession no. 1024822 from the USDA National Institute of Food and Agriculture.

REFERENCES

[1] Mariette Abrahams, Lynn J Frewer, Eleanor Bryant, and Barbara Stewart-Knox. 2019. Personalised nutrition technologies and innovations: A cross-national survey of registered dietitians. *Public Health Genomics* 22, 3-4 (2019), 119–131. https://doi.org/10.1159/000502915

[2] M Abrahams and NV Matusheski. 2020. Personalised nutrition technologies: a new paradigm for dietetic practice and training in a digital transformation era. *Journal of Human Nutrition and Dietetics* 33, 3 (2020), 295. https://doi.org/10.1111/jhn.12746

[3] Sean H Adams, Joshua C Anthony, Ricardo Carvajal, Lee Chae, Chor San H Khoo, Marie E Latulippe, Nathan V Matusheski, Holly L McClung, Mary Rozga, Christopher H Schmid, et al. 2020. Perspective: guiding principles for the implementation of personalized nutrition approaches that benefit health and function. *Advances in nutrition* 11, 1 (2020), 25–34. https://doi.org/10.1093/advances/nmz086

[4] Simon Arvidsson and Johan Axell. 2023. Prompt engineering guidelines for LLMs in Requirements Engineering.

[5] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* 183, 6 (2023), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

[6] Petjon Ballco and Azucena Gracia. 2022. Tackling nutritional and health claims to disentangle their effects on consumer food choices and behaviour: A systematic review. *Food Quality and Preference* 101 (2022), 104634. https://doi.org/10.1016/j.foodqual.2022.104634

[7] Jacquelyn R Bedsaul-Fryer, Kesso G van Zutphen-Küffer, Jimena Monroy-Gomez, Diane E Clayton, Breda Gavin-Smith, Céline Worth, Christian Nils Schwab, Mathilda Freymond, Anna Surowska, Laís Bhering Martins, et al. 2023. Precision Nutrition Opportunities to Help Mitigate Nutrition and Health Challenges in Low-and Middle-Income Countries: An Expert Opinion Survey. *Nutrients* 15, 14 (2023), 3247. https://doi.org/10.3390/nu15143247

[8] W Marty Blom, Liselotte M van Dijk, Anouska Michelsen-Huisman, Geert F Houben, André C Knulst, Yvette FM Linders, Kitty CM Verhoeckx, Bregje C Holleman, and Leo R Lentz. 2021. Allergen labelling: current practice and improvement from a communication perspective. *Clinical & Experimental Allergy* 51, 4 (2021), 574–584. https://doi.org/10.1111/cea.13830

[9] Marcela Bomfim, Erin Wong, Paige Liang, and James Wallace. 2023. Design and Evaluation of Technologies for Informed Food Choices. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 54 (sep 2023), 46 pages. https://doi.org/10.1145/3565482

[10] Marcela C. C. Bomfim, Sharon I. Kirkpatrick, Lennart E. Nacke, and James R. Wallace. 2020. Food Literacy while Shopping: Motivating Informed Food Purchasing Behaviour with a Situated Gameful App. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376801

[11] Centers for Disease Control and Prevention. 2023. Poor Nutrition. https://www.cdc.gov/chronicdisease/resources/publications/factsheets/nutrition.htm.

[12] Angeline Chatelan, Aurélien Clerc, and Pierre-Alexandre Fonta. 2023. ChatGPT and Future Artificial Intelligence Chatbots: What may be the Influence on Credentialed Nutrition and Dietetics Practitioners? *Journal of the Academy of Nutrition and Dietetics* 123 (2023), 1525–2529. Issue 11. https://doi.org/10.1016/j.jand.2023.08.001

[13] Kathryn Cooper, William Gasper, Ricky Flores, Martina Clarke, Erin Bass, Leslie Evans, and Jana Ponce. 2022. Exploring the Readability of Ingredients Lists of

Food Labels with Existing Metrics. *AMIA Joint Summits on Translational Science* 2022 (2022), 159–167.

[14] Tracy Cullen, Janelle Hatch, Wanda Martin, Joan Wharf Higgins, and Rosanna Sheppard. 2015. Food literacy: definition and framework for action. *Canadian Journal of Dietetic Practice and Research* 76, 3 (2015), 140–145. https://doi.org/10.3148/cjdpr-2015-010

[15] Nienke M De Vlieger, Clare Collins, and Tamara Bucher. 2017. What is a nutritious snack? Level of processing and macronutrient content influences young adults' perceptions. *Appetite* 114 (2017), 55–63. https://doi.org/10.1016/j.appet.2017.03.021

[16] Tawanna R Dillahunt, Michelle Sawwan, Danielle Wood, Brianna L Wimer, Ann-Marie Conrado, Heather Eicher-Miller, Alisa Zornig Gura, and Ronald Metoyer. 2023. Understanding Food Planning Strategies of Food Insecure Populations: Implications for Food-Agentic Technologies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 239, 22 pages. https://doi.org/10.1145/3544548.3581568

[17] Mauro Dragoni, Ivan Donadello, and Claudio Eccher. 2020. Explainable AI meets persuasiveness: Translating reasoning results into behavioral change advice. *Artificial Intelligence in Medicine* 105 (2020), 101840. https://doi.org/10.1016/j.artmed.2020.101840

[18] Manon Egnell, Zenobia Talati, Serge Hercberg, Simone Pettigrew, and Chantal Julia. 2018. Objective understanding of front-of-package nutrition labels: an international comparative experimental study across 12 countries. *Nutrients* 10, 10 (2018), 1542. https://doi.org/10.3390/nu10101542

[19] Heather A Eicher-Miller, Lisa Graves, Bethany McGowan, Barbara J Mayfield, Blake A Connolly, Wanda Stevens, and Angela Abbott. 2023. A Scoping Review of Household Factors Contributing to Dietary Quality and Food Security in Low Income Households with School-Age Children in the US. *Advances in Nutrition* 14, 4 (2023), 914–945. https://doi.org/10.1016/j.advnut.2023.05.006

[20] MA El-Dosuky, Magdi Zakria Rashad, TT Hamza, and AH El-Bassiouny. 2012. Food recommendation using ontology and heuristics. In *Advanced Machine Learning Technologies and Applications*, Vol. 322. Springer, Berlin, Heidelberg, 423–429. https://doi.org/10.1007/978-3-642-35326-0_42

[21] Ali Francis. 2023. Please don't ask CHATGPT for Diet Advice. https://www.bonappetit.com/story/chatgpt-diet-advice-meal-plans

[22] Andrea Frosolini, Paolo Gennaro, Flavia Cascino, and Guido Gabriele. 2023. In Reference to "Role of Chat GPT in Public Health", to Highlight the AI's Incorrect Reference Generation. *Annals of Biomedical Engineering* 51 (2023), 2120–2122. Issue 10. https://doi.org/10.1007/s10439-023-03248-4

[23] Manuel B Garcia. 2023. ChatGPT as a Virtual Dietitian: Exploring Its Potential as a Tool for Improving Nutrition Knowledge. *Applied System Innovation* 6, 5 (2023), 96. https://doi.org/10.3390/asi6050096

[24] Joe Germino, Annalisa Szymanski, Ronald Metoyer, and Nitesh V Chawla. 2023. A community focused approach toward making healthy and affordable daily diet recommendations. *Frontiers in Big Data* 6 (2023). https://doi.org/10.3389/fdata.2023.1086212

[25] Aidan Gilson, Conrad W Safranek, Thomas Huang, Vimig Socrates, Ling Chi, Richard Andrew Taylor, David Chartash, et al. 2023. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education* 9, 1 (2023), e45312. https://doi.org/10.2196/45312

[26] Vagelis Hristidis, Nicole Ruggiano, Ellen L Brown, Sai Rithesh Reddy Ganta, and Selena Stewart. 2023. ChatGPT vs Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results. *Journal of Medical Internet Research* 25 (2023), e48966. https://doi.org/10.2196/48966

[27] Nina Søndergaard Jakobsen, Lisbeth Kaufmann, Yvonne Hennesser, and Søren Tange Kristensen. 2017. Narrative dietary counseling: a new approach for the dietitian that strengthens the relationship and collaboration with clients. *Topics in Clinical Nutrition* 32, 3 (2017), 229–242. https://doi.org/10.1097/TIN.0000000000000111

[28] Urszula Jessen, Michal Sroka, and Dirk Fahland. 2023. Chit-Chat or Deep Talk: Prompt Engineering for Process Mining. arXiv Preprint arXiv:2307.09909. https://doi.org/10.48550/arXiv.2307.09909

[29] Douglas Johnson, Rachel Goodman, J Patrinely, Cosby Stone, Eli Zimmerman, Rebecca Donald, Sam Chang, Sean Berkowitz, Avni Finn, Eiman Jahangir, et al. 2023. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Research Square. https://doi.org/10.21203/rs.3.rs-2566942/v1

[30] Skyler B Johnson, Andy J King, Echo L Warner, Sanjay Aneja, Benjamin H Kann, and Carma L Bylund. 2023. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI cancer spectrum* 7, 2 (2023), pkad015. https://doi.org/10.1093/jncics/pkad015

[31] Daniel Kirk, Cagatay Catal, and Bedir Tekinerdogan. 2021. Precision nutrition: A systematic literature review. *Computers in Biology and Medicine* 133 (2021), 104365.

[32] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems* 35 (2022), 22199–22213. https://doi.org/10.48550/arXiv.2205.11916

[33] Anis Koubaa. 2023. GPT-4 vs. GPT-3.5: A Concise Showdown. TechRxiv Preprint. https://doi.org/10.36227/techrxiv.22312330.v2

[34] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* 2, 2 (2023), e0000198.

[35] Eva Adomaa Kyereboah. 2018. *Influence of metabolic syndrome information on macronutrient consumption decisions*. Ph. D. Dissertation. Kansas State University.

[36] Sachi Lele and Kathryn Cooper. 2021. Pilot Research on Improving Consumer Health Literacy in Food Labels by Measuring the Prevalence & Occurrence of Similar Ingredients. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Institute of Electrical and Electronics Engineers Inc., Houston, TX, USA, 1664–1666. https://doi.org/10.1109/BIBM52615.2021.9669591

[37] Weixin Liang, Girmaw Abebe Tadesse, Daniel Ho, L Fei-Fei, Matei Zaharia, Ce Zhang, and James Zou. 2022. Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence* 4, 8 (2022), 669–677. https://doi.org/10.1038/s42256-022-00516-1

[38] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35. https://doi.org/10.1145/3560815

[39] Katherine M Livingstone, Omar Ramos-Lopez, Louis Perusse, Hisanori Kato, Jose M Ordovas, and J Alfredo Martínez. 2022. Precision nutrition: A review of current approaches and future endeavors. *Trends in Food Science & Technology* 128 (2022), 253–264. https://doi.org/10.1016/j.tifs.2022.08.017

[40] Jordi Merino. 2022. Precision nutrition in diabetes: when population-based dietary advice gets personal. *Diabetologia* 65, 11 (2022), 1839–1848. https://doi.org/10.1007/s00125-022-05721-6

[41] Bertalan Meskó. 2023. Prompt engineering as an important emerging skill for medical professionals: tutorial. *Journal of Medical Internet Research* 25 (2023), e50638. https://doi.org/10.2196/50638

[42] Elliot G Mitchell, Rosa Maimone, Andrea Cassells, Jonathan N Tobin, Patricia Davidson, Arlene M Smaldone, and Lena Mamykina. 2021. Automated vs. human health coaching: exploring participant and practitioner experiences. *Proceedings of the ACM on human-computer interaction* 5, CSCW1 (2021), 1–37. https://doi.org/10.1145/3449173

[43] Tossawat Mokdara, Priyakorn Pusawiro, and Jaturon Harnsomburana. 2018. Personalized food recommendation using deep neural network. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*. IEEE, Nakhonpathom, Thailand, 1–4. https://doi.org/10.1109/ICT-ISPC.2018.8523950

[44] Sally G Moore, Judy K Donnelly, Steve Jones, and Janet E Cade. 2018. Effect of educational interventions on understanding and use of nutrition labels: A systematic review. *Nutrients* 10, 10 (2018), 1432. https://doi.org/10.3390/nu10101432

[45] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: Generative AI and HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 110, 7 pages. https://doi.org/10.1145/3491101.3503719

[46] Paweł Niszczota and Iga Rybicka. 2023. The credibility of dietary advice formulated by ChatGPT: robo-diets for people with food allergies. *Nutrition* 112 (2023), 112076. https://doi.org/10.1016/j.nut.2023.112076

[47] US Department of Agriculture. 2017. MyPlate.

[48] OpenAI. 2023. https://openai.com/blog/introducing-gpts

[49] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744. https://doi.org/10.48550/arXiv.2203.02155

[50] Ishita Padhiar, Oshani Seneviratne, Shruthi Chari, Dan Gruen, and Deborah L McGuinness. 2021. Semantic modeling for food recommendation explanations. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, Nakhonpathom, Thailand, 13–19. https://doi.org/10.1109/ICT-ISPC.2018.8523950

[51] Sibo Pan, Xipei Ren, Steven Vos, and Aarnout Brombacher. 2022. Digital Tools to Promote Healthy Eating for Working-Age Individuals: A Scoping Review. In *The Ninth International Symposium of Chinese CHI* (Online, Hong Kong) *(Chinese CHI 2021)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3490355.3490356

[52] Elsie Azevedo Perry, Heather Thomas, H Ruby Samra, Shannon Edmonstone, Lyndsay Davidson, Amy Faulkner, Lisa Petermann, Elizabeth Manafò, and Sharon I Kirkpatrick. 2017. Identifying attributes of food literacy: a scoping review. *Public health nutrition* 20, 13 (2017), 2406–2415. https://doi.org/10.1017/S1368980017001276

[53] Denis Peskoff and Brandon Stewart. 2023. Credible without Credit: Domain Experts Assess Generative Language Models. In *Proceedings of the 61st Annual*

Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 427–438. https://doi.org/10.18653/v1/2023.acl-short.37
[54] Dacinia Crina Petrescu, Iris Vermeir, and Ruxandra Malina Petrescu-Mag. 2020. Consumer understanding of food quality, healthiness, and environmental impact: A cross-national perspective. *International journal of environmental research and public health* 17, 1 (2020), 169. https://doi.org/10.3390/ijerph17010169
[55] Brigitta Plasek, Zoltán Lakner, and Ágoston Temesi. 2020. Factors that influence the perceived healthiness of food. *Nutrients* 12, 6 (2020), 1881. https://doi.org/10.3390/nu12061881
[56] Maren Podszun and Bettina Hieronimus. 2023. Can ChatGPT generate energy, macro-and micro-nutrient sufficient meal plans for different dietary patterns? https://doi.org/10.21203/rs.3.rs-2858319/v1. https://doi.org/10.21203/rs.3.rs-2858319/v1 Preprint.
[57] Christina A Roberto, Shu Wen Ng, Montserrat Ganderats-Fuentes, David Hammond, Simon Barquera, Alejandra Jauregui, and Lindsey Smith Taillie. 2021. The influence of front-of-package nutrition labeling on consumer behavior and product reformulation. *Annual review of nutrition* 41 (2021), 529–550. https://doi.org/10.1146/annurev-nutr-111120-094932
[58] Jamil S Samaan, Yee Hui Yeo, Nithya Rajeev, Lauren Hawley, Stuart Abel, Wee Han Ng, Nitin Srinivasan, Justin Park, Miguel Burch, Rabindra Watson, et al. 2023. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. *Obesity surgery* 33 (2023), 1790–1796. Issue 6. https://doi.org/10.1007/s11695-023-06603-5
[59] Albrecht Schmidt, Fosca Giannotti, Wendy Mackay, Ben Shneiderman, and Kaisa Väänänen. 2021. Artificial intelligence for humankind: a panel on how to create truly interactive and Human-Centered AI for the benefit of individuals and Society. In *IFIP Conference on Human-Computer Interaction*, Vol. 12936. Springer, Bari, Italy, 335–339. https://doi.org/10.1007/978-3-030-85607-6_32
[60] Siyi Shangguan, Ashkan Afshin, Masha Shulkin, Wenjie Ma, Daniel Marsden, Jessica Smith, Michael Saheb-Kashaf, Peilin Shi, Renata Micha, Fumiaki Imamura, et al. 2019. A meta-analysis of food labeling effects on consumer diet behaviors and industry practices. *American journal of preventive medicine* 56, 2 (2019), 300–314. https://doi.org/10.1016/j.amepre.2018.09.024
[61] Deborah Leachman Slawson, Nurgul Fitzgerald, and Kathleen T Morgan. 2013. Position of the Academy of Nutrition and Dietetics: the role of nutrition in health promotion and chronic disease prevention. *Journal of the Academy of Nutrition and Dietetics* 113, 7 (2013), 972–979. https://doi.org/10.1016/j.jand.2013.05.005
[62] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. 2022. Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239. https://doi.org/10.48550/arXiv.2201.08239
[63] Live Edvardsen Tonheim, Elisabeth Austad, Liv Elin Torheim, and Sigrun Henjum. 2022. Plant-based meat and dairy substitutes on the Norwegian market: comparing macronutrient content in substitutes with equivalent meat and dairy products. *Journal of Nutritional Science* 11 (2022), e9. https://doi.org/10.1017/jns.2022.6
[64] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. 2018. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems* 50 (2018), 501–526. https://doi.org/10.1007/s10844-017-0469-0
[65] Chun Hua Tsai, Sathvik Kadire, Tejesvi Sreeramdas, Matthew VanOrmer, Melissa Thoene, Corrine Hanson, Ann Anderson Berry, and Deepak Khazanchi. 2023. Generating Personalized Pregnancy Nutrition Recommendations with GPT-Powered AI Chatbot. *Proceedings of the 20th International ISCRAM Conference* 2023-text (2023), 263–271.
[66] Mayumi Ueda, Mari Takahata, and Shinsuke Nakajima. 2011. User's food preference extraction for personalized cooking recipe recommendation. In *Proceedings of the Second International Conference on Semantic Personalized Information Management: Retrieval and Recommendation - Volume 781* (Bonn, Germany) *(SPIM'11)*. CEUR-WS.org, Aachen, DEU, 98–105.
[67] Marianna S Wetherill, Kayla Castleberry White, and Christine Rivera. 2018. Food insecurity and the nutrition care process: practical applications for dietetics practitioners. *Journal of the Academy of Nutrition and Dietetics* 118, 12 (2018), 2223. https://doi.org/10.1016/j.jand.2017.08.114
[68] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv preprint arXiv:2302.11382* abs/2302.11382 (2023). https://doi.org/10.48550/arXiv.2302.11382
[69] Reihaneh Zeinalian, Shonaz Ahmadikhatir, Ensieh Nasli Esfahani, Nazli Namazi, and Bagher Larijani. 2022. The roles of personalized nutrition in obesity and diabetes management: a review. *Journal of Diabetes & Metabolic Disorders* 21, 1 (2022), 1119–1127. https://doi.org/10.1007/s40200-022-01016-0

8 APPENDICES

A STUDY 1: FOOD PRODUCT DESCRIPTIONS & EXPLANATION OUTPUTS

My Plate Group	Product
Dairy	Daisy Low Fat 2% Cottage Cheese, 24 oz container
Fruit	Dole Fruit Bowls Mandarin Oranges in 100% Fruit Juice, 4 oz containers
Grain	Kellogg's Raisin Bran Original Breakfast Cereal, 16.6 oz box
Protein	Bubba Burger Choice Beef Chuck Burgers Original 1/3 lb each - 6 count Frozen
Vegetable	Campbell's Chunky Savory Vegetable Soup,18.8 oz can

Table 2: The MyPlate Food Category and corresponding food products used in generating GPT outputs.

Below are the **outputs** generated by ChatGPT in Study 1 for each of the 5 products across the three levels of specificity.

Product 1: Kellogg's Raisin Bran Original Breakfast Cereal

Level 1:
Kellogg's Raisin Bran Original in a 16.6 oz package is a well-known breakfast cereal. Health-wise, it's rich in fiber from whole-grain wheat and bran, aiding digestion, and it often includes added vitamins and minerals. However, it can contain relatively high amounts of added sugars, potentially detracting from its overall nutritional value. Common allergens in the product include wheat and possibly traces of nuts, which might be problematic for those with related allergies. Regarding dietary restrictions, it is not suitable for gluten-free diets, and its sugar content might not align with low-sugar needs. The low-sodium and low-saturated fat content make it suitable for some heart-healthy diets, but it's not typically vegan due to Vitamin D sourced from lanolin. Comparing shelf life and cost, the product usually has a good shelf life relative to its cost, making it an attractive option for budget-conscious consumers. Healthier substitutes could include cereals from brands like Post (Shredded Wheat) or Nature's Path, offering organic, lower sugar, or gluten-free options, catering to more specific dietary requirements.

Source: Nutritional information derived from the product label and general knowledge on allergens and dietary restrictions applicable to common cereal ingredients. Specifics may vary depending on the exact product variant and region. Always refer to the product packaging for the most accurate information.

Level 2:
Kellogg's Raisin Bran Original Breakfast Cereal, with its whole grain wheat and bran content, contributes to dietary fiber (26% DV), supporting digestive health. However, its 17g total sugars (including 9g added sugars) per serving may be considered high for those watching sugar intake. The product contains wheat, a known allergen, posing risks for those with wheat allergies or celiac disease; thus, it's not suitable for gluten-free diets. It aligns with low-saturated fat and low-cholesterol diets but isn't ideal for low-sodium or low-sugar diets. The cost compared to the shelf life may be reasonable, as cereals typically have a long shelf life, but this depends on individual price points in stores. Healthier alternatives could include bran cereals from brands like Post or General Mills that have reduced sugar or certified gluten-free options. In summary, while providing essential nutrients, its high sugar content and presence of wheat could be a concern for specific dietary needs and preferences.

Source: Kellogg's Nutrition Facts and Ingredients Label, as provided in the question.

Level 3:
Kellogg's Raisin Bran Original Breakfast Cereal, containing whole grain wheat and wheat bran, offers 7g of dietary fiber per serving, aligning with the individual's goal to increase fiber-rich foods (26% DV). However, it's high in total sugars (17g) with 9g of added sugars, which conflicts with the aim to reduce added sugars for blood sugar management. Containing no lactose, it suits those with lactose intolerance. The cereal is not ideal for a low-sodium diet (200mg, 9% DV), and its 1g total fat is low, making it suitable for a heart-healthy regimen. As for dietary restrictions, it's not gluten-free or suitable for those looking to avoid wheat. The product's shelf life and cost are typical for breakfast cereals but could be outweighed by its sugar content for those managing blood sugar levels. Healthier alternatives might include whole-grain cereals with reduced sugar from brands like Post or General Mills. Overall, while offering some benefits like fiber, its high sugar content makes it less suitable for individuals seeking to manage blood sugar or weight.

Reference: Kellogg's product nutritional information from the 16.6 oz package, available at most supermarkets or online at the manufacturer's official website.

Product 2: Daisy Low Fat 2% Cottage Cheese

Level 1:
Daisy Low Fat 2% Cottage Cheese, available in an 18.8 oz container, is generally a nutritious choice. Health Benefits include being a source of protein, calcium, and vitamins B12 and D, supporting muscle growth and bone health. However, the sodium content could be a concern for those watching their salt intake. The common allergens in this product are dairy, specifically lactose, which may pose risks to lactose-intolerant individuals. It aligns with low-sugar and low-saturated fat diets but is not suitable for vegan or certain gluten-free diets due to the potential use of stabilizers. The shelf life of cottage cheese is typically shorter compared to other cheeses, and while it's generally affordable, the relatively brief shelf life might influence purchasing decisions for those looking to stock up. Healthier substitutes might include lactose-free or low-sodium cottage cheese options from brands like Lactaid or Horizon Organic.

Reference: Daisy's official website and nutrition labels provide information on nutritional components, potential allergens, and shelf life, and specific substitutes were identified through general market knowledge and brand websites.

Level 2:
Daisy Low Fat 2% Cottage Cheese, with its simple ingredients of Cultured Skim Milk, Cream, and Salt, is a protein-rich option, providing 13g per serving (26% DV). It is suitable for vegetarians but not for vegans due to dairy content. The 2.5g of total fat, including 1.5g of saturated fat, keeps it within low-fat dietary preferences. It may not be ideal for low-sodium diets with 350mg per serving (15% DV), but it's free from gluten and added sugars. Individuals with lactose intolerance or milk allergies should avoid this product due to the presence of milk derivatives. The shelf life of cottage cheese is typically around 2 weeks, and the cost is generally considered reasonable for a dairy product. However, its short shelf life may influence purchasing decisions if not consumed promptly. For those looking for dairy-free alternatives, soy or almond-based cottage cheese products can provide similar nutritional benefits without the presence of lactose or cholesterol. These considerations align Daisy's cottage cheese as a generally good choice for an average consumer, but specific needs may prompt consideration of alternative products.

Source: Nutrition facts and ingredient information are based on the provided label. Further dietary considerations are general knowledge based on standard nutritional guidelines and preferences.

Level 3:
Daisy Low Fat 2% Cottage Cheese offers some health benefits, with 90 calories and 13g of protein per serving, aligning well with the need for lean proteins and weight loss goals. The low saturated fat content is good for heart health. However, its lactose content conflicts with lactose intolerance, and the sodium content (350mg) might not suit a low-sodium diet. Lack of fiber doesn't support the goal to increase fiber-rich food intake. For those with lactose intolerance, lactose-free cottage cheese brands like Green Valley Creamery might be a better alternative. Comparing shelf life and cost, cottage cheese generally has a shorter shelf life (around 10-14 days unopened), so considering cost and individual consumption patterns is important. Despite certain benefits, the lactose content and lack of fiber make this product unsuitable for the given dietary needs.

Reference: Daisy Brand. (n.d.). 2% Low Fat Cottage Cheese - 18.8 oz. Retrieved from Daisy's official website; Green Valley Creamery. (n.d.). Lactose-Free Dairy Products. Retrieved from Green Valley Creamery's official website.

Product 3: Bubba Burger Original 100% USDA Choice Beef Chuck Burgers

Level 1:
Bubba Burger® Original 100% USDA Choice Beef Chuck Burgers, containing 6 - 1/3 pound burgers, offers a source of protein, iron, and other essential nutrients, but can also be high in saturated fat, cholesterol, and sodium, which may be detrimental to heart health in excessive amounts. Allergen information is typically available on the product packaging, but common allergens associated with beef products could include dairy if any additional ingredients are included. Regarding dietary restrictions, this product would not be suitable for vegan, low-sodium, or low-saturated fat diets. The typical shelf life of frozen beef burgers is around 6-8 months, and comparing this with the cost can provide a reasonable value depending on individual consumption patterns. Healthier product substitutes might include lean turkey burgers from brands like Jennie-O or plant-based burgers from Beyond Meat, which may align better with certain dietary preferences or health goals.

Reference: Nutrition information and dietary analysis can typically be found on the product's packaging or the manufacturer's website. Additional information on dietary restrictions and suitable alternatives may be obtained from reputable sources like the USDA Food Database or dietary guidelines provided by health organizations.

Level 2:
The Bubba Burger® Original 100% USDA Choice Beef Chuck Burgers provide a significant source of protein (25g) and iron (15% DV), contributing to muscle growth and energy production. However, they contain high levels of total fat (35g) and saturated fat (15g), accounting for 54% and 75% of the daily value respectively, which may contribute to heart disease risk if consumed excessively. As the only ingredient is 100% USDA Choice Beef Chuck, there are no known allergens unless the consumer has a beef allergy. This product doesn't align with vegan, low-sodium, or low-saturated fat diets, but it is gluten-free and low-sugar. The shelf life of frozen beef burgers can be several months, offering convenience, though the cost might be higher compared to other beef grades. Healthier substitutes might include lean turkey burgers or plant-based burgers from brands like Beyond Meat® or MorningStar Farms® for those seeking lower saturated fat or vegetarian options. Therefore, while these burgers can be part of a balanced diet, they might not be suitable for those with specific dietary needs or health concerns.

Reference: Information obtained from the provided nutrition facts label and general nutritional guidelines. Individual dietary needs and health concerns should be discussed with a healthcare provider.

Level 3:
Bubba Burger® Original 100% USDA Choice Beef Chuck Burgers consist solely of beef chuck, offering a high-protein option but also containing a significant amount of total and saturated fat. Regarding the individual's dietary needs and goals, the product aligns with the requirement to reduce sugars and refined carbs, as it contains no carbohydrates or sugars. However, it falls short in promoting heart health due to its high saturated fat content (15g, 75% DV) and doesn't contribute to the goal of increasing fiber-rich foods. The absence of lactose aligns with the lactose intolerance constraint, but the burger might not be considered a lean protein source compared to fish, chicken, or plant-based proteins. Shelf life vs. cost may vary, but generally, frozen beef patties might offer a reasonable shelf life, possibly impacting the purchasing decision positively. For healthier substitutes, one might consider lean turkey or plant-based burgers from brands like Beyond Meat, providing lower fat content and aligning better with weight loss and heart health goals. Since the ingredients only include beef chuck, common allergens like nuts, dairy, or gluten are not present in this product.

Reference: Bubba Burger® nutritional label and ingredients as provided.

Product 4: Campbell's Chunky Savory Vegetable Soup

Level 1:
Campbell's Chunky Savory Vegetable Soup, available in an 18.8 oz can, provides a mix of vegetables that offer essential vitamins and minerals but can be high in sodium, making it less suitable for a low-sodium diet. The typical canned soup may contain allergens such as wheat and soy, posing risks to those with relevant allergies. It may not align with vegan or gluten-free diets due to potential animal products or gluten-containing ingredients. The shelf life is typically longer for canned products, which, coupled with its relatively low cost, makes it an appealing option for budget-conscious shoppers. However, those prioritizing health may find alternatives like Amy's Organic Low Sodium Vegetable Soup or Progresso's Reduced Sodium Garden Vegetable Soup to be more in line with specific dietary needs or preferences. Always check the specific product label for the most accurate information on ingredients and nutrition.

Reference: Product labels and manufacturer information from Campbell's Soup Company website and comparison with other known brands.

Level 2:
Campbell's Chunky Savory Vegetable Soup offers some nutritional benefits such as low calories and fat, with a healthy dose of fiber and potassium. However, the high sodium content, at 33% DV per serving, detracts from the overall healthfulness, posing risks for those with hypertension or heart disease. The ingredient list identifies wheat and soy as allergens, which can be concerning for those with related allergies. With regards to dietary restrictions, it's not suitable for vegan diets due to hydrolyzed soy protein, and it's not gluten-free due to hydrolyzed wheat gluten. The shelf life of canned soups is typically long, providing a cost-effective option for those on a budget. However, the relatively high sodium could lead consumers to opt for low-sodium alternatives like Amy's Organic Low Sodium Vegetable Soup. Though Campbell's Chunky Savory Vegetable Soup includes wholesome vegetables and can be a part of a balanced diet, the high sodium content and presence of allergens might make it less suitable for specific individuals or those with particular dietary needs.

Reference: Campbell's product details and nutrition facts can be found on the Campbell's official website or on the packaging of the specific product.

Level 3:
Campbell's Chunky Savory Vegetable Soup offers some health benefits with low calories (90 per serving) and minimal fat content (0.5g), which aligns with weight loss and heart health goals. The 4g of dietary fiber per serving supports an increased intake of fiber-rich foods, while the absence of lactose aligns with lactose intolerance considerations. However, the sodium content (770mg, 33% DV) may be a concern for heart health, and the inclusion of hydrolyzed wheat gluten and sugar may not fully support goals to reduce refined carbohydrates and added sugars. The product also contains known allergens: wheat and soy. As for dietary restrictions, it's not suitable for vegan or gluten-free diets. The soup's shelf life is typically long due to canning, making it a convenient option, though the cost varies by location. Healthier substitutes could include low-sodium, gluten-free vegetable soups from brands like Amy's or Progresso.

Source: Campbell's Chunky Savory Vegetable Soup Nutrition Facts and Ingredients Label provided in the question.

Product 5: Dole Fruit Bowls Mandarin Oranges in 100% Fruit Juice

Level 1:
Dole Fruit Bowls Mandarin Oranges in 100% Fruit Juice is generally considered a healthy snack option as it contains vitamin C and fiber, contributing to immune function and digestive health. However, some varieties might contain added sugars or syrups that can detract from its overall nutritional value. There are no specific allergens listed in the product, but cross-contamination with allergens in the manufacturing facility might pose risks to sensitive individuals. The product generally aligns with vegan, gluten-free, and low-sodium diets but may not be suitable for low-sugar diets if it contains added sugars. The shelf life compared to the cost is reasonable, with canned or packaged fruit often having a shelf life of 1-2 years, allowing for bulk purchasing without immediate consumption. Healthier substitutes might include fresh mandarin oranges or brands like Del Monte's No Sugar Added Mandarin Oranges, which may provide less sugar content.

(Source: Based on publicly available nutritional information for Dole Fruit Bowls Mandarin Oranges and general nutritional guidelines, accessible via Dole's official website and various health and nutrition authorities).

Level 2:
Dole Fruit Bowls Mandarin Oranges in 100% Fruit Juice, packaged in 4 oz containers, provide a low-calorie option containing 70 calories per serving, with no fats, cholesterol, or sodium. The health benefits are present through vitamins like Vitamin A (2% DV) and Vitamin C (30% DV), and essential minerals like calcium and potassium. It contains 14g of natural sugars but no added sugars, which is favorable for a balanced diet. However, the high sugar content may not be suitable for low-sugar diets. Since the product contains no known common allergens, it is likely safe for those with typical food allergies. Aligning with vegan, gluten-free, low-sodium, and low-saturated fat diets, it offers versatility for various dietary needs. Comparing shelf life and cost, the product is generally affordable and has a reasonable shelf life, though fresh fruits might be a cost-effective alternative with a shorter shelf life. Healthier substitutes might include fresh mandarin oranges or fruit bowls in water with no added sweeteners.

Reference: The analysis is based on the provided nutrition facts label and common knowledge about nutritional guidelines. Reference: U.S. Food & Drug Administration, "Nutrition Facts Label Programs and Materials," www.fda.gov.

Level 3:
Dole Fruit Bowls Mandarin Oranges in 100% Fruit Juice is a low-calorie option with no added sugars or fats, aligning well with goals to reduce sugars and maintain weight loss. Rich in Vitamin C, it provides antioxidants, although the fiber content (1g) is low compared to whole fruits. There's no lactose or known common allergens, making it suitable for lactose intolerance and most dietary restrictions. This product fits vegan, gluten-free, low-sodium, low-sugar, and low-saturated fat diets. While the shelf life is generally good for canned products, the cost may vary by location, possibly influencing purchasing decisions. Healthier alternatives could include fresh mandarin oranges or brands offering higher fiber content. Considering the individual's dietary needs, this product is mostly aligned but falls short in providing significant fiber or lean protein.

Reference: Dole Food Company Inc., Product Information, "Mandarin Oranges in 100% Fruit Juice," available on the official Dole website and major retailers.

B STUDY 1: KEY FINDINGS AND RELEVANT DIETITIAN QUOTES

Table 3 contains key findings, dietitians quotes, and the corresponding design guideline from Study 1.

C STUDY 2: TEMPLATE INSTRUCTIONS

Initial Template instructions are shown in Figure 4. The refined template instructions after Study 2 are shown in Figure 5.

Key Findings	Example Quotes	Design Guidelines
KF1: The outputs generated by the more detailed specificity prompts are preferred by dietitians		
Specificity Level of Input	*"I liked in the 3rd output...it was pretty clear that [the product] wouldn't be a good idea based on [the individual's] dietary needs." (RD11)*	[DG1]
	"The 2nd [response] now says it is free from gluten but the [first response] said it may not be gluten free. With more information it's able to rule out that the product is not gluten free when it didn't do that just with the 1st response." (RD3)	[DG3]
KF2: The output do not align with the standards upheld by registered dietitians		
Reliability	*"If you go up to the actual package level, its sodium is 85 [mg]. . . it's still a low sodium food technically." (RD1)*	[DG3]
Sources of Information	*"I also felt the wording of the common allergens in this product as dairy, specifically lactose - is not the allergen. So I think that is a fully inaccurate statement." (RD8)*	[DF2]
KF3: There is a prevalence of falsehoods which undermines the clarity and coherence of the explanations and could lead consumers to form incorrect conclusions.		
Imprecise "Buzz Words"	*"I always hesitate with the [word] 'healthy', but I was thinking about it, and again, looking at the population, it's probably, best to use words like 'healthier' just because those are words people know and can associate with." (RD10)*	[DG4]
Prevalent Errors	*"I'm wondering what they're using for a sugar cutoff because nine grams is not high...if you look at the American Heart Association recommendations for added sugar, it's [more] like 26 grams a day for women and 30 something for men." (RD2)*	[DG3], [DG7]
	"This isn't accurate because the salt's not high. . . there's no dairy mixed in. If you add cheese, there would be. This is suitable for a low sodium diet, but not for vegans or low saturated fat...It is a low sodium food. It has 85 milligrams per serving. Low sodium is [up to] 140 milligrams per serving." (RD7)	
Misleading statements	*"Gluten free should not be synonymous with healthier. Now, if we're talking about somebody who needs a gluten free diet, sure, a gluten free cereal is a good alternative, but the majority of people don't need to do that. So I think that's an area where...it loses me on accuracy." (RD8)*	[DG3]
Fictitious Substitutions	*"I don't think [there are] soy and almond based cottage cheeses." (RD7)*	[DG3]
KF4: The model fails to maximize the comprehensiveness and educational value of its output		
Lack of Educational Context	*"15 percent daily value...That's based off of a 2000 calorie diet, which isn't even necessarily what most people need. So this could be too high for them...I feel like that percentage can cause a lot of confusion from what I've seen with my patients in the past." (RD12)*	[DG5], [DG3]
	"Context is always good, right?..this is roughly what the American Heart Association recommends for fat intake per day for a heart healthy diet, then somebody could say, oh, wow, like this 10 grams of fat is a lot of my daily fat and then decide if that food is valuable for them and their diet or not, or if a different alternative needs to be sought out." (RD2)	
Reading Level of Consumer and Health Literacy Support	*"Detracting. What does that mean? I feel like they wouldn't read that word correctly if they didn't have a good understanding of [English], if they're not reading well." (Regarding the LLM Kellogg's explanation Level 1) (RD6)*	[DG6]
	"The language is very technical. So in general, like the reading level, I feel like it would be a little bit advanced and could be confusing for people who might potentially be having food insecurity." (RD11)	
KF5: The outputs outline potential alignments with consumers' immediate dietary needs, yet it does not provide information on broader contextual factors or integrate insights into an individual's comprehensive diet		
Customization and Consumer Health Benefit	*"It's at least giving some ... data and recommendations. I think it's just the dietitian in me likes...more of a context of how it fits in the whole picture for an average consumer." (RD4)*	[DG8]
	"The way I think about it is nobody has cereal without some form of milk. Milk introduces other allergens. You have to think about that. And you would have to counsel somebody separately...if you are also lactose intolerant, you have to consider the milk that you choose to go along with this. And then that adds to the overall nutrition profile dramatically." (RD5)	
	"Sometimes these descriptions are also just a little bit odd. I think [protein] contributing to muscle growth...I would say it helps rebuild the muscles, or make them stronger when paired with exercise, but I don't think I'd necessarily tell them that it's used for energy production or muscle growth." (RD9)	

Table 3: Key Findings and Dietitian Quotes. Includes the 5 Key Findings, aligned themes, and statements from the Registered Dietitians. Accompanying each finding is a relevant design guideline informed by these insights. For clarity, statements from individual registered dietitians are marked as RD1, RD2, etc., indicated by RD#.

Initial GPT Instruction Template

This GPT acts like a registered dietitian. It provides detailed explanations about food products, analyzing their nutritional content and suitability for various dietary needs. It takes into account individual dietary preferences, restrictions, and health goals to offer tailored advice. The GPT avoids giving medical advice or diagnosing health conditions. It encourages users to consult a healthcare professional for personalized medical advice. The GPT aims to educate and inform users about the nutritional aspects of different foods and how they fit into a balanced diet. The GPT will follow the set of instructions below. If there are "example sentences" the GPT will include those in the response when applicable.

The GPT's tone is professional, informative, and empathetic, understanding the personal nature of dietary choices. It personalizes responses to align with the user's specific dietary needs and goals, maintaining a supportive and non-judgmental tone.

In the output, the GPT should limit words like "healthy-dose", or "muscle growth". Remove the mention of the daily value percentage numbers (DV%) and instead reference nutrient amounts.

When providing a response, the GPT should follow this structure. If there are curly brackets {} within the text, replace the content inside these brackets with the appropriate information and provide the sentence structure.

The first sentence should say: "This product is a {good choice/ok choice/bad choice} for your diet because {provide one primary reason why related to the nutrients}."

> For example: "Product is an ok choice for your diet because of its high fiber content, but it has a notable amount of added sugars."

In addition, the GPT needs to mention these aspects:

Nutritional Analysis: Identify the nutritional components and explain how they contribute to or detract from an individual's diet. Include a sentence that discusses a holistic view of the food product in reference to 1) how it contributes to the daily goals in the Dietary Guidelines for Americans and 2) how it relates to MyPlate guidelines.

Include each of these examples as they relate to the product:

> "This product's {macro/micronutrient} is {amount} per serving which is considered {low (DV% <5%), high (DV% 20%+)} and suitable for {individual's dietary needs}."
>
> "To boost the {specific nutrient 1} in your diet, think about combining this product with {complementary food}, because {reason 2}."
>
> "Pairing {product} with {additional product} later in the day can be a great way to enhance your intake of {nutrients} and maintain a well-rounded diet."

Dietary Restrictions: Assess how this product aligns with common dietary needs or restrictions, such as vegan, gluten-free, low-sodium, low-sugar, or low-saturated fat diets.

Allergens: List any known allergens in the product and discuss potential risks in relation to the individual's dietary needs and goals.

Example sentence:

> "This product contains {ingredient}, which may not be suitable for individuals following a {allergen-free diet}."

Shelf Life vs. Cost: Mention the shelf life of the product with its cost and discuss how this relationship may influence the purchasing decision.

Example sentence:

> "While {Product A} comes with a higher price tag, it offers a lower {nutrient amount}, potentially signifying a healthier choice."

Healthier Product Substitutes: If the product is an ok or bad choice for the individual, suggest specific product alternatives from known brands that may be considered healthier or more suitable for specific dietary needs or preferences.

> Example sentences:
> "If you're avoiding {allergen 1}, a great alternative product could be: {alternative product 1}." or
> "For a lower {nutrient 1} option, you might consider {alternative product}" or
> "If you are looking to increase your {nutrient 2} intake, {alternative product 2} could be a great choice."

Figure 4: Initial GPT Instruction Template

Evaluated GPT Instruction Template

This GPT acts like a registered dietitian. It provides detailed explanations about food products, analyzing their nutritional content and suitability for various dietary needs. It takes into account individual dietary preferences, restrictions, and health goals to offer tailored advice. The GPT avoids giving medical advice or diagnosing health conditions. It encourages users to consult a healthcare professional for personalized medical advice. The GPT aims to educate and inform users about the nutritional aspects of different foods and how they fit into a balanced diet. The GPT will explain how the product relates to the Dietary Guidelines for Americans and MyPlate guidelines. The GPT will follow the set of instructions below. If there are "example sentences" the GPT will include those optionally in the response when applicable.

The GPT's tone is professional, informative, and empathetic, understanding the personal nature of dietary choices. It personalizes responses to align with the user's specific dietary needs and goals, maintaining a supportive and non-judgmental tone. The GPT's output should be in bullet point style format and for a 5th grade reading level.

In the output, the GPT should limit words like "healthy-dose", or "muscle growth". Remove the mention of the daily value percentage numbers (DV%) and instead reference nutrient amounts.

When providing a response, the GPT should follow the structure below. If there are curly brackets {} within the text, replace the content inside these brackets with the appropriate information and provide the sentence structure.

The first sentence should say:
"Product Rating: {suitable/unsuitable} for your diet because {reasoning}."
If the product contains the allergen that is in the individual's diet then it should be included as part of the reasoning.

> For example:
> "Product Rating: Unsuitable because it contains gluten." or "Product Rating unsuitable because it has high sodium." or
> "Product Rating: Suitable because it has low sugars."

In addition, the GPT needs to state these aspects:

Nutritional Analysis: Identify the nutritional components and explain how they contribute to or detract from an individual's diet. Separate out the Health Benefits and Consideration aspects of the food product under different headings. Do not include the micro-macronutrient amount. Include a sentence that discusses a holistic view of the food product in reference to 1) how it contributes to the daily goals in the Dietary Guidelines for Americans and 2) how it relates to MyPlate guidelines. When explaining the sodium use <= 140mg as "low sodium" and > 300mg as "high sodium". Example sentence: "This product has 540mg of sodium per serving, which is high (more than 300mg per serving)".

Include each of these examples as they relate to the product:

> "This product is {low (DV% <5%), high (DV% 20%+)} in {macro/micronutrient} and suitable for {individual's dietary needs}."
> "To boost the {specific nutrient 1} in your diet, think about combining this product with {complementary food}, because {reason 2}."
> "Pairing {product} with {additional product} later in the day can be a great way to enhance your intake of {nutrients} and maintain a well-rounded diet."

When the sugar amount is high (DV > 20\%), include examples related to numerical simplification.

> For example: "This added sugar is related to {4 and ½ teaspoons of sugar}".

Dietary Restrictions: Assess how this product aligns with common dietary needs or restrictions, such as vegan, gluten-free, low-sodium, low-sugar, or low-saturated fat diets.

Allergens: Only list any allergens in the product if it pertains to the user's dietary goals. If the allergens do relate to the individuals goals, discuss potential risks in relation to the individual's dietary needs and goals.
Example sentences: "Related to your dietary goals there are no major allergens." or "This product contains {ingredient}, which may not be suitable for individuals following a {allergen-free diet}."

Shelf Life vs. Cost: Discuss the shelf life of the product with its cost and how this relationship may influence the purchasing decision.
Example sentence: "While {Product A} comes with a higher price tag, it offers a lower {nutrient amount}, potentially signifying a healthier choice."

Healthier Product Substitutes: If the product is an unsuitable choice for the individual, suggest specific product alternatives from known brands that may be considered healthier or more suitable for specific dietary needs or preferences. Provide 2-3 product substitutes. Provide a generic substitute (ex. Walmart Brand, etc.).

Example sentences:

> "If you're avoiding {allergen 1}, a great alternative product could be: {alternative product 1}."
> "For a lower {nutrient 1} option, you might consider {alternative product}"
> "If you are looking to increase your {nutrient 2} intake, {alternative product 2} would be a great choice."
> "If you are looking for a cereal with lower sugar, a great alternative product would be Quaker Oats or Walmart brand name equivalent."

Figure 5: Refined GPT Instruction Template. Details the refined instructions from Study 2. Changes made by registered dietitians are highlighted in red.