

# Personalized Diabetes Management Using Large Language Models and CGM Data

Amani Matook Alhozali\*

Department of Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

\*Correspondence to: Amani Matook Alhozali (E-mail: aalhuthali@kau.edu.sa)

(Submitted: 27 January 2025 – Revised version received: 07 February 2025 – Accepted: 28 February 2025 – Published online: 26 April 2025)

## Abstract

**Objectives:** To develop a novel framework that leverages Large Language Models (LLMs) for generating personalized, context-aware diabetes management recommendations using Continuous Glucose Monitoring (CGM) data combined with patient lifestyle logs.

**Methods:** The proposed system transforms structured CGM data into natural language text and generates synthetic contextual log data, including meal descriptions and activity levels, based on health metrics. These textual inputs are then integrated and processed by an LLM to produce individualized, actionable recommendations tailored to each patient's unique glucose patterns and lifestyle context.

**Results:** The framework was evaluated through expert review, which assessed the clinical relevance, clarity, and practicality of the generated recommendations. The findings suggest that the model can provide coherent, understandable, and personalized guidance that may support improved self-management of diabetes.

**Conclusion:** This study demonstrates the potential of LLMs to enhance personalized diabetes care by converting complex medical data into accessible, patient-centric recommendations. The integration of CGM data with LLMs represents a promising direction for advancing intelligent, user-friendly digital health interventions.

**Keywords:** Diabetes management, continuous glucose monitoring, large language models, personalized recommendations

## Introduction

Diabetes, particularly Type 1 diabetes, is a chronic disease that requires continuous management and poses significant challenges to both patients and healthcare systems. Individuals with Type 1 diabetes must manage their condition daily by monitoring their diet, measuring their blood glucose, and administering insulin injections to maintain stable glucose levels. Fluctuation in blood sugar can cause immediate complications, such as hypoglycemia and hyperglycemia, or lead to long-term health risks such as cardiovascular disease, kidney failure, neuropathy, and vision. Despite these efforts, most adults with Type 1 diabetes do not achieve their recommended glucose control target levels. This lack of optimal management increases the risk of acute and chronic complications, with elevated levels of HbA1c recently being associated with worse outcomes in COVID-19.<sup>1</sup>

Effective diabetes management requires continuous monitoring and personalized interventions that take into consideration individual glucose levels, dietary habits, and physical activities. Technological advances in health care, such as Continuous Glucose Monitoring (CGM) systems, have transformed diabetes care by providing real-time insight into glucose trends.<sup>2</sup> However, despite their advanced capabilities, these systems alone cannot bridge the gap in diabetes management. Standard prediction models lacked the ability to fully utilize the huge amounts of data collected by diabetes patients, leaving many of the actionable insights unexplored.<sup>3</sup> Although researchers have successfully developed machine learning models that can predict glucose levels up to 30 minutes in advance, these models do not adequately incorporate contextual patient data, such as meal composition and physical activity, which are critical for personalized care.<sup>4</sup>

Despite the vast amount of data obtained from CGM systems and other health tracking devices, integrating this information with other contextual patient data

and transforming them into actionable personalized recommendations remains a significant challenge that has not been explored. Traditional diabetes management tools often rely on generalized guidelines that do not account for the dynamic nature of glucose fluctuations influenced by individual-specific factors, such as meal composition, stress, sleep, and exercise.<sup>5</sup> In addition, providing patients with a consistent and timely recommendation and feedback is a challenging task as manually reviewing CGM data by healthcare providers is time consuming and prone to variability. Although existing systems are capable of alerting patients and caregivers when glucose levels reach a critical point, they are limited in their ability to provide comprehensive context-sensitive recommendations.<sup>6</sup>

Recent advancements in artificial intelligence (AI) and natural language processing (NLP) have paved the way for new opportunities for innovation in healthcare. Large Language Models (LLMs), such as GPT and LLaMA, have shown remarkable capabilities to comprehend and produce human-like text.<sup>7</sup> These models offer a new dimension to make complex raw medical data more accessible and interpretable. By transforming glucose readings and contextual patient input, such as meal descriptions and activity logs, into natural language summaries and personalized recommendations, LLMs have the potential to bridge the gap between raw health data and actionable insights.<sup>8</sup> However, less attention has been given to the application of LLMs in diabetes management, with only a few studies exploring their integration of CGM data to provide dynamic, real-time advice.<sup>9</sup>

This study addresses these limitations by proposing a novel framework that integrates CGM data and patient-reported logs with LLMs to generate personalized, context-aware recommendations. The framework transforms structured CGM and patient log data into textual representations, enabling LLMs to analyze glucose trends and provide actionable recommendations. For example, the system can detect glucose spikes after meals and suggest personalized

dietary adjustments or identify periods of hypoglycemia and recommend immediate interventions. Using pre-trained LLMs fine-tuned for healthcare-specific tasks, the framework aims to improve patient understanding, adherence, and engagement with their diabetes management plans.<sup>10</sup>

To achieve this goal, we used the publicly available HUPA-UCM dataset,<sup>11</sup> which includes a diverse range of health metrics such as CGM data, insulin administration records, meal intake details (quantified in grams of carbohydrates) and physical activity metrics. To provide contextual input to the model (i.e., a description of meal content and physical activities, the framework first utilizes LLMs to generate synthetic contextualized log data based on carbohydrate intake, time, and glucose readings.

The generated synthetic logs were structured as natural language descriptions simulating real-world patient inputs. Each log combined information on meal content (e.g., “grilled chicken sandwich for lunch”), physical activity (e.g., “30 minutes of jogging in the evening”), and time-specific glucose observations. Variability in the logs was introduced based on the variability and values of insulin doses, carbohydrate intake, and physical activity recorded in HUPA-UCM dataset. This ensured that the synthetic data captured a wide range of realistic patient behaviors and physiological responses. To ensure medical plausibility and consistency, a two-stage validation process was employed. First, rule-based filters checked logical consistency between CGM trends and synthetic logs. Second, domain expert review was conducted to qualitatively assess a sample of the logs for realism and contextual fidelity. The logs were refined as needed to ensure consistency with domain knowledge and real-world patterns.

The framework then combines CGM-transformed and synthetic log data with LLM to generate patient-specific recommendations. The produced recommendations were evaluated through expert validation, demonstrating the framework's utility in delivering clear, accurate, relevant, and feasible recommendations. By focusing on personalized, data-driven insights, this study aims to empower patients, reduce the burden on clinicians, and promote proactive disease management. This work represents a significant step toward transforming diabetes care by integrating advanced AI and NLP technologies.

## Related Works

Large language model (LLM) integration in healthcare is developing quickly and has significant implications for personalized patient treatment, especially when it pertains to chronic diseases like diabetes. The literature reveals a growing recognition of the potential of LLMs to enhance clinical decision making, patient engagement, and overall care quality.

Connor et al.,<sup>12</sup> highlight the transformative capabilities of LLMs in sports medicine, noting their ability to augment practitioner knowledge and facilitate personalized training recommendations. However, they also caution against potential risks, such as data biases and confidentiality concerns, emphasizing the need for ongoing research to explore the full spectrum of LLM applications in specialized domains. This foundational understanding sets the stage for considering LLMs within medical contexts, as exemplified by the work of Li et al.,<sup>13</sup> who discuss the role of LLMs in advancing artificial

general intelligence (AGI) for medical imaging and diagnostics. They envision a future where LLMs assist in integrating diverse data modalities, thereby promoting a more comprehensive approach to patient care.

Yang et al.,<sup>14</sup> further elaborate on the practical applications of LLMs in medical settings, demonstrating their utility in optimizing documentation and enhancing patient consultations. They suggest that LLMs can streamline research processes and improve customer service in healthcare, indicating a broadening scope of LLM applications beyond mere data analysis to active patient interaction and support. This theme of enhancing communication and understanding continues in the work of Omiye et al.,<sup>15</sup> who investigate the performance of domain-specific LLMs like Med-PaLM, which exhibit strong clinical reasoning capabilities and improved accuracy in medical tasks. Their findings underscore the importance of tailored training in maximizing the efficacy of LLMs in healthcare.

He et al.,<sup>16</sup> provide a comprehensive survey of LLMs in healthcare, examining their development, capabilities, and ethical considerations. They highlight the need for accountability and transparency in deploying LLMs, which is particularly relevant in managing sensitive health data. This ethical lens is echoed by Lawson et al.,<sup>17</sup> who address the challenges of patient privacy and informed consent when utilizing LLMs for shared decision-making in neuro-oncology. Their insights stress the importance of safeguarding patient information while leveraging LLMs to enhance clinical interactions.

Andrew et al.,<sup>18</sup> discusses the potential for LLMs, such as Med-PaLM 2, to provide personalized healthcare recommendations and improve patient understanding of complex medical information. By simplifying medical terminology and tailoring advice to individual needs, LLMs can foster better patient engagement and adherence to treatment plans. Shahab et al.,<sup>19</sup> further emphasize the importance of ensuring that LLMs are accessible to patients with varying levels of digital literacy, advocating for clear communication and safety protocols in their deployment.

Finally, Healey et al.,<sup>20</sup> present a focused case study on applying LLMs to analyze continuous glucose monitoring (CGM) data. Their findings suggest that LLMs can effectively interpret complex time-series data and generate narrative summaries, thereby facilitating meaningful discussions between clinicians and patients. This study illustrates the potential for LLMs to bridge the gap between intricate data analysis and actionable patient insights, ultimately promoting improved glycemic control and behavioral change.

Collectively, these articles illustrate the multifaceted role of LLMs in healthcare, particularly in the realm of personalized diabetes management. As the literature reveals, while the promise of LLMs is considerable, careful consideration of ethical implications, data privacy, and the need for tailored applications will be crucial in harnessing their full potential for patient care.

## Methodology

This study presents a framework for using continuous glucose monitoring (CGM) data and patient-reported logs to develop personalized diabetes management recommendations using large language models (LLM). The pipeline, as shown in

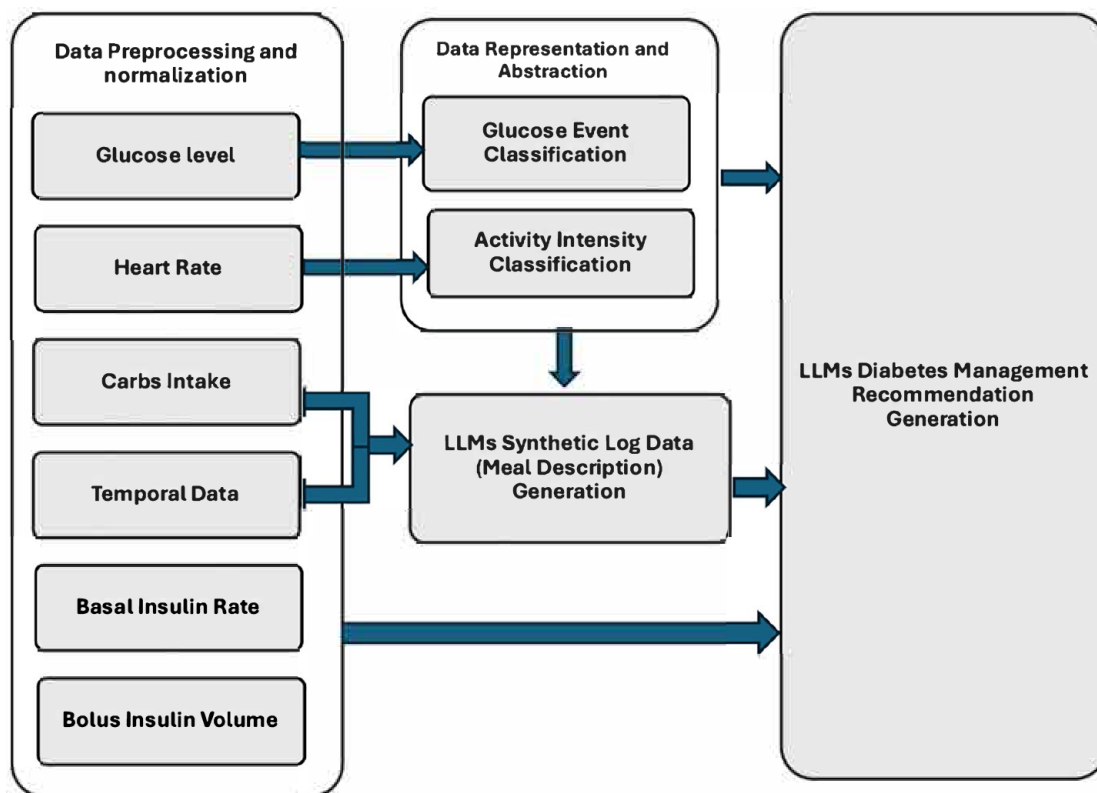


Fig. 1 An overview of the methodology used to generate diabetes management recommendations based continuous glucose monitoring (CGM) data and patient-reported logs.

Figure 1 encompasses six interconnected stages: data acquisition, pre-processing, data representation, synthetic and log data generation, recommendation generation and evaluation.

### Data Acquisition

The dataset used in this study combines real-world CGM data with synthetically generated patient logs to create a robust foundation for training and evaluation. The primary source of CGM data is the publicly available HUPAUCM dataset,<sup>11</sup> which provides continuous glucose readings, timing, insulin dosages, carbohydrates intake, and activity levels data acquire from 25 people with type 1 diabetes mellitus (T1DM). This dataset covers diverse patient profiles and glycemic patterns, making it suitable for training personalized diabetes management models.

The dataset includes data from 25 individuals with Type 1 Diabetes Mellitus (T1DM), monitored over a period of 4 weeks per patient, with CGM readings recorded every 15 minutes. It provides information on glucose levels, insulin doses (basal and bolus), carbohydrate intake, and physical activity. The dataset covers a range of ages (18–65 years), with a balanced representation of male and female participants. The average missing data rate across CGM records was approximately 6%, primarily due to short-term sensor disconnects or data upload issues. Key dataset statistics presented in Table 1.

To supplement the CGM data, synthetic patient logs were generated using LLMs based on the provided time, carbohydrates, and glucose level. These logs simulate realistic patient detailed meal descriptions and activity levels.

The inclusion of synthetic logs ensures the dataset captures a wide range of real-world scenarios, enhancing the generalizability of the system.

### Preprocessing Techniques

Preprocessing is a crucial step to align, normalize, and de-noise the data, ensuring it is compatible with transformer-based

Table 1. Summary statistics of the HUPA-UCM dataset

Characteristic	Value
Number of Patients	25
Age Range (years)	18–65
Gender Distribution	52% Male / 48% Female
Monitoring Duration per Patient	28 days
CGM Reading Frequency	Every 15 minutes
Total Number of CGM Records	~160,000
Average Daily Glucose (mg/dL)	156.2 ± 42.5
Average Daily Carbohydrate Intake (grams)	186.7 ± 56.4
Average Daily Bolus Insulin Dose (units)	25.4 ± 9.8
Average Daily Basal Insulin Rate (units/hour)	0.85 ± 0.35
Physical Activity Records Available	Yes (steps/hour, activity intensity level)
Missing CGM Data Rate	~6%

models. Given the noisy and unstructured nature of the health data, several pre-processing steps were applied.

### Temporal Alignment

Continuous Glucose Monitoring (CGM) data and patient logs were temporally aligned to maintain consistency across sources. Weekly cycles were adopted as the base unit to capture recurring patterns in glucose trends and lifestyle behaviors.

The dataset contained records from different sources (e.g., CGM, meal logs) with varying timestamps. All timestamps were converted to a consistent format (YYYY-MM-DD HH:MM:SS) to facilitate synchronization.

From the standardized timestamps, additional features were created to capture temporal patterns.

The time of day is categorized into morning, afternoon, and evening based on the hour.

### Glucose Normalization

Glucose readings were normalized to remove biases introduced by individual variations using the following formula:

$$G_{norm} = \frac{G_t - G_{mean}}{G_{std}}$$

where  $G_t$  is the glucose level at time  $t$ ,  $G_{mean}$  is the mean glucose level over the observation period, and  $G_{std}$  is the standard deviation of glucose levels.

### Trend and Event Detection

Key glucose events such as spikes, drops, and stable levels were identified using the following piecewise function:

$$G(t) = \begin{cases} \text{Spike, if } G_t > 180 \\ \text{Drop, if } G_t < 70 \\ \text{Stable, otherwise} \end{cases}$$

These trends were later converted into natural language summaries for the LLM. For example, an event where  $G_t = 220$  after breakfast was summarized as: "The glucose spiked to 220 mg/dL after breakfast."

### Noise Reduction

To minimize random fluctuations in the glucose data, a moving average filter was applied:

$$G_{smooth(t)} = \frac{1}{w} \sum_{i=t-w+1}^t G_i$$

where  $w$  is the window size used for smoothing,  $G_i$  represents individual glucose readings within the window.

### Activity Intensity

mapped to levels such as sedentary, light activity, or high intensity. To implement this classification, the following transformation is applied:

$$A = \begin{cases} \text{Sedentary, if heart rate} \leq 80 \\ \text{Moderate, if } 80 < \text{heart rate} \leq 120 \\ \text{Active, if heart rate} > 120 \end{cases}$$

### Handling Missing Data

Missing glucose readings, a common issue in time-series health data, were imputed using linear interpolation. This approach ensured continuity in the data while maintaining the integrity of the time-series patterns.

### Model Development

In this study, we utilized OpenAI's GPT-3.5 model, accessed via the API, as the primary Large Language Model (LLM) to generate natural language recommendations. The model was not fine-tuned on domain-specific datasets; instead, we applied prompt engineering to guide the model's behavior for healthcare-specific tasks. Prompt-tuning involved the use of carefully constructed instruction prompts that embedded contextual information—such as carbohydrate intake, CGM readings, and time of day—to simulate realistic patient scenarios.

LLMs typically rely on a prompt that specifies the task to be performed. For our setting, a suitable prompt would entail providing the historical blood glucose values for time  $t$  and a window of historical nutritional intake and insulin needed to drive the decision process. However, LLMs struggle with modeling structured tabular data, like patient logs, especially when they include histograms or other statistical overviews across multiple attributes. Therefore, we employ a prompt-tuning mechanism to preprocess the structured log input into a node-specific prompt for each of the trees in our the model architecture. The proposed system integrates two LLMs designed for distinct but complementary tasks:

#### Dietary Model

The dietary model generates synthetic meal descriptions based on time, glucose levels, carbohydrate intake, and insulin bolus data.

- Inputs: Temporal summaries of glucose trends, meal timings, carbohydrate intake, and activity levels.
- Outputs: Natural language meal descriptions.
- For example, for Glucose Levels: Before Meal: 110 mg/dL; After Meal: 220 mg/dL, Timing duration: Morning, Activity Intensity: Sedentary, if heart rate <80 the generated meal Meal: 2 slices of white bread, 1 tablespoon of butter, 1 glass of orange juice.

The dietary model also incorporates a decentralized simulator that evaluates patient responses to the generated recommendations. The simulator creates synthetic data to enhance the model's ability to refine its predictions iteratively, ensuring that the recommendations are both personalized and adaptive to real-world variations in patient behavior.

The generation of synthetic meal descriptions was achieved using a predefined prompt tailored to guide the large language model (LLM) in creating realistic and contextually accurate meal details. The prompt was designed to incorporate key variables to produce detailed and coherent outputs. The structure of the prompt is as follows:

*Write a meal detail that someone has taken to cause their glucose level to be {glucose} mg/dL. The meal occurred in the {time\_period}, and included a carbohydrate intake of {carb\_input} grams. The activity intensity during this period was {activity\_intensity}. Generate a description of the meal. The description should include the type of meal, its components, and an appropriate beverage. Provide one single meal detail.*

Where:

- **{glucose}**: Represents the individual's glucose level in mg/dL at the specified time.
- **{time\_period}**: Indicates the meal timing, such as morning, afternoon, or evening.
- **{carb\_input}**: Specifies the carbohydrate intake in grams.
- **{activity\_intensity}**: Describes the physical activity level during the period, categorized as Sedentary, Moderate, or Active.

### Conversational/Advisory Model

The Conversational/Advisory Model provides personalized guidance on meal planning, glycemic control, and insulin management by leveraging historical recommendations and patient outcomes. This model facilitates ongoing patient engagement and adapts its advice based on feedback and previous actions.

- Inputs: Summaries of historical recommendations, glucose levels, insulin doses, and outcomes, formatted as conversational prompts.
- Outputs: Contextual strategies for better glycemic control.
- For instance, "Based on your recent glucose readings, consider reducing your carbohydrate intake during lunch and increasing fiber-rich foods."

To generate personalized diabetes management recommendations, a prompt was crafted to guide the large language model (LLM) using relevant patient data. The structured prompt provides detailed input variables and outlines the expected output in the form of actionable advice. The prompt is as follows:

*You are an expert in diabetes management. Given the following patient data:*

- Glucose Level: {glucose normalized} mg/dL
- Glucose Event: {glucose event}
- Activity Intensity: {activity intensity}
- Basal Insulin Rate: {basal rate} U/hr
- Heart Rate: {heart rate} bpm
- Bolus Insulin Volume Delivered: {bolus volume delivered} U
- Meal Description: {meal description}

*Provide a personalized recommendation to the patient for one of the following areas:*

- **Insulin Dosing:** Suggest adjustments to bolus or basal insulin doses based on the patient's glucose level, glucose event, and meal description.
- **Meal Planning:** Provide advice on modifying meal composition or portion sizes for better glycemic control, considering the glucose event.
- **Timing Adjustments:** Recommend changes to the timing of insulin administration or meal consumption based on the glucose event and the patient's data.
- **Physical Activity:** Suggest exercise routines or adjustments based on steps and heart rate data, considering the glucose event.
- **Managing Hypoglycemia:** Provide guidance on treating or preventing low glucose levels based on the glucose event and other patient data. Choose the best type of recommendation given the provided scenario. Your recommendation should be concise, actionable, and clinically relevant. Ensure that your response is easy to understand while maintaining a professional tone.

Both models are built on transformer architectures, employing multi-head self-attention mechanisms to capture complex dependencies between glucose levels, patient behaviors, and contextual factors. The integration of these models allows for a comprehensive approach to diabetes management, addressing both immediate dosing needs and longer-term behavioral adjustments.

### Performance Measures

We evaluated the proposed approach through the use of medical professionals who reviewed the generated outputs for clinical relevance and practicality. The evaluation was conducted across four key categories: clinical accuracy, relevance, clarity, and feasibility. Clinical accuracy focuses on ensuring the recommendations align with established medical guidelines and evidence-based practices. Relevance measures the alignment of the recommendations with the patient's provided data, such as glucose trends, activity levels, and carbohydrate intake, ensuring the outputs are meaningful and applicable to the patient's context. Clarity evaluates the readability and comprehensibility of the recommendations, emphasizing the use of clear and concise language that avoids ambiguity. Feasibility assesses the practicality of implementing the recommendations, considering the patient's lifestyle, resources, and typical capabilities, such as accessible meal planning or achievable physical activity suggestions.

## Results

The proposed framework was evaluated on the HUPA-UCM Diabetes Dataset using qualitative expert validation. Three domain experts have participated in this evaluation, three are diabetes and endocrine consultants with more than ten years of experience. They evaluated a set of recommendations categorized into different categories, such as meal planning and insulin dosing based on four criteria: clinical precision, relevance, clarity, and feasibility. The model achieved an average clinical accuracy score of 4.6, reflecting its strong adherence to established medical guidelines. Relevance was rated at 4.5, highlighting the model's ability to generate recommendations tailored to patient-specific data. Clarity received a score of 4.4, showing that the outputs were easy to understand and free from ambiguity. Feasibility was rated slightly lower at 4.4, suggesting that while most recommendations were practical, there were occasional challenges in addressing real-world constraints, as shown in Table 2.

Table 2. Average domain expert scores across key evaluation metrics for each recommendation category and overall

Category	Clinical Accuracy	Relevance	Clarity	Feasibility
Overall Average	4.6	4.5	4.5	4.4
Meal Planning	4.7	4.4	4.7	4.6
Insulin Dosing	4.7	4.5	4.7	4.7
Managing Hypoglycemia	4.8	4.7	4.8	4.5
Physical Activity	3.8	4.3	3.7	3.5
Timing Adjustment	4.7	4.7	4.3	4.7

Cohen's Kappa coefficient was used to assess the agreement among the expert reviewers. The scores obtained were: Clinical Accuracy (0.41), indicating moderate agreement; Relevance (1.0), showing perfect agreement; Clarity (0.40), reflecting moderate agreement; and Feasibility (0.38), which also demonstrated moderate agreement. These scores highlight the reliability of the evaluation process, particularly highlighting the strong alignment on the relevance of the recommendations, which is crucial for ensuring their personalized and meaningful nature. The moderate agreement observed in clinical accuracy, clarity, and feasibility suggests some variability in interpretations.

Table 2 presents examples of synthetic meal descriptions generated by the Dietary Model. These descriptions were contextualized using inputs such as glucose level, carbohydrate intake, and time of day. For example, given a glucose level of 150 mg/dL, carbohydrate intake of 60 g, and a meal occurring in the evening, the model generated the following description: "Grilled chicken sandwich with sweet potato fries and a glass of lemonade." These synthetic descriptions were integral to enhancing the dataset's representativeness and were subsequently used as inputs for generating personalized recommendations.

Table 2 also shows the model's recommendations across different scenarios, categorized into insulin dosing, meal planning, timing adjustments, physical activity, and managing hypoglycemia. For instance, in response to the aforementioned synthetic meal description, the model recommended increasing bolus insulin by 1.0 U to address the post-meal glucose spike, reducing carbohydrate intake to 45 g during lunch, and incorporating a 15-minute walk post-meal to stabilize glucose levels.

The results demonstrate the potential of the proposed framework in generating personalized diabetes management

recommendations. The high expert evaluation scores validate the effectiveness of using a pre-trained LLM in a zero-shot setting, allowing the model to capture contextual nuances without the need for extensive fine-tuning. The inclusion of synthetic meal descriptions and activity logs, as shown in Table 3, enhanced the dataset's diversity, enabling the model to generalize effectively across various patient scenarios.

The model gave good evaluation in generating insulin dosing and meal planning recommendations, which require precise numerical reasoning and domain-specific knowledge. However, challenges were noted in generating physical activity recommendations, where some suggestions were oversimplified, such as recommending light exercise for patients with high glucose spikes despite low activity intensity. Expert feedback underscored the clarity and clinical accuracy of the recommendations as significant strengths, but slightly lower feasibility scores, as noted in Table 2, indicated a need to refine the model's understanding of real-world constraints, such as patient access to recommended foods or time for exercise.

## Discussion

An analysis of the results show that recommendations for patients with "Spike" glucose events generally scored well in Clinical Accuracy and Relevance, indicating that the model effectively addressed high glucose scenarios.

While, recommendations for "Stable" glucose events were clear and feasible but occasionally lacked adjustments for practical or patient-specific considerations. However, the management of hypoglycemia (for example, was handled effectively for some patients, with high scores in all categories, reflecting clear and actionable recommendations. For

Table 3. **Personalized recommendations for diabetes management based on patientspecific data, including glucose levels, activity intensity, and meal descriptions. Each recommendation addresses key areas such as insulin dosing, meal planning, timing adjustments, physical activity, or managing hypoglycemia**

Patient Data	Meal Description	Recommendation
Glucose Level: 89 mg/dL Glucose Event: Stable Activity Intensity: Sedentary Basal Insulin Rate: 0.0 U/hr Heart Rate: 61 bpm Bolus Insulin Volume Delivered: 0.0 U	A light breakfast of whole-grain toast, a small banana, and a cup of unsweetened green tea.	Meal Planning: Maintain this balanced meal composition but consider adding a protein source, such as a boiled egg, to prevent mid-morning glucose drops.
Glucose Level: 271 mg/dL Glucose Event: Spike Activity Intensity: Sedentary Basal Insulin Rate: 0.1 U/hr Heart Rate: 71 bpm Bolus Insulin Volume Delivered: 0.0 U	A large dinner including pasta with a creamy alfredo sauce, garlic bread, and a sweetened iced tea.	Insulin Dosing: Increase bolus insulin by 2.0 U before such meals to manage post-meal glucose spikes. Meal Planning: Replace pasta with a smaller portion of whole-grain pasta and opt for water instead of a sweetened beverage.
Glucose Level: 182 mg/dL Glucose Event: Spike Activity Intensity: Moderate Basal Insulin Rate: 0.0 U/hr Heart Rate: 93 bpm Bolus Insulin Volume Delivered: 0.0 U	A moderate lunch of grilled chicken salad with vinaigrette dressing, accompanied by a fruit smoothie.	Timing Adjustments: Administer bolus insulin 15 minutes before lunch to better manage the glucose spike. Physical Activity: Consider a short walk post-lunch to stabilize glucose levels.
Glucose Level: 44 mg/dL Glucose Event: Drop Activity Intensity: Sedentary Basal Insulin Rate: 0.07 U/hr Heart Rate: 76 bpm Bolus Insulin Volume Delivered: 0.0 U	A small breakfast of scrambled eggs, a slice of whole-grain toast, and a cup of black coffee.	Managing Hypoglycemia: Consume a fast-acting carbohydrate source, such as a small glass of fruit juice, to quickly elevate glucose levels. Recheck glucose in 15 minutes.

example, the input data indicated a glucose level of 44 mg/dL, classified as a drop, with a sedentary activity intensity and a meal consisting of scrambled eggs, a slice of whole-grain toast, and a cup of black coffee. Based on this scenario, the recommendation focused on managing hypoglycemia, advising the patient to consume a fast-acting carbohydrate source, such as a small glass of fruit juice, to quickly elevate glucose levels. It also recommended rechecking the glucose in 15 minutes to ensure stabilization. This recommendation was evaluated highly by domain experts. Other hypoglycemia recommendation received lower scores due to challenges in clarity and feasibility. For example, the input data indicated a glucose level of 77 mg/dL, classified as stable, with a sedentary activity level and a meal consisting of Greek yogurt topped with granola, a handful of almonds, and a glass of fresh orange juice. Based on this scenario, the recommendation suggested encouraging light physical activity, such as a short walk after breakfast, to maintain stable glucose levels. It also advised reducing the intake of orange juice to prevent potential glucose spikes later in the day. This recommendation received lower scores by domain experts, suggesting that the guidance may not have fully aligned with the patient's context or constraints.

Despite these strengths, the study has certain limitations. The reliance on synthetic data, while beneficial for diversifying the dataset, may introduce bias if the generated meal descriptions lack sufficient variability. Additionally, the system's performance has yet to be validated in real-world clinical settings, where patient outcomes could provide a more comprehensive evaluation of its effectiveness. Future work could focus on incorporating reinforcement learning to dynamically adapt recommendations based on patient feedback. Expanding the dataset with real-world patient logs would enhance the system's realism and accuracy. Furthermore, exploring multilingual capabilities could improve accessibility for diverse patient populations, broadening the impact of this approach.

## Ethical Considerations

This study used the publicly available and fully anonymized HUPA-UCM dataset, requiring no ethical approval or informed consent. While no personal data were involved, we emphasize the importance of privacy in future applications. Any clinical deployment would require compliance with relevant data protection regulations (e.g., HIPAA, GDPR).

The AI-generated recommendations are designed to support—not replace—clinical decision-making. To mitigate risks, outputs were reviewed by domain experts to ensure clinical relevance and safety.

The synthetic log generation process followed institutional ethical guidelines and was based on structured clinical variables without involving human subjects. Rule-based validation and internal expert review ensured the appropriateness of generated data.

## Conclusion

This study presents a novel framework for personalized diabetes management by leveraging Continuous Glucose Monitoring (CGM) data, patient-reported logs, and Large Language Models (LLMs). By integrating structured health data with the generative capabilities of LLMs, the framework demonstrated its ability to generate realistic meal descriptions and actionable recommendations across critical categories, including insulin dosing, meal planning, timing adjustments, physical activity, and managing hypoglycemia. Qualitative expert reviews validated the clinical relevance, clarity, and feasibility of the generated outputs, with an average score of 4.6 for clinical accuracy. These results underscore the potential of LLMs to provide precise, tailored recommendations for individuals with diabetes, addressing real-world scenarios through synthetic data generation and robust input representations.

Despite these promising outcomes, the study acknowledges several limitations. The reliance on synthetic logs, while enhancing dataset diversity, may introduce biases that affect the generalizability of the recommendations. Furthermore, the framework has not yet been validated in real-world clinical settings, where patient outcomes and environmental variables could provide a more comprehensive understanding of its efficacy. Future work will focus on expanding the dataset with real-world patient logs, integrating reinforcement learning to refine recommendations dynamically, and exploring multilingual capabilities to ensure accessibility for diverse patient populations. These advancements aim to enhance the system's reliability, adaptability, and scalability for widespread clinical application.

## Conflict of Interest

None. ■

## References

- Prattichizzo, F., Candia, P., Nicolucci, A. & Ceriello, A. Elevated HbA1c levels in pre-Covid-19 infection increases the risk of mortality: a systematic review and meta-analysis. *Diabetes/Metabolism Research And Reviews*. 38, e3476 (2022).
- Vashist, S. Continuous glucose monitoring systems: a review. *Diagnostics*. 3, 385–412 (2013).
- Makroum, M., Adda, M., Bouzouane, A. & Ibrahim, H. Machine learning and smart devices for diabetes management: Systematic review. *Sensors*. 22, 1843 (2022).
- Plis, K., Bunesco, R., Marling, C., Shubrook, J. & Schwartz, F. A machine learning approach to predicting blood glucose levels for diabetes management. *Workshops At The Twenty-Eighth AAAI Conference On Artificial Intelligence*. (2014).
- Bertsimas, D., Kallus, N., Weinstein, A. & Zhuo, Y. Personalized diabetes management using electronic medical records. *Diabetes Care*. 40, 210–217 (2017).
- Boland, E., Monsod, T., Delucia, M., Brandt, C., Fernando, S. & Tamborlane, W. Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. *Diabetes Care*. 24, 1858–1862 (2001).
- Wang, D. & Zhang, S. Large language models in medical and healthcare fields: applications, advances, and challenges. *Artificial Intelligence Review*. 57, 299 (2024).
- Alowais, S., Alghamdi, S., Alsuhbany, N., Alqahtani, T., Alshaya, A., Almohareb, S., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. & Others Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*. 23, 689 (2023).
- Contreras, I. & Vehi, J. Artificial intelligence for diabetes management and decision support: literature review. *Journal Of Medical Internet Research*. 20, e10775 (2018).
- Anisuzzaman, D., Malins, J., Friedman, P. & Attia, Z. Fine-Tuning LLMs for Specialized Use Cases. *Mayo Clinic Proceedings: Digital Health*. (2024).

11. Hidalgo, J., Alvarado, J., Botella, M., Aramendi, A., Velasco, J. & Garnica, O. HUPA-UCM Diabetes Dataset. *Data In Brief*. pp. 110559 (2024)
12. Connor, M. & O'Neill, M. Large language models in sport science & medicine: Opportunities, risks and considerations. *ArXiv Preprint ArXiv:2305.03851*.
13. Li, X., Zhao, L., Zhang, L., Wu, Z., Liu, Z., Jiang, H., Cao, C., Xu, S., Li, Y., Dai, H. & Others Artificial general intelligence for medical imaging analysis. *IEEE Reviews In Biomedical Engineering*. (2024).
14. Yang, J., Li, H. & Wei, D. The impact of ChatGPT and LLMs on medical imaging stakeholders: perspectives and use cases. *Meta-Radiology*. pp. 100007 (2023).
15. Omiye, J., Gui, H., Rezaei, S., Zou, J. & Daneshjou, R. Large language models in medicine: the potentials and pitfalls: a narrative review. *Annals Of Internal Medicine*. 177, 210–220 (2024).
16. He, K., Mao, R., Lin, Q., Ruan, Y., Lan, X., Feng, M. & Cambria, E. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *ArXiv Preprint ArXiv:2310.05694*. (2023).
17. Lawson McLean, A., Wu, Y., Lawson McLean, A. & Hristidis, V. Large language models as decision aids in neuro-oncology: a review of shared decisionmaking applications. *Journal of Cancer Research and Clinical Oncology*. 150, 139 (2024).
18. Andrew, A. Potential applications and implications of large language models in primary care. *Family Medicine And Community Health*. 12 (2024).
19. Shahab, O., El Kurdi, B., Shaukat, A., Nadkarni, G. & Soroush, A. Large language models: a primer and gastroenterology applications. *Therapeutic Advances In Gastroenterology*. 17 pp. 17562848241227031 (2024).
20. Healey, E., Tan, A., Flint, K., Ruiz, J. & Kohane, I. Leveraging Large Language Models to Analyze Continuous Glucose Monitoring Data: A Case Study. *MedRxiv*. (2024).

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License which allows users to read, copy, distribute and make derivative works for non-commercial purposes from the material, as long as the author of the original work is cited properly.