

Can Artificial Intelligence (AI) Chatbots Tools be Used Effectively for Nutritional Management in Obesity?

Abstract

Background: Artificial Intelligence (AI), particularly Chat Generative Pre-Trained Transformer (ChatGPT), has been suggested as a tool for dietary planning in different diseases. **Aim:** The study aimed to compare the energy, macro and micronutrients of the sample menu components presented by ChatGPT-4o and ChatGPT-4 for obesity with the Turkish Dietary Guidelines (TDG)-2022, evaluating their accuracy and clarity in medical nutrition management. Due to higher accuracy levels and the most preferred AI, ChatGPT-4o and ChatGPT-4 were selected for comparison. **Methods:** A comparative content analysis was conducted using ChatGPT-4o, and ChatGPT-4 to generate 1800 kcal daily diet plans for a 20-year-old female with obesity. AI models provided recommendations for dietary management, the nutrition care process, and menu planning. Three dietitians evaluated the outputs. Data were analyzed using SPSS 24.0. **Results:** ChatGPT-generated menus were inconsistent with dietary recommendations. Both ChatGPT-4o and ChatGPT-4 offered five-day menu samples with low calorie content of "1800 kcal prompt" compared to the Turkish Dietary Guidelines (TDG)-2022 ($p < 0.001$ for ChatGPT-4o). Additionally, key macronutrients, particularly fats ($p: 0.003$), carbohydrates (%), potassium, and calcium ($p < 0.05$ for all) were inadequately compared to the TDG-2022. Nutrient analysis revealed that both models underperformed in meeting recommended intakes for critical micronutrients such as calcium and iron and had an unbalanced distribution of macronutrients. **Conclusion:** ChatGPT-4o and ChatGPT-4 have limitations when used to provide accurate dietary management. While AI chatbots offer useful insights, they cannot replace expertise of dietitians in clinical planning; as a result, caution is advised when using these tools in this context.

Keywords

Artificial intelligence, diet, nutritional treatment, obesity

Introduction

Obesity is a chronic and complex disease, characterized by excessive adipose tissue that can impair health. According to the World Health Organization (WHO), in 2022, 1 in 8 people in the world were living with obesity. Obesity is associated with higher risk and earlier onset of several non-communicable diseases (NCDs), including cardiovascular diseases, type 2 diabetes, and certain cancers (WHO, 2021). Based on recent data from WHO, and other global health organizations, Türkiye has the highest obesity rate in Europe (World Obesity Federation, 2021; WHO, 2022; World Population Review, 2024). According to Eurostat, 16 years and over population in European countries overweight rate determined as 58.7% for males and 43% for females. The overweight population rates varied between 31.3% in Italy and 56.7% in Latvia for females and between 51.5% in France and 69.4% in Croatia, Malta and Slovakia for males (Eurostat, 2022). As of 2024, more than 66.8% of Türkiye's adult population is considered overweight, and around 32.1% are classified as obese. Globally, Türkiye ranks among the top four countries in terms of obesity (WHO, 2022; World Population Review, 2024).

Adequate energy and nutrient intake as well as healthy lifestyle habits (regular physical activity, adequate sleep, stress-releasing methods etc.) are essential in the treatment of obesity. Changes in these two aspects are the first steps, which often prove lifesaving, for the overweight and obese population¹. The role of dietitians in managing obesity is critical, as these professionals are uniquely positioned to provide evidence-based nutritional guidance and support for individuals struggling with weight management (Morgan-Bathke et al., 2023). Dietitians possess specialized training that equips them to address the complex interplay of dietary habits, lifestyle factors, and psychological aspects related to obesity. Their expertise is essential in developing personalized interventions aimed at promoting sustainable weight loss and improving overall health outcomes (European Federation of the Association of Dietitians, 2024).

The digital environment is becoming increasingly prevalent in the healthcare system, and Artificial Intelligence (AI) offers unique opportunities for advancement and application in many disciplines (Li et al., 2021; Car et al., 2022; Knights et al., 2023). AI in nutrition can be defined as the application of conceptual algorithms, machine learning (ML), large language models (LLM), and deep learning techniques to analyze, interpret and make informed decisions from various datasets related to nutritional data, dietary patterns and other health factors (Bayram and Ozturkcan, 2024; Theodore et al., 2024). They can be used to understand how nutrients affect

health status and analyze dietary data using machine learning algorithms LLMs like ChatGPT are a specific type of ML models that focuses on understanding and generating human-like text (Hamanjuk, 2019; Kosaraju et al., 2022; Naja et al., 2024). Furthermore, AI can be used for the management of chronic diseases as obesity. AI-based techniques can help to determine risk of obesity especially for childhood (Alghalyini, 2023). Machine learning models have demonstrated superior accuracy in predicting obesity outcomes compared to traditional statistical methods, allowing healthcare providers to effectively tailor prevention strategies (Colmenarejo, 2020). Moreover, AI-based techniques can enhance adherence to dietary treatment and follow the exercise regime. AI can offer personalized diet and exercise recommendations with offering real-time feedback which helps to increase the motivation. In traditional treatment the motivation of the patient is especially crucial to achieve successful diet therapy and AI can help to maintain the diet motivation (Chew and Achananuparp, 2022). However, there is lack of information and inconsistency about advice for the nutrition treatment from AI (Ponzo et al., 2024).

ChatGPTs are increasingly being recognized for their applicability in the management of chronic diseases, which require the processing and synthesis of large amounts of data related to diagnosis and treatment (Thirunavukarasu et al., 2023). The AI company OpenAI developed ChatGPT, a conversational interface based on GPT 3 on 20 November 2022 (GPT 4 as of 14 March 2023, for paid users only). ChatGPT utilizes deep learning and is pre-trained on a large text dataset originating from online sources such as articles, books, websites and other publicly available sources to provide human-like responses to large amounts of text-based prompts to learn the nuances of the language (Barlas et al., 2024; Bayram and Ozturkcan, 2024). Although ChatGPT is considered potentially promising in providing personalized advice on obesity treatment, its use has been associated with several challenges such as lack of medical expertise, personalization limitations, ethical concerns, etc. (Arslan, 2023; Barlas et al., 2024; Naja et al., 2024). Additionally, the AI chatbot may produce references that do not exist. It may produce 'hallucinations', where the output is meaningless relative to the 'prompts' or input provided. The confident natural language used by dialogical AI may cause users to overestimate the scientific rigor of the information presented (Au Yeung et al., 2023; Haupt and Marks, 2023). Furthermore, ChatGPT-4o builds upon the foundation of ChatGPT-4 by introducing multimodal processing, faster response times, broader accessibility, and enhanced interactive features, making it a more comprehensive and user-friendly AI model. Although AI-assisted tools in obesity management

offer an innovative approach to individualised dietary recommendations, there are limited studies in the literature on the accuracy of these recommendations and their compliance with guidelines. In particular, the comparison of the Turkish Dietary Guidelines (TDG) (2022), and AI-generated menus may fill an important gap in the usability and reliability of such technologies in the Turkish context (Turkish Government Ministry of Health, 2022). Therefore, the aim of the study was to compare the energy, macro and micronutrients of the sample menu components presented by ChatGPT-4o and ChatGPT-4 for obesity when asked in Turkish with the TDG-2022 and to evaluate their accuracy and quality/clarity in providing medical nutrition treatment and management in obesity.

Methods

Study design

In this content analysis and comparative study, we evaluated the accuracy and quality/clarity in providing nutritional management for obesity of AI models ChatGPT-4o, and ChatGPT-4. According to literature ChatGPT-4o had higher consistency rates when compares other AI chatbots. Moreover, ChatGPT is widely and most used chatbot when compare others (Ponzo et al., 2024; Qarajeh, 2023). Due to higher user rates and reach the accuracy Chat GPT-4o and ChatGPT were selected.

Assessment of AI models performance

In consideration of the distinguishing features of these dietary regimens, the research team planned a five-day sample menu with an energy intake of 1800 kcal per day.

ChatGPT-4o and ChatGPT-4 (developed by OpenAI) were used to generate AI-based diet plans on 05.09.2024. ChatGPT-4o is the latest multimodal version, capable of processing both text and images with improved reasoning abilities, while ChatGPT-4 is a text-based model with enhanced contextual understanding compared to its predecessors. Both models were accessed through OpenAI's official web interface. To minimize the randomness inherent in GPT-generated outputs, each prompt was submitted three times in separate, refreshed sessions. The variability of responses was analyzed, and the averaged data was used for nutrient analysis. Prior to the initial input, the model was refreshed, and two key sentences were provided consecutively, without any refreshing between the inputs.

Three primary domains related to nutritional management were examined: (1) dietary management, (2) nutrition care process (NCP) recommendations, and (3) menu planning. Three zero-shot prompt templates were designed for each of the aforementioned areas in order to receive suggestions. In designing the prompts for this study, the methodology proposed by Ray (2023) was followed, with the objective of obtaining responses that are accurate, relevant, and useful. The following steps were undertaken:

- (1) It is recommended that the process commence with clear and specific prompts. To ensure clarity and specificity, the prompts were constructed in a way that referenced key terms related to nutrition management and dietetics, including "food," "diet," "nutrition," and "menu."
- (2) A clear context was provided. The disease or condition is explicitly stated in all prompts to provide contextual background for the subsequent nutrition and dietary guidelines.
- (3) The use of open-ended questions was employed. All questions were designed to be open-ended, employing verbs such as "list," "provide," and "give an example."
- (4) Additionally, constraints and limitations were applied.

For example, in the "menu planning" field, constraints such as age (20 years), gender (female), and calorie amount (1800 kcal) were specified. In Turkish Subsequently, four cues were developed for each condition, as shown in Figure 1. The four different prompts was created in Turkish. For example the first prompt was “Obeziteyi tedavi etmek için önerilen diyet nedir?”. Prompts were crafted in Turkish and explicitly included instructions to align with the Turkish Dietary Guidelines (TDG-2022). No memory function was used; all necessary information was included directly in the prompts.

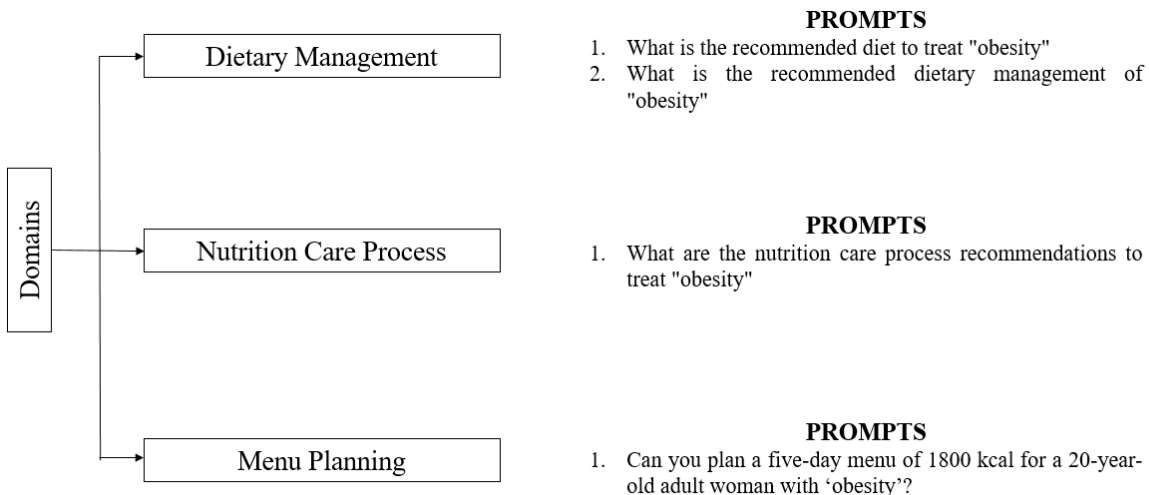


Figure 1. The domains and corresponding prompts for obesity.

The chatbot's recommendations for the domains of “dietary management” and “NCP” were compared with those outlined in the Academy of Nutrition and Dietetics' Nutrition Care Manual (NCM)[®] (Academy of Nutrition and Dietetics, 2022). Three dietitians evaluated the compliance of the chatbot output with the guidelines. The main focus of the assessment was accuracy. Dietitians indicated their responses using a Likert scale from 1 to 4, corresponding to ‘none’ (i.e. no agreement between output and guidelines), ‘some but not all’, ‘most but not all’ and ‘all’ respectively.

In qualitative research, particularly in content analysis and expert-based evaluations, traditional power calculations are not always applicable. Instead, expert reliability is often assessed using statistical measures such as the Intraclass Correlation Coefficient (ICC) or Fleiss’ Kappa, which evaluate agreement among evaluators. Literature suggests that for expert-based evaluations, a panel of 3-5 experts is generally considered sufficient (Bertens et al., 2013; Vasileiou et al., 2018). The selection of dietitians was based on their expertise and experience in the field of clinical nutrition and obesity management. The inclusion criteria required dietitians to have at least five years of professional experience in developing and evaluating dietary plans for individuals with obesity. Additionally, familiarity with international and national dietary guidelines, including the Nutrition Care Process (NCP), was mandatory. Dietitians were also required to be proficient in analyzing and interpreting dietary data using tools like BeBiS program. These criteria ensured that the evaluations were performed by highly qualified professionals capable of critically assessing the AI-generated diet plans. Furthermore, the reliability of the expert evaluations for the NCP domain was confirmed by an ICC value of 0.904 ($p < 0.001$), indicating excellent inter-rater agreement. Given the high inter-rater reliability and alignment with expert panel methodologies, the number of dietitians included in this study is methodologically sound and appropriate for expert-based content analysis.

For the NCP domain, dietitians analyzed chatbot responses separately according to the following four NCP pillars: “Nutritional assessment”, “Diagnosis”, “Intervention” and “Monitoring/Evaluation”. The “Dietary management” domain was assessed against the intervention section of the NCP guidelines. For the “Menu planning” domain, an 1800 kcal diet for a 20-year-old adult female was used in the referrals. The nutrients and their corresponding

amounts in the five-day menus created by the chatbot were calculated by entering them into the Nutrient Database (BeBiS, EBISpro for Windows, Willstaett, Germany; Turkish Version, BeBiS 9). Menus were analyzed for energy, macro- and micronutrient content and compared with the recommendations set out in the TDG-2022 (Turkish Government Ministry of Health, 2022).

Statistical analysis

Data was analyzed using SPSS 24.0. The inter-observer correlation was evaluated by the Intraclass correlation coefficient (ICC). The difference between AI models in the energy, macro and micronutrients content of the five-day sample menus of diet models was tested using Repeated Measures ANOVA. Energy, macro and micronutrient analyses of the five-days menu plans created by ChatGPT and GPT 40 were compared with RDI using one-sample-T-test. A p-value ≤ 0.05 was considered statistically significant for the statistical test.

Results

Table 1 shows the evaluation of three dietitians for the AI chatbots responses for dietary management, and NCP domains for obesity. For dietary management, for both ChatGPT-4o and ChatGPT-4 responses, 2 dietitians scored 2 and one scored 3. The main gap was that macronutrient intake (fat, carbohydrate especially), saturated fatty acids (SFA), omega 3 were not addressed. In this part, the ICC value was 1, and there was complete agreement between the observers. This result reveals that the dietitians evaluated the recommendations of ChatGPT in the same way and there were no subjective differences in the evaluations.

The scores of the dietitians according to the responses of ChatGPT-4o for nutritional assessment were 1, 2 and 1, while the scores for ChatGPT-4 were 2, 3 and 2 respectively. The deficiencies in the responses of ChatGPT-4o consisted of dietary intake assessment (partial), and popular diets offered as part of the therapy. Details such as body fat percentage, body fat (kg), muscle mass, lean body mass (kg), total body water content (kg), waist/hip ratio, neck circumference, etc. were missing when using the body impedance analysis (BIA) device. For ChatGPT-4, deficiencies were dietary intake assessment (partial), popular diets offered as part of the therapy, and biochemical parameters not being mentioned as well as drug use and anthropometric measurements not being mentioned. The assessment process is described very superficially. In this part, the ICC value was 0.82, indicating a high level of agreement among the

dietitians. This suggests that while the chatbot responses exhibited notable deficiencies, the expert raters provided consistent evaluations, reinforcing the validity of the assessment outcomes.

The diagnosis section for both ChatGPT-4o and ChatGPT-4 was quite inadequate and received 1 point from all dietitians. Sample Problem, Etiology, Signs & Symptoms (PES) statements are unspecified and incomplete. Similarly, for the intervention section, the responses of both ChatGPT-4o and ChatGPT-4 received a score of 2 from all dietitians and the gaps in both were a lack of information on macro- and micronutrients, no mention of approved medication or bariatric surgery. No mention of using plate models, food pyramids or food replicas etc. was made. There was no information about food consumption records. The ICC for the diagnosis and intervention sections was perfect, with an ICC value of 1.00, indicating complete agreement among the dietitians. This result confirms that all raters consistently identified the same deficiencies in the chatbot-generated responses, reinforcing the robustness of the evaluation process and eliminating subjective variation in expert assessment.

For the monitoring/evaluation section, all dietitians gave a score of 2 for ChatGPT-4o responses, while two dietitians gave a score of 2 and one dietitian gave a score of 3 for ChatGPT-4 responses. The gaps in the ChatGPT-4o responses were that there was no mention of reassessment of nutritional status (anthropometry), weight maintenance, self-monitoring, changes in body composition or biochemical parameters. What the dietitian should do in case of encountering problems during the reassessment process was not mentioned. Diet sustainability, changes in dietary and lifestyle habits were not mentioned. The gaps in ChatGPT-4 were that there was no mention of self-monitoring, and there was no mention of what the dietitian should do in case of problems in the reassessment process. In this part, the ICC value was 1, indicating complete agreement among the dietitians. This result demonstrates that all raters consistently identified the same shortcomings, confirming the robustness and reliability of the expert assessments.

Table 1. Evaluation of AI chatbots responses for dietary management and nutrition care process for obesity*

ChatGPT-4o	Dietitian 1	Dietitian 2	Dietitian 3	Gaps
Dietary management	2	3	2	Macronutrient intake (fat, carbohydrate especially), SFA, omega 3 are not addressed.
Nutrition care process				

Nutritional assessment	1	2	1	Dietary intake assessment (partial), popular diets presented as part of the therapy. Details for using the BIA device, body fat percentage, body fat (kg), muscle mass, lean body mass (kg), total body water content (kg), waist/hip ratio, neck circumference, etc. are missing.
Diagnosis	1	1	1	Problem, Etiology, Signs & Symptoms (PES) statements are unspecified and incomplete.
Intervention	2	2	2	Information on macro and micronutrients is missing, approved medication and bariatric surgery are not mentioned. There is no mention of plate models, food pyramids, food replicas, etc. that could be used. There is no information on food consumption records.
Monitoring/Evaluation	2	2	2	Reassessment of nutritional status (anthropometry), weight maintenance, self-monitoring, changes in body composition, biochemical parameters are not mentioned There is no mention of what the dietitian should do in case of problems experienced during the re-evaluation process. Sustainability of the diet, changes in dietary and lifestyle habits are not mentioned.
ChatGPT-4				
Dietary management	2	3	2	Macronutrient intake (fat, carbohydrate especially), SFA, omega 3 not addressed.
Nutrition care process				

Nutritional assessment	1	1	1	There is no mention of dietary intake assessment (partial), popular diets presented are presented as part of the therapy, and biochemical parameters were not mentioned. There is no mention of medication use or anthropometric measurements. The evaluation process is described very superficially.
Diagnosis	1	1	1	Problem, Etiology, Signs & Symptoms (PES) statements are unspecified and incomplete
Intervention	2	2	2	Information on macro and micronutrients is missing, physical activity, approved medication and bariatric surgery are not mentioned. There is no mention of plate models, food pyramids, food replicas, etc. that could be used. There is no information on food consumption records.
Monitoring/Evaluation	3	2	2	Self-monitoring was not mentioned There is no mention of what the dietitian should do in case of problems experienced during the re-evaluation process.

*1 = none; 2 = some but not all; 3 = most but not all; 4 = all.

Table 2 shows the mean energy, macro- and micronutrients of five-day menu samples according to AI chatbots analysis and intake percentages according to TDG-2022. Both ChatGPT-4o and ChatGPT-4 offered five-day menu samples with low calorie content compared to the "1800 kcal prompt", and the five-day menu samples mean energy of ChatGPT-4o was 985.51 ± 113.52 kcal, and of ChatGPT-4 was 1427.65 ± 563.96 kcal. When the mean of the five-day menu samples provided by ChatGPT-4o is compared with TDG-2022, the amounts of energy ($54.75 \pm 6.31\%$), sodium ($59.16 \pm 27.71\%$), potassium ($49.41 \pm 7.00\%$), calcium ($62.84 \pm 7.31\%$) and iron ($58.10 \pm 14.98\%$) are below the Recommended Dietary Allowances (RDA). According to the five-day

menu samples provided by ChatGPT-4, calcium intake was below the RDA with $62.69 \pm 19.07\%$. In addition, when nutrient analyses of ChatGPT-4 and ChatGPT-4o menu samples were compared with the Recommended Dietary Intakes (RDI), it was found that ChatGPT-4o menu plans had significantly lower energy, carbohydrate (both g and %), fiber, vitamin E, thiamin, folic acid, niacin, sodium, potassium, calcium and iron, while they had significantly higher protein (%) and fat (%) and phosphorus. When ChatGPT-4 menus were compared with RDI, it was determined that these menu plans had significantly higher protein (%), fat (%) and phosphorus. In contrast, they had significantly lower carbohydrate (%), potassium, and calcium.

Table 2. Mean energy, macro- and micronutrients of five-day menu samples according to AI chatbots analysis and intake percentages according to TDG-2022

	ChatGPT-4o	RDA %	ChatGPT-4	RDA %	p-value ¹	p-value ²	p-value ³
Energy (kcal)	985.51 $\pm$ 113.52	54.75 $\pm$ 6.31	1427.65 $\pm$ 563.96	79.31 $\pm$ 31.33	0.602	<0.01	0.214
Protein (g)	59.91 $\pm$ 4.95	108.54 $\pm$ 8.98	101.43 $\pm$ 67.99	183.75 $\pm$ 123.18	0.532	0.100	0.203
Protein (E%)	25.2 $\pm$ 2.58	148.23 $\pm$ 15.23	27.8 $\pm$ 6.83	163.52 $\pm$ 40.20	0.808	0.002	0.024
Fat (g)	40.38 $\pm$ 7.41	N.A.	56.18 $\pm$ 16.28	N.A.	0.226		
Fat (E%)	36.4 $\pm$ 4.97	161.77 $\pm$ 22.13	36.4 $\pm$ 4.72	161.77 $\pm$ 20.99	0.758	0.003	0.003
Carbohydrate (g)	92.26 $\pm$ 14.49	70.97 $\pm$ 11.15	122.79 $\pm$ 46.10	94.19 $\pm$ 35.46	0.792	0.004	0.744
Carbohydrate (E%)	38.2 $\pm$ 3.27	72.76 $\pm$ 6.23	25.8 $\pm$ 6.41	68.19 $\pm$ 12.23	0.281	<0.001	0.004
Fiber (g)	19.41 $\pm$ 1.09	77.67 $\pm$ 4.39	24.51 $\pm$ 6.07	98.06 $\pm$ 24.29	0.647	<0.001	0.867
SFA (g)	12.18 $\pm$ 2.99	N.A.	16.41 $\pm$ 9.03	N.A.	0.285		
Polyunsaturated fatty acids (g)	8.25 $\pm$ 4.35	N.A.	13.45 $\pm$ 3.69	N.A.	0.144		
Monounsaturated fatty acids (g)	17.25 $\pm$ 4.30	N.A.	22.2 $\pm$ 9.05	N.A.	0.526		
Cholesterol (mg)	170.61 $\pm$ 78.67	56.87 $\pm$ 26.23	393.00 $\pm$ 255.93	131.00 $\pm$ 85.31	0.641	0.021	0.462
Omega 3 (g)	2.55 $\pm$ 1.70	N.A.	2.57 $\pm$ 2.99	N.A.	0.057		
Omega 6 (g)	5.55 $\pm$ 2.76	N.A.	10.76 $\pm$ 2.89	N.A.	0.122		
Vitamin A (μ g)	803.66 $\pm$ 241.01	123.45 $\pm$ 37.08	1701.75 $\pm$ 1029.70	261.80 $\pm$ 158.42	0.457	0.227	0.084
Vitamin E (mg)	8.19 $\pm$ 1.47	74.45 $\pm$ 13.43	14.65 $\pm$ 4.74	133.20 $\pm$ 43.10	0.189	0.013	0.160
Vitamin K (μ g)	127.13 $\pm$ 13.86	141.25 $\pm$ 15.31	252.69 $\pm$ 31.03	280.77 $\pm$ 34.48	0.569	0.579	0.306
Vitamin C (mg)	66.49 $\pm$ 37.85	69.99 $\pm$ 39.85	172.89 $\pm$ 75.22	181.98 $\pm$ 79.18	0.191	0.168	0.082
Thiamin (mg)	0.74 $\pm$ 0.17	67.81 $\pm$ 16.09	1.07 $\pm$ 0.30	97.27 $\pm$ 27.73	0.904	0.011	0.837
Riboflavin (mg)	1.28 $\pm$ 0.27	116.54 $\pm$ 25.41	1.52 $\pm$ 0.63	138.18 $\pm$ 57.84	0.845	0.219	0.214
Vitamin B6 (mg)	1.32 $\pm$ 0.37	101.69 $\pm$ 28.95	2.28 $\pm$ 1.06	175.84 $\pm$ 81.96	0.469	0.902	0.107
Folic acid (μ g)	226.37 $\pm$ 29.27	68.59 $\pm$ 8.87	349.15 $\pm$ 161.45	105.80 $\pm$ 48.93	0.503	<0.001	0.804
Niacin (mg)	23.12 $\pm$ 2.55	192.68 $\pm$ 21.29	43.00 $\pm$ 39.18	358.40 $\pm$ 326.54	0.462	<0.001	0.152
Vitamin B5 (mg)	3.23 $\pm$ 0.91	N.A.	5.88 $\pm$ 2.67	N.A.	0.836		
Biotin (μ g)	33.79 $\pm$ 11.33	N.A.	57.51 $\pm$ 18.93	N.A.	0.948		
Vitamin B12 (μ g)	4.5 $\pm$ 1.973	113.30 $\pm$ 49.27	3.87 $\pm$ 2.55	96.85 $\pm$ 63.77	0.092	0.519	0.917
Sodium (mg)	1360.70 $\pm$ 637.40	59.16 $\pm$ 27.71	1780.27 $\pm$ 632.24	77.40 $\pm$ 27.49	0.501	0.030	0.240
Potassium (mg)	2322.56 $\pm$ 329.03	49.41 $\pm$ 7.00	3235.21 $\pm$ 530.30	77.40 $\pm$ 27.49	0.605	<0.001	0.003

Calcium (mg)	628.43 ± 73.09	62.84 ± 7.31	626.92 ± 190.70	62.69 ± 19.07	0.087	<0.001	0.012
Magnesium (mg)	278.43 ± 35.22	92.81 ± 11.74	357.86 ± 131.41	119.28 ± 43.81	0.712	0.243	0.381
Phosphorus (mg)	1127.82 ± 133.56	205.05 ± 24.28	1506.66 ± 617.09	273.93 ± 112.20	0.570	<0.001	0.026
Iron (mg)	7.84 ± 2.02	58.10 ± 14.98	12.65 ± 4.77	93.76 ± 35.36	0.740	0.003	0.713
Zinc (mg)	6.95 ± 2.80	69.50 ± 28.05	9.49 ± 3.22	94.92 ± 32.28	0.807	0.072	0.743

RDA: recommended dietary allowance, SFA: saturated fatty acids. p-value¹: Repeated Measures ANOVA analysis of ChatGPT-4 and ChatGPT-4o ; p-value²: comparison of ChatGPT-4o and RDA, one-sample-T-test, significant at p<0.05; p-value³: comparison of ChatGPT-4 and RDA, one-sample-T-test, significant at p<0.05.

Discussion

In this study, two chatbots, ChatGPT-4o and ChatGPT-4, were tasked with planning five-day menu samples of 1800 calories for a 20-year-old woman with obesity and explaining the nutritional recommendations and NCP steps suitable for this individual. When ChatGPT was asked questions focusing on the dietary management of obesity and NCP, ChatGPT-4o and 4 responses were found to have low consistency with dietary management and obesity assessment guidelines. Additionally, it is our contention that the most striking result of the study is that even though a diet plan of 1800 calories was requested, both AI chatbots planned five-day menu samples which contained calories far below the requested amount. In particular, the menu planned by ChatGPT-4o is found to meet only $54.75 \pm 6.31\%$ of the RDA.

Despite the growing interest of AI across all sectors, ChatGPT is increasingly being recognized for their applicability in diagnosis and treatment, especially in the management of chronic diseases that require the processing and synthesis of large amounts of data (Thirunavukarasu et al., 2023). However, the results of studies evaluating accuracy in the field of nutrition have shown that there are differences between the outputs of different AI models and variations in accuracy ranges, thus emphasizing the need for human supervision when using AI in the field of nutrition education and recommendations (Qarajeh et al., 2023; Bayram and Ozturkcan, 2024). Prior research has demonstrated that ChatGPT is capable of effectively assessing obesity in individuals with type 2 diabetes. However, its treatment recommendations, particularly those pertaining to nutrition and surgical interventions, remain incomplete. AI also holds promise for personalized obesity management, offering tailored advice. Nevertheless, ethical and safety concerns emerge when considering the use of AI in this context. Furthermore, although when providing dietary guidance for individuals with diabetes and metabolic syndrome ChatGPT offers valuable insights, it falls short of providing comprehensive nutrition and energy recommendations (Arslan, 2023; Barlas et al., 2024; Naja et al., 2024). Our analysis revealed several shortcomings

in both AI models when it comes to “Dietary management”. Neither model adequately addressed key aspects of macronutrient intake, particularly fat, carbohydrate, SFA, and omega-3. While both AI models partially covered the steps involved in dietary intake assessment, they failed to provide sufficient information in the diagnosis section, with none of the PES statements specified. Both models also lacked critical information regarding macro- and micronutrient content, approved medications, and bariatric surgery options. Furthermore, essential tools like plate models, food pyramids, and food replicas that could aid in patient education were completely omitted. No details were provided on food consumption records either. In the “Monitoring/Evaluation” section, the gaps between the two AI models varied, indicating inconsistencies in their approach to this critical component of nutritional management. The results highlight significant limitations in the AI models' ability to provide comprehensive dietary management, particularly in areas crucial for clinical practice. Inconsistent outputs can lead to confusion among healthcare providers and patients, undermining the trust that is essential for effective clinical decision-making (Gong et al. 2024). When patients receive conflicting advice from AI systems, they may be less likely to adhere to treatment plans, which can adversely effect their health outcomes. This inconsistency can also create a situation where patients may disregard professional medical advice in favor of AI-generated recommendations, leading to potentially harmful decisions (Kim et al., 2023). To improve AI performance, personalized approaches could be useful. By utilizing machine learning algorithms that analyze personal health data, including metabolic responses and dietary preferences, AI can generate more relevant and effective nutritional advice. This approach aligns with the concept of precision nutrition, which aims to customize dietary interventions based on individual characteristics (Sempionatto et al., 2021).

In this study, ChatGPT-4o, and 4 were prompted to provide an 1800 kcal five-day menu for obesity. Previous studies showed that AI-generated menus contained non-starchy vegetables, lean protein foods and foods with carbohydrates (CHO), but nutritional analyses showed that the energy amounts in these menus were sometimes different from those specified, and the amounts of CHO and fat were often inconsistent with dietary recommendations (Chatelan et al., 2023; Naja et al., 2024). Similar findings were observed in this study, where both ChatGPT-4o models demonstrated notable discrepancies in energy and nutrient calculations for a five-day sample menu of 1800 kcal. The energy values provided were significantly lower than expected, with means of 985.51 ± 113.52 kcal and 1427.65 ± 563.96 kcal, respectively. Additionally, the carbohydrate

percentages were lower than recommended, calculated at $38.2 \pm 3.27\%$ and $25.8 \pm 6.41\%$. Fat intake was consistent across both models, with averages of $36.4 \pm 4.97\%$ and $36.4 \pm 4.72\%$. However, key nutrients such as energy ($54.75 \pm 6.31\%$), sodium ($59.16 \pm 27.71\%$), potassium ($49.41 \pm 7.00\%$), calcium ($62.84 \pm 7.31\%$), and iron ($58.10 \pm 14.98\%$) for ChatGPT-4o fell below the RDA. While calcium intake for the second model was also below the RDA at $62.69 \pm 19.07\%$, these findings suggest that both AI models require improvements in accuracy to meet the nutritional needs set by dietary guidelines. These results highlight the importance of cautious use of AI tools in dietary planning, as they may underperform in providing adequate nutrient distributions.

Another important consideration in evaluating AI-generated diet plans is their cultural appropriateness within the Turkish context. Traditional Turkish cuisine is rich in a variety of whole grains (such as bulgur and whole wheat bread), legumes, dairy products (yogurt and white cheese), and fresh vegetables, aligning well with Mediterranean dietary principles (Turkish Government Ministry of Health, 2022). The five-day menu created by ChatGPT-4o offers a content partially suitable for Turkish cuisine, but it also has some deficiencies. The inclusion of foods belonging to traditional Turkish cuisine such as whole wheat bread, feta cheese, lentil meatballs, cacık, bulgur pilaf with chickpeas, and beans with olive oil in the menu was evaluated as a positive aspect. Furthermore, the regular inclusion of fermented milk products such as yoghurt and ayran is noteworthy in terms of compliance with Turkish eating habits. However, olives, one of the indispensable elements of Turkish breakfast, were recommended only on one day and were not included on other days. In addition, breakfast options that include Western-oriented foods such as chia seeds and peanut butter offer alternatives that are not common in Turkish cuisine. In addition, in terms of bread and cereal consumption, it was observed that traditional cereals such as bulgur were not sufficiently emphasized. When the five-day menu created by ChatGPT-4 was analyzed, it was seen that it contained some basic nutritional elements of Turkish cuisine, but it was not fully culturally appropriate. The inclusion of traditional dishes such as lentil soup, kidney bean stew, menemen, cacık and chickpea dish in the menu can be considered as a positive feature. However, it was observed that foods such as whole grain bread, bulgur, vegetable dishes with olive oil and yoghurt, which are commonly consumed in daily eating habits, were not included sufficiently on some days. Additionally, considering that breakfast is an important meal in Turkish cuisine, it is noteworthy that the recommended breakfasts were culturally unusual on some days. For example,

breakfasts containing oatmeal and strawberries show a more Western-oriented approach compared to traditional Turkish breakfasts. However, in some meals, more global foods such as quinoa salad or pineapple are offered instead of flavours specific to Turkish cuisine. This may limit the applicability and sustainability of the menu for Turkish individuals. Therefore, further optimisation on regional dietary habits is needed to better adapt the menus created by AI to the local culinary culture.

While this study provides insights into AI-generated short-term meal planning, one critical limitation is the lack of consideration for long-term adherence. Sustainable dietary changes are essential for effective obesity management; as short-term modifications often fail to result in lasting health benefits. Long-term adherence to dietary recommendations is influenced by factors such as behavioral habits, environmental constraints, social influences, and psychological motivations (Cacau et al., 2021; Jung et al., 2023), which are not adequately addressed in AI-generated menu planning. Current research suggests that sustainable dietary changes require flexibility and behavioral support mechanisms, which AI tools lack (Morgan-Bathke et al., 2023). Chatbots can generate structured meal plans, but adherence over weeks or months depends on personalization, ongoing monitoring, and motivational strategies—elements that AI does not yet integrate effectively. Moreover, behavioral reinforcement strategies, such as goal setting, self-monitoring, and dietary modifications based on individual progress, are crucial for long-term success but are missing in AI-generated meal plans (Chatelan et al., 2023). Without these adaptive elements, individuals may struggle to maintain adherence, limiting the real-world applicability of AI-driven dietary interventions. A systematic review showed that AI chatbots have demonstrated the effectiveness of health behaviour change interventions among large and diverse populations; however, it was emphasised that future studies should adopt robust randomized control trials to draw firm conclusions (Aggarwal et al., 2023). To improve the effectiveness of AI in dietary management, future research should explore how AI models can integrate long-term adherence strategies, such as adaptive dietary modifications, user-specific feedback loops, and behavioral coaching techniques. Additionally, longitudinal studies assessing real-world adherence to AI-generated dietary recommendations could provide valuable insights into whether these tools can support sustainable dietary changes beyond short-term menu planning. These findings highlight the need for dietitians to play an active role in bridging the gap between AI-generated dietary recommendations and real-world adherence strategies. While AI chatbots may offer a structured

starting point for dietary planning, their role should be complementary rather than substitutive, reinforcing the necessity of human expertise in ensuring long-term adherence and sustainable health outcomes.

The study has some limitations and strengths. First, due to the regular updating of the ChatGPT, responses to the same question may change over time, making it difficult to know exactly what information a person will receive when additional data (e.g. gender, age, food preferences and medical history) are provided (Chatelan et al., 2023; Barlas et al., 2024). Secondly, the prompts asked to the ChatGPT-4o, and ChatGTP 4 were made in Turkish. The answers may show differences when searched in other languages and this affects the generalizability. Thirdly, it is focused on short-term menu planning that does not take into account the long-term adherence and sustainability of AI-generated diet plans. Forth, the dietitians evaluating the AI-generated outputs were not blinded to the source of the responses. Awareness of the AI origin may have influenced their scoring, either through skepticism toward AI-generated content or preconceived expectations. Future research should consider implementing blinded evaluation processes to mitigate potential bias and improve the objectivity of assessments. Despite all these limitations, the strength of this study is that ChatGPT-4o, and ChatGTP 4 include the analysis of obesity-related dietary management and NCPs and sample menus for Turkish language prompts.

Conclusions

In conclusion, despite their clarity, the ChatGPT-4o, and ChatGPT-4-derived outputs in the field of obesity nutritional management presented several deficiencies. These findings suggest that ChatGPT-4o and ChatGPT-4 respond to user prompts in a 'human-like' manner. Chatbots can be helpful for healthcare professionals to create education plans for patients, a variety of diet planning, monitor the progress of the treatment, nutritional data analysis and facilitate clinical nutrition therapy. However, they cannot replace the expertise and critical judgement of dietitians. Caution should be exercised when relying on AI chatbots in clinical practice, there is a need to raise awareness of the risks involved. There is lack of the research about the accuracy of the chatbots in clinical nutrition management. Future studies should include horizontal comparisons with existing dietary guidance programs to better contextualize the performance of AI chatbots. Such comparisons could enhance the understanding of how AI-based tools perform relative to traditional, validated software used in nutritional planning.

REFERENCES

Academy of Nutrition and Dietetics (2022) Nutrition Care Manual. Available at: <https://www.nutritioncaremanual.org>.

Alghalyini B (2023) Applications of artificial intelligence in the management of childhood obesity. *Journal of Family Medicine and Primary Care*, 12(11):2558-2564.

Aggarwal A, Tam CC, Wu D, et al. (2023) Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research* 25:e40789.

Arslan S (2023) Exploring the Potential of Chat GPT in Personalized Obesity Treatment. *Annals of Biomedical Engineering* 51: 1887–8.

Au Yeung J, Kraljevic Z, Luintel A, et al. (2023) AI chatbots not yet ready for clinical use. *Frontiers in Digital Health* 2023;5:1-5.

Barlas T, Altinova AE, Akturk M, and Toruner FB (2024) Credibility of ChatGPT in the assessment of obesity in type 2 diabetes according to the guidelines. *International Journal of Obesity* 48(2):271–5.

Bayram HM, and Ozturkcan A (2024) AI showdown: info accuracy on protein quality content in foods from ChatGPT 3.5, ChatGPT 4, bard AI and bing chat. *British Food Journal* 126(9):3335–46.

Bertens LC, Broekhuizen BD, Naaktgeboren CA, et al. (2013) Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS medicine* 10(10), e1001531.

Cacau LT, Benseñor IM, Goulart AC, et al. (2021) Adherence to the planetary health diet index and obesity indicators in the Brazilian longitudinal study of adult health (ELSA-Brasil). *Nutrients* 13(11):3691.

Car LT, Poon S, Kyaw BM, et al. (2022) Digital Education for Health Professionals: An Evidence Map, Conceptual Framework, and Research Agenda. *Journal of Medical Internet Research* 24(3):1-30.

Chatelan A, Clerc A, Fonta PA (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(11):1525–31.

Chew HSJ, Achananuparp P (2022) Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. *Journal of Medical Internet Research* 24(1):e32939.

Colmenarejo G (2020) Machine learning models to predict childhood and adolescent obesity: a review. *Nutrients* 12(8):2466.

European Federation of the Association of Dietitians (2024) Definition of Dietitian. Available at: <https://www.efad.org/definition-of-a-dietitian>

Eurostat (2022) Overweight and obesity - BMI statistics. Available at: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Overweight_and_obesity - BMI statistics#Obesity by age group](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Overweight_and_obesity_-_BMI_statistics#Obesity_by_age_group)

Gong Y, Min H, Jing X, et al. (2024) Challenges and opportunities of artificial intelligence in cdss and patient safety. *Studies in Health Technology and Informatics* 316:1250-1254.

Hamaniuk VA (2019) Educational Dimension: a new journal for research on education, learning and training. *Educational Dimension* 1:1–4.

Haupt CE, and Marks M (2023) AI-Generated Medical Advice-GPT and Beyond. *JAMA* 16(329):1349–50.

Jung S, Young HA, Simmens SJ, et al. (2023) The cross-sectional association between a sustainable diet index and obesity among US adults. *Obesity* 31(7):1962-1971.

Kim CS, Samaniego C, Melo SLS, et al. (2023) Artificial intelligence (a.i.) in dental curricula: ethics and responsible integration. *Journal of Dental Education* 87(11):1570-1573.

Knights V, Kolak M, Markovikj G, and Gajdoš Kljusurić J (2023) Modeling and Optimization with Artificial Intelligence in Nutrition. *Applied Sciences* 1(13):1-17.

Kosaraju C, Chiruguri S, Mohammad N, Jetty BT, and Gali G (2022) A Model for Analysis of Diseases Based on Nutrition Deficiency Using Random Forest. In: 7th International Conference

on Communication and Electronics Systems, ICCES 2022- Proceedings. Institute of Electrical and Electronics Engineers Inc; p. 1527–31.

Li L, Novillo-Ortiz D, Azzopardi-Muscat N, and Kostkova P (2021) Digital Data Sources and Their Impact on People's Health: A Systematic Review of Systematic Reviews. *Frontiers in Public Health* 9:1-19.

Morgan-Bathke M, Baxter SD, Halliday TM, et al. (2023) Weight Management Interventions Provided by a Dietitian for Adults with Overweight or Obesity: An Evidence Analysis Center Systematic Review and Meta-Analysis. *Journal of the Academy of Nutrition and Dietetics* 23(11):1621-61.

Naja F, Taktouk M, Matbouli D, et al. (2024) Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. *European Journal of Clinical Nutrition* 1-10.

Ponzo V, Rosato R, Scigliano MC, et al. (2024) Comparison of the Accuracy, Completeness, Reproducibility, and Consistency of Different AI Chatbots in Providing Nutritional Advice: An Exploratory Study. *Journal of Clinical Medicine* 13(24):7810.

Ray PP (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 3: 121-154.

Sempionatto JR, Montiel VR, Vargas E, et al. (2021) Wearable and mobile sensors for personalized nutrition. *ACS Sensors* 6(5):1745-1760.

Qarajeh A, Tangpanithandee S, Thongprayoon C, et al. (2023) AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. *Clinics and Practice* 13:1160–72.

Theodore Armand TP, Nfor KA, Kim JI, and Kim HC (2024) Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review. *Nutrition* 16:1-24.

Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, and Ting DSW (2023) Large language models in medicine. *Nature Medicine* 29:1930–40.

Turkish Government Ministry of Health (2022) Türkiye Dietary Guideline 2022. Available at: https://hsgm.saglik.gov.tr/depo/birimler/saglikli-beslenme-ve-hareketli-hayat-2022/db/Dokumanlar/Rehberler/Turkiye_Beslenme_Rehber_TUBER_2022_min.pdf.

World Health Organization (2021) Obesity and overweight fact sheet. Available at: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

World Health Organization (2022) WHO European regional obesity report 2022. Available at: <http://apps.who.int/bookorders>.

World Obesity Federation (2021) Global Obesity Observatory 2024. Available at: <https://data.worldobesity.org/country/turkey-219>

World Population Review (2024) Obesity Rates by Country 2024. Available at: <https://worldpopulationreview.com/country-rankings/obesity-rates-by-country>

Vasileiou K, Barnett J, Thorpe S, et al. (2018) Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology* 18:1-18.