NutriGen: Personalized Meal Plan Generator Leveraging Large Language Models to Enhance Dietary and Nutritional Adherence

Saman Khamesian[1,2], Asiful Arefeen[1,2], Stephanie M. Carpenter[1], and Hassan Ghasemzadeh[1]

Abstract— Maintaining a balanced diet is essential for overall health, yet many individuals struggle with meal planning due to nutritional complexity, time constraints, and lack of dietary knowledge. Personalized food recommendations can help address these challenges by tailoring meal plans to individual preferences, habits, and dietary restrictions. However, existing dietary recommendation systems often lack adaptability, fail to consider real-world constraints such as food ingredient availability, and require extensive user input, making them impractical for sustainable and scalable daily use. To address these limitations, we introduce NutriGen, a framework based on large language models (LLM) designed to generate personalized meal plans that align with user-defined dietary preferences and constraints. By building a personalized nutrition database and leveraging prompt engineering, our approach enables LLMs to incorporate reliable nutritional references like the USDA nutrition database while maintaining flexibility and ease-of-use. We demonstrate that LLMs have strong potential in generating accurate and user-friendly food recommendations, addressing key limitations in existing dietary recommendation systems by providing structured, practical, and scalable meal plans. Our evaluation shows that Llama 3.1 8B and GPT-3.5 Turbo achieve the lowest percentage errors of 1.55% and 3.68%, respectively, producing meal plans that closely align with user-defined caloric targets while minimizing deviation and improving precision. Additionally, we compared the performance of DeepSeek V3 against several established models to evaluate its potential in personalized nutrition planning. Our results showed that it struggled with accuracy, exhibiting a higher MAE of 10.45%, and was significantly slower in processing time compared to models like GPT-3.5 Turbo and Llama 3.1 8B, highlighting the need for further optimization.

Index Terms— personalized food recommendation, diet, large language models, prompt engineering, AI, machine learning, eHealth, behavioral health.

I. INTRODUCTION

Maintaining a healthy diet is fundamental to preventing and managing a range of chronic diseases, including obesity, cardiovascular conditions, and diabetes [1]. Despite its importance, many individuals face significant challenges in adhering to a healthy eating regimen. These difficulties often stem from the need to calculate daily caloric intake, evaluate the nutritional value of meals, and consistently make informed decisions about what to eat. Additionally, planning balanced meal schedules or preparing meals with the ingredients available at home can be overwhelming, especially when trying to align with specific dietary goals and restrictions. These barriers highlight the pressing need for tools that simplify these decision making processes, making it easier for individuals to adopt and sustain healthier eating habits over time [2].

To address the aforementioned challenges, personalized dietary recommendation systems have emerged as a promising solution. Unlike conventional food recommendation systems, which primarily consider whether users enjoy specific foods, personalized approaches cater to both individual dietary preferences and health needs, such as gluten-free diets, allergy sensitivities, and vegetarian lifestyles [3]. These systems aim to meet unique nutritional requirements, ensuring better adherence and promoting long-term health outcomes [4], [5]. Furthermore, nutrition-oriented food recommendations go beyond basic food preferences by integrating the broader impacts of dietary patterns on individual health. By aligning personal preferences with evidence-based nutritional guidelines, these systems offer actionable insights to help users make informed and health-conscious dietary decisions [6].

The rise of machine learning (ML) and data analysis techniques has further enhanced the effectiveness of personalized dietary recommendations [7]–[9]. By leveraging user data, these advanced systems can generate insights into optimal eating patterns while considering individual preferences. For example, DietOS employs optimization algorithms to develop customized diet plans tailored to the management of chronic diseases [10]. Similarly, IoMT-assisted systems, such as the approach by Iwendi et al., utilize ML to create precise, patient-specific dietary recommendations, excelling in clinical environments [11]. ChatDiet integrates large language models (LLM) to deliver interactive and engaging dietary advice via chatbot interfaces, significantly enhancing user interaction [6]. Additionally, optimization techniques like Particle Swarm Optimization (PSO) have been applied to generate meal plans that efficiently balance nutritional goals [12]. These innovative methodologies illustrate the potential of computational tools in addressing dietary challenges.

However, despite their potential, existing dietary recommendation systems face notable limitations that restrict their practical application. Methods such as PSO, while efficient at balancing nutritional objectives, often fail to account for user-specific preferences like cultural restrictions or diverse dietary needs, which reduces their real-world usability [12]. Similarly, IoMT-assisted systems, though highly precise and robust in clinical settings, are not well-suited for the general population seeking practical, everyday dietary guidance. Their reliance on continuous medical monitoring, clinical

---

[1]College of Health Solutions, Arizona State University, Phoenix, AZ 85004, USA.

[2]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA.

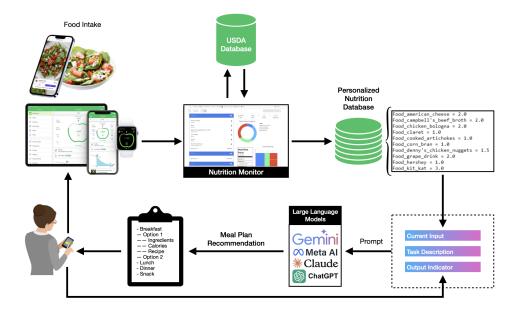Corresponding author: skhamesi@asu.edu

Fig. 1. The overall architecture of the NutriGen framework for personalized meal plan recommendations. The system integrates user inputs, external nutrition databases, and LLMs to generate structured and practical meal plans.

validation, and specialized sensors limits their usability outside structured healthcare environments, reducing their practicality for everyday meal planning [11]. ChatDiet excels in user engagement but struggles with scalability due to high computational costs, platform dependencies, and reliance on continuous user input [6]. Moreover, many existing systems focus solely on recommending individual food items rather than offering comprehensive and cohesive diet menus. This leaves users with the burden of assembling balanced meal plans on their own. Additionally, these systems frequently lack features such as alternative food suggestions, consideration of ingredient availability, or integrated recipes and cooking instructions. As a result, users may find it challenging to implement the recommendations in their daily lives. These challenges often lead to steady declines in engagement over time, with many participants discontinuing their use of the tools altogether. Addressing these gaps requires the development of adaptable, user-centered frameworks that provide structured, actionable, and personalized dietary guidance [13]–[16].

In this paper, we introduce NutriGen, a personalized meal plan recommendation system designed to overcome the limitations of existing methods by generating comprehensive and user-friendly daily meal plans. The system offers multiple meal plans, each including options for breakfast, lunch, dinner, and snacks, along with calorie counts, nutritional information and recipes. By leveraging LLMs and prompt engineering, the framework incorporates user preferences, restrictions, and ingredient availability to deliver fully personalized and practical dietary recommendations. This approach bridges the gap between nutrition science and usability, offering a flexible solution to promote sustainable healthy eating habits.

II. METHODOLOGY

In this section, we present our NutriGen framework for generating personalized meal plans based on users' food intake history, habits, preferences, and dietary restrictions. Existing dietary recommendation systems often lack adaptability, failing to accommodate diverse user needs and real-world constraints. NutriGen addresses these limitations by leveraging LLMs and prompt engineering to integrate user-specific inputs into a structured nutrition database. This approach ensures that meal plans align with individual dietary goals while incorporating features such as calorie counts, alternative food options, and suggested recipes. These enhancements improve the practical usability of personalized dietary recommendations [17]–[19]. The overall architecture of NutriGen is illustrated in Fig. 1.

A. Input Data

The system collects input from multiple sources, including food trackers that utilize image capture, manual text entry, and voice input. Image-based food logging employs optical character recognition (OCR) and ML models to estimate portion sizes and extract nutritional details, automating the process and reducing user effort. Text and voice-based inputs leverage natural language processing (NLP) to interpret descriptions, enabling flexible and user-friendly data entry. Together, these methods create a comprehensive record of a user's dietary intake, improving the accuracy and relevance of generated meal plans [20].

To enhance reliability, user-reported data is supplemented with standardized nutritional values retrieved from external sources, such as the United States Department of Agriculture (USDA) database[1]. Cross-referencing with established

---

[1]https://fdc.nal.usda.gov/

dietary guidelines helps mitigate potential inaccuracies in manual entries while ensuring consistency in nutrient estimations. Additionally, the system can integrate multiple data repositories, including branded food databases and nutrition labels, allowing for greater adaptability across different dietary contexts. Over time, user interactions, preferences, and feedback are incorporated into the nutrition database, refining recommendations dynamically to better align with individual dietary goals.

B. Architecture

Once the personalized nutrition database is constructed, it serves as the foundation for prompt engineering. The input prompt for the language model is structured into three components: the current input, the task description, and the output indicator. The structured prompt P is formally defined as:

$$P = \{I_{\text{current}}, T_{\text{task}}, O_{\text{output}}\}$$

where:

- I_{current} represents the user's dietary profile, including food intake history, preferences, and constraints.
- T_{task} defines the task with structured instructions, such as generating meal plans that meet specific nutritional targets (e.g., calorie goals, macronutrient distribution, dietary restrictions).
- O_{output} specifies the expected format, ensuring clarity and adherence to user-defined requirements.

Using this structured prompt, the meal plan generation process is formulated as:

$$M_i = \text{LLM}(P)$$

This approach enables self-contained meal generation without external dependencies, ensuring that outputs are consistent, personalized, and aligned with user-defined goals.

A potential extension to improve the accuracy of nutritional estimations in meal plan generation is the Retrieval-Augmented Generation (RAG) framework. RAG integrates a retriever module, $Q(x)$, which queries an external knowledge base such as the USDA database for a given meal x, with a generator module, $G(Q(x))$, powered by an LLM. This approach allows the system to dynamically retrieve up-to-date nutritional values, reducing discrepancies between estimated and actual calorie counts. Incorporating RAG into the system would allow the meal plan generation process to be formalized as:

$$M_i = G(Q(x), P) = \text{LLM}(Q(x), P)$$

where:

- $Q(x)$ retrieves the USDA-referenced nutritional values for meal x, ensuring accuracy.
- $G(Q(x))$ generates a meal plan that aligns with the retrieved nutritional information and user specifications.
- P serves as the structured prompt, integrating user preferences and constraints into the generation process.

While RAG offers the advantage of retrieving real-time USDA nutritional values, it also introduces practical challenges such as increased latency, token limits, and dependency on external APIs. Future work may explore hybrid approaches, where retrieved nutritional data is pre-processed and integrated into the structured prompt before querying the LLM.

C. Prompt Engineering

In designing the prompt, we structured it to ensure meal plans are personalized, nutritionally balanced, and aligned with user-defined goals. The prompt accommodates diverse dietary preferences, restrictions (e.g., gluten-free, vegetarian), and specific nutritional targets by explicitly defining constraints such as calorie limits, macronutrient distributions, and portion sizes. To maintain consistency and clarity across model outputs, we employed a modular structure, where different sections progressively refine the generated meal plans.

To further enhance reliability, we incorporated few-shot prompting by providing structured examples within the prompt. These examples serve as reference points, guiding the LLM to follow a consistent format while allowing diversity in meal recommendations. Additionally, our prompt design supports flexible constraints, enabling users to prioritize specific aspects such as high-protein meals or low-sugar alternatives. Below is the core section of the structured template prompt used in all our tests. The complete prompt and implementation code [2] is available for reference.

```
User's goal is to create a meal plan with:
- Total calories: {total_calories} kcal.
- Total protein: {target_protein} g.
- Total sugar: {target_sugar} g.
The plan must include:
- Breakfast, lunch, dinner, and snacks.
- The calorie count for each meal (e.g.,
"Breakfast: 400 kcal").
- At the end of each meal plan option,
provide the total calories, total fat, total
protein, and total carbohydrate.
- For each meal component, specify portion
sizes, including the number of items or
volume (e.g., "1 Kit Kat bar (45g)," "1
hamburger with a 150g beef patty, bun, and
lettuce").
- Provide a short recipe for each item,
detailing how it can be prepared (e.g.,
"Grill the patty for 5 minutes, then
assemble with lettuce, tomato, and a bun").
Provide three different meal plan options
for diversity.
Use familiar dishes instead of listing
individual food items. For example, use
"hamburger" instead of "150 grams of meat
with bun and lettuce."
Ensure the plan adheres to the user's
preferences and restrictions and meets
the specified targets while maintaining a
balanced nutritional profile.
Here are the available items: {menu_input}
```

III. EXPERIMENTAL SETUP

A. Dataset

The input prompt requires information on the user's consumption patterns, target calorie intake, and intervention duration as inputs to the system. To generate meal plans using LLMs without conducting a user study, we simulated realistic initial consumption patterns. Instead of collecting dietary data from human participants, we designed a system to generate ten diverse and plausible daily food intake profiles using a set of 200 randomly selected meals from the USDA dataset. To introduce variability and ensure a broad representation of dietary scenarios, calorie goals and intervention durations were randomly selected from a reasonable range (1500–3500 kcal). These synthetic consumption patterns, along with corresponding diet goals and durations, were then incorporated into input prompts to generate personalized meal plan recommendations. This approach allowed us to bypass the logistical and ethical challenges associated with human data collection while still evaluating the system's ability to tailor recommendations. The following box presents a sample nutrition profile with predefined food items and nutritional targets. These values were used as input for the LLMs, which were then tasked with generating meal plans that align with these specifications.

```
Food_Barbequeue_Lays = 1.0
Food_Garden_Pizza = 1.0
Food_Milano_double_chocolate = 1.0
Food_baked_cheddar_ruffles = 1.0
Food_beef_angus_burger_patty = 1.0
Food_chocolate_milkshake = 1.0
Food_eggs_benedict = 0.5
Food_tortilla_chips = 1.0
_____
calories = 1573.25
protein = 54.0
sugar = 58.3
```

B. Selected Models

In our experimental evaluations, we selected several advanced language models from leading organizations:

- **Llama 3.1-70B [21]**: Developed by Meta, this model comprises 70 billion parameters and has been trained on approximately 15 trillion tokens, demonstrating enhanced performance in language understanding and generation tasks.
- **Llama 3.1-8B [21]**: A smaller variant from Meta, this 8 billion parameter model is designed for efficiency while maintaining robust language processing abilities.
- **Claude 3.5 Sonnet**: Anthropic's model, with an undisclosed number of parameters, emphasizes safety and reliability, exhibiting improved capabilities in coding and reasoning tasks.
- **Claude 3.5 Haiku**: A streamlined version focusing on rapid response generation and concise outputs, tailored for applications requiring brevity and speed.

- **DeepSeek-V3 [22]**: Released by DeepSeek in December 2024, this model features 671 billion active parameters and a context length of 128K tokens, excelling in multilingual tasks, particularly in English and Chinese.
- **Gemini 2.0 Flash Exp [23]**: Developed by Google DeepMind, this model combines language understanding with reinforcement learning to enhance reasoning and problem-solving.
- **Gemini 1.5 Pro [23]**: Offers advanced language understanding and reasoning capabilities, optimized for professional application.
- **GPT-4o [24]**: OpenAI's latest model supports both text and image inputs, maintaining high intelligence levels across various tasks.
- **GPT-4o Mini**: A smaller model designed for high performance with lower resource usage.
- **GPT-3.5 Turbo [25]**: An optimized version of GPT-3, offering faster response times and improved efficiency in language tasks.

These models were chosen to provide a comprehensive assessment of current advancements in language model development. In selecting the models, we aimed to include at least two from each model family—one representing the latest and most powerful model and another optimized for speed and efficiency. This approach ensures that our evaluation covers a diverse range of models, from large-scale, high-performance architectures to lightweight, fast alternatives, thereby capturing the full spectrum of capabilities available in modern language models.

IV. RESULTS

A. Processing Time Comparison

To evaluate the computational efficiency of each model, we measured the total processing time required to generate 10 outputs. The results, presented in Fig. 2, show significant variations in execution time across different models. As expected, smaller and optimized models such as **GPT-3.5 Turbo, Gemini 2.0 Flash, and Claude 3.5 Haiku** demonstrated the fastest processing times, completing the task in significantly less time compared to larger models.
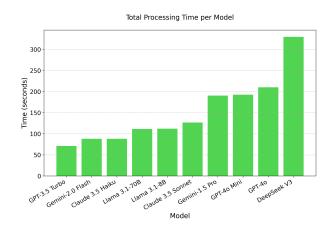

Fig. 2.   Comparison of total processing time

TABLE I
EVALUATION OF NUTRITION FACTS GENERATED BY LLMs: THIS TABLE PRESENTS THE ERROR BETWEEN THE AVERAGE REPORTED TOTAL CALORIES AND THE USDA DATABASE FOR VARIOUS MEAL PLANS ACROSS DIFFERENT MODELS. BOLD NUMBERS INDICATE THE MINIMUM ERROR, WHILE A DASH (-) SIGNIFIES THAT THE MODEL FAILED TO GENERATE A MEAL PLAN.

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8	Input 9	Input 10
Claude 3.5 Haiku	570.69	124	88	103.5	-	75	76.75	100	83	113
Claude 3.5 Sonnet	138.09	124	88	105	96	75	77	96	83	113
DeepSeek V3 [22]	138.09	96.6	88	-	-	-	-	100	82.84	113
Gemini 1.5 Pro [23]	588	173.59	87.75	153.5	445.5	75	76.75	96	83	52.59
Gemini 2 Flash [23]	589.33	173.59	87.75	153.5	445.5	75	103.41	96	83	48
GPT-3.5 Turbo [25]	138.09	73.59	37.75	53.5	**45.5**	**25**	**26.75**	46	**32.84**	62.59
GPT-4o [24]	50	**24**	**38**	**45**	563	41.6	50	**36**	74.6	52.59
GPT-4o Mini	60.03	27.5	50	50	161.66	55	**26.75**	47.33	35	53.33
Llama 3.1 8B [21]	88.09	58.33	50	60	46	58.33	**26.75**	-	50	**43.33**
Llama 3.1 70B [21]	**45**	60	50	50	46.66	40	48.33	46	50	63

In contrast, larger models required approximately twice the time to generate outputs. DeepSeek V3 exhibited the longest runtime, indicating higher computational demands. The comparison also highlights that while certain high-performance models, such as GPT-4o and Gemini 1.5 Pro, provide advanced capabilities, they require considerably more time to process outputs. These findings emphasize the trade-offs between model size, computational efficiency, and response time, which are critical considerations for real-time applications.

B. Reported Nutritional Values vs. USDA Dataset

One crucial aspect of our analysis is verifying the accuracy of nutrition facts generated by LLMs. Each model was asked to generate three meal plans for each input, and we computed the average absolute error between the total reported calories and the USDA reference values. When we ask a model to create a meal plan and compute total nutritional values, such as calories, fats, and protein, it is essential to assess whether the provided information is accurate. To validate these outputs, we used the USDA database as a reference and examined whether the nutritional values generated for meal plan items align with the actual USDA values. The Mean Absolute Error (MAE) was calculated as follows:

$$MAE = \frac{1}{M} \sum_{j=1}^{M} \left| \sum_{k=1}^{P} c_{\text{reported},j,k} - \sum_{k=1}^{P} c_{\text{USDA},j,k} \right| \quad (1)$$

where $M = 3$ represents the number of meal plans generated per input, and P is the number of meal items in each meal plan. Here, $c_{\text{reported},j,k}$ denotes the reported calorie value of item k in meal plan j, while $c_{\text{USDA},j,k}$ represents the corresponding USDA reference value. This formula ensures that the total calories for each meal plan are first computed by summing up the calorie values of individual meal items. Then, the absolute difference between the reported total and the USDA reference is calculated for each meal plan. The average across the three meal plans provides a robust measure of the model's accuracy in estimating calorie values.

Table I presents the error between the average reported total calories per input and the corresponding USDA values.

The results demonstrate that **GPT-4o** and **GPT-3.5 Turbo** achieve the lowest error, alongside other models that exhibit their capability to retrieve accurate nutritional information from reliable sources.

C. Adherence of Meal Plans to User-Specified Targets

Another valuable analysis that provides insight into how well LLMs incorporate user preferences into their generated meal plans involves comparing the average total nutrition for each input against the specified target values. Fig. 3 illustrates the comparison between the average total calories in meal plans generated by each model and the target for each input.

To quantify this difference, we computed the MAE between the actual total calories and the target values, as presented in Table II. The MAE is calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{P} c_{\text{actual},i,j,k} - C_{\text{target},i} \right| \quad (2)$$

where:
- $N = 10$ represents the total number of inputs.
- $M = 3$ represents the number of meal plans per input.
- P represents the number of items in each meal plan.
- $c_{\text{actual},i,j,k}$ is the calorie content of the k-th food item in the j-th meal plan for input i.
- $C_{\text{target},i}$ is the target calorie value for input i.

This formulation ensures that the actual total calories for each meal plan are computed as the sum of the calorie content of all individual food items, providing a more detailed and accurate assessment of how closely generated meal plans align with the intended calorie targets.

As shown in Table II, **Llama 3.1 8B** and **GPT-3.5 Turbo** achieved the lowest MAE of 34.14 and 54.16 calories, respectively, with corresponding percentage errors of **1.55%** and **3.68%**. These results indicate that both models effectively generate meal plans that closely adhere to user-defined caloric targets, with minimal deviation. Meanwhile, GPT-4o Mini exhibited the highest absolute error of 329.06
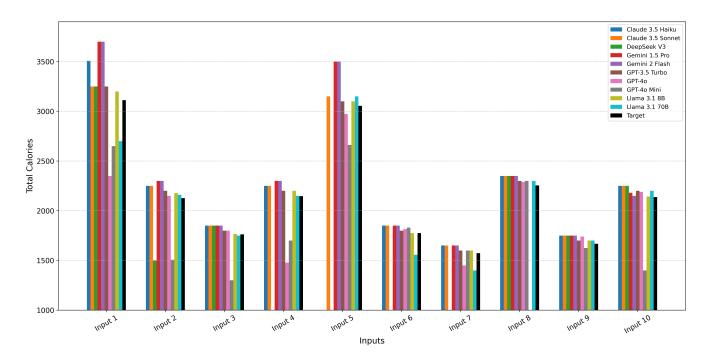
Fig. 3. Comparison of the average total calories in meal plans generated by each model against the specified target for each input. The black bars represent the target values, while colored bars indicate the total calories predicted by different models. This visualization highlights the variation in model performance and their ability to generate meal plans that align with user-defined nutritional goals.

TABLE II

MEAN ABSOLUTE AND PERCENTAGE ERROR BETWEEN TOTAL CALORIES AND TARGET FOR EACH MODEL

Model	MAE	MAE (%)
Claude 3.5 Haiku	128.23	8.99
Claude 3.5 Sonnet	99.16	4.85
DeepSeek V3 [22]	190.61	4.85
Gemini 1.5 Pro [23]	182.16	10.44
Gemini 2 Flash [23]	179.16	9.74
GPT-3.5 Turbo [25]	54.16	3.68
GPT-4o [24]	189.76	13.47
GPT-4o Mini	329.06	24.67
Llama 3.1 8B [21]	**34.14**	**1.55**
Llama 3.1 70B [21]	109.21	8.08

calories and the largest percentage error of 24.67%, highlighting substantial discrepancies in its caloric estimations. These findings indicate that while some models effectively minimize deviations in absolute values, others struggle with maintaining proportional accuracy, impacting their reliability in personalized meal planning.

D. Evaluation of Meal Plan Quality and Completeness

Beyond evaluating the nutritional accuracy of generated meal plans, it is equally important to assess whether they adhere to the instructions specified in the prompts. In this analysis, we focus on the structural completeness and richness of the meal plans, including their ability to provide three distinct meal plans, maintain alignment with the personalized nutrition profile, and include detailed recipes.

Our findings indicate that DeepSeek V3 failed to generate meal plans for inputs 4, 5, 6, and 7, while Claude 3.5 Haiku was unable to produce a meal plan for input 5, and Llama 3.1 8B failed to generate a meal plan for input 8. Additionally, Claude 3.5 Haiku produced only one meal plan instead of three for inputs 6, 8, 9, and 10. Another notable issue was observed with Gemini 1.5 Pro, which did not include nutritional values such as fat and protein in any of the generated meal plans, despite these targets being explicitly provided in the input. One possible explanation for why some models, particularly smaller versions, struggled to generate complete meal plans is the constraint on output token length. Each meal plan requires detailed recipes, nutritional values, and food items, resulting in lengthy outputs that may exceed the token limit imposed on smaller models.

It is also worth noting that all models except DeepSeek V3 and GPT-3.5 Turbo included an important disclaimer stating that the provided nutritional values may vary depending on portion sizes and could differ slightly from actual values. This suggests that most models recognize inherent variability in nutritional estimations, contributing to more realistic and informative meal plan generation.

V. FUTURE WORK

Despite its strengths, NutriGen has limitations that require further improvement. Some language models encountered output token constraints, leading to incomplete meal plans missing key nutritional details. Inconsistencies in calorie estimation also highlight the need for a more robust retrieval mechanism. Additionally, the use of simulated user data may limit the generalizability of the findings.

To address these issues, future work will explore RAG to dynamically fetch real-time USDA data, reducing retrieval errors and improving nutrient calculations. NutriGen also lacks interactivity, limiting user control over meal plan customization. Integrating a ChatBot interface would allow real-time updates to preferences, restrictions, and nutrition profiles, enhancing engagement and personalization. Further, incorporating multimodal LLMs with food image recognition could improve recommendations by analyzing visual dietary input, assessing ingredient availability, and suggesting alternatives. Although currently limited to English, NutriGen's language-agnostic design allows adaptation to other languages through localized databases and multilingual LLMs like GPT-3.5 Turbo.

VI. CONCLUSIONS

This study introduced NutriGen, a personalized meal plan generation framework that leverages LLMs and prompt engineering to bridge the gap between dietary science and practical food recommendation systems. By integrating user-defined dietary preferences, target nutritional goals, and USDA-referenced nutrition data, our approach enables LLMs to generate structured, comprehensive, and nutritionally aligned meal plans. Through rigorous evaluation, we demonstrated that LLMs can effectively generate diverse meal plans while maintaining flexibility, ease of use, and adherence to dietary constraints. Notably, Llama 3.1 8B and GPT-3.5 Turbo achieved the lowest percentage errors of 1.55% and 3.68%, respectively, closely matching user-defined caloric targets with minimal deviation. Additionally, GPT-3.5 Turbo exhibited one of the fastest response times, making it a practical choice for real-time meal generation while also demonstrating strong overall performance in minimizing MAE in calorie estimation.

VII. ACKNOWLEDGMENT

REFERENCES

[1] R. C. Brownson and F. S. Bright, "Chronic disease control in public health practice: looking back and moving forward," *Public Health Reports*, vol. 119, no. 3, pp. 230–238, 2004.

[2] N. Alshurafa, H. Kalantarian, M. Pourhomayoun, J. J. Liu, S. Sarin, B. Shahbazi, and M. Sarrafzadeh, "Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor," *IEEE sensors journal*, vol. 15, no. 7, pp. 3909–3916, 2015.

[3] J. Farquhar, P. Wood, H. Breitrose, W. Haskell, A. Meyer, N. Maccoby, J. Alexander, B. Brown JR, A. Mcalister, J. Nash *et al.*, "Community education for cardiovascular health," *The Lancet*, vol. 309, no. 8023, pp. 1192–1195, 1977.

[4] J.-C. Kim and K. Chung, "Knowledge-based hybrid decision model using neural network for nutrition management," *Information Technology and Management*, vol. 21, no. 1, pp. 29–39, 2020.

[5] A. Arefeen, N. Jaribi, B. J. Mortazavi, and H. Ghasemzadeh, "Computational framework for sequential diet recommendation: Integrating linear optimization and clinical domain knowledge," in *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2022, pp. 91–98.

[6] Z. Yang, E. Khatibi, N. Nagesh, M. Abbasian, I. Azimi, R. Jain, and A. M. Rahmani, "Chatdiet: Empowering personalized nutrition-oriented food recommender chatbots through an llm-augmented framework," *Smart Health*, vol. 32, p. 100465, 2024.

[7] N. E. Sherwood, M. L. Butryn, E. M. Forman, D. Almirall, E. M. Seburg, A. L. Crain, A. S. Kunin-Batson, M. G. Hayes, R. L. Levy, and R. W. Jeffery, "The bestfit trial: a smart approach to developing individualized weight loss treatments," *Contemporary clinical trials*, vol. 47, pp. 209–216, 2016.

[8] N. Hezarjaribi, S. Mazrouee, S. Hemati, N. S. Chaytor, M. Perrigue, and H. Ghasemzadeh, "Human-in-the-loop learning for personalized diet monitoring from unstructured mobile data," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 9, no. 4, pp. 1–24, 2019.

[9] N. Hezarjaribi, S. Mazrouee, and H. Ghasemzadeh, "Speech2health: a mobile framework for monitoring dietary composition from spoken data," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 252–264, 2017.

[10] G. Agapito, M. Simeoni, B. Calabrese, I. Caré, T. Lamprinoudi, P. H. Guzzi, A. Pujia, G. Fuiano, and M. Cannataro, "Dietos: A dietary recommender system for chronic diseases monitoring and management," *Computer methods and programs in biomedicine*, vol. 153, pp. 93–104, 2018.

[11] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realizing an efficient iomt-assisted patient diet recommendation system through machine learning model," *IEEE access*, vol. 8, pp. 28 462–28 474, 2020.

[12] V. Narendra, J. Christopher, and V. Arunachalam, "Particle swarm optimization for dietary recommendations," in *2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*. IEEE, 2023, pp. 1–5.

[13] X.-W. Wang, S. T. Weiss, F. B. Hu, and Y.-Y. Liu, "Optimization-based dietary recommendations for healthy eating," *bioRxiv*, pp. 2024–12, 2024.

[14] S. Khamesian, A. Arefeen, A. Grando, B. Thompson, and H. Ghasemzadeh, "Type 1 diabetes management using glimmer: Glucose level indicator model with modified error rate," *arXiv preprint arXiv:2502.14183*, 2025.

[15] M. Story, K. M. Kaphingst, R. Robinson-O'Brien, and K. Glanz, "Creating healthy food and eating environments: policy and environmental approaches," *Annu. Rev. Public Health*, vol. 29, no. 1, pp. 253–272, 2008.

[16] A. Mamun, L. D. Devoe, M. I. Evans, D. W. Britt, J. Klein-Seetharaman, and H. Ghasemzadeh, "Use of what-if scenarios to help explain artificial intelligence models for neonatal health," *arXiv preprint arXiv:2410.09635*, 2024.

[17] L. Wu, Z. Zheng, Z. Qiu, H. Wang, H. Gu, T. Shen, C. Qin, C. Zhu, H. Zhu, Q. Liu *et al.*, "A survey on large language models for recommendation," *World Wide Web*, vol. 27, no. 5, p. 60, 2024.

[18] J. Li, T. Tang, W. X. Zhao, J.-Y. Nie, and J.-R. Wen, "Pre-trained language models for text generation: A survey," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–39, 2024.

[19] P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, "A systematic survey of prompt engineering in large language models: Techniques and applications," *arXiv preprint arXiv:2402.07927*, 2024.

[20] L. Yang, C.-K. Hsieh, H. Yang, J. P. Pollak, N. Dell, S. Belongie, C. Cole, and D. Estrin, "Yum-me: a personalized nutrient-based meal recommender system," *ACM Transactions on Information Systems (TOIS)*, vol. 36, no. 1, pp. 1–31, 2017.

[21] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.

[22] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan *et al.*, "Deepseek-v3 technical report," *arXiv preprint arXiv:2412.19437*, 2024.

[23] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican *et al.*, "Gemini: a family of highly capable multimodal models," *arXiv preprint arXiv:2312.11805*, 2023.

[24] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford *et al.*, "Gpt-4o system card," *arXiv preprint arXiv:2410.21276*, 2024.

[25] J. Ye, X. Chen, N. Xu, C. Zu, Z. Shao, S. Liu, Y. Cui, Z. Zhou, C. Gong, Y. Shen *et al.*, "A comprehensive capability analysis of gpt-3 and gpt-3.5 series models," *arXiv preprint arXiv:2303.10420*, 2023.