frontiers in Sustainable Food Systems

TYPE Original Research
PUBLISHED 06 October 2026
DOI 10.3389/fsufs.2026.1984101

Check for updates

OPEN ACCESS

EDITED BY
Sedat Arslan,
Bursa Uludag Universitesi, Türkiye

REVIEWED BY
Mahmut Bodur,
Ankara University, Türkiye
Marco Antonio Diaz Martinez,
Instituto Tecnologico Superior de Panuco, Mexico

*CORRESPONDENCE
Elvan Yılmaz Akyüz
elvan.yilmazakyuz@sbu.edu.tr

RECEIVED 31 August 2026
REVISED 21 September 2026
ACCEPTED 21 September 2026
PUBLISHED 06 October 2026

CITATION
Sarıyer ET, Özlü Karahan T, Ural B and Yılmaz Akyüz E (2026) AI-generated menus versus the Türkiye Dietary Guidelines: nutrient adequacy, ultra-processed food burden, cost, and carbon footprint.
Front. Sustain. Food Syst. 10:1984101.
doi: 10.3389/fsufs.2026.1984101

COPYRIGHT
© 2026 Sarıyer, Özlü Karahan, Ural and Yılmaz Akyüz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

AI-generated menus versus the Türkiye Dietary Guidelines: nutrient adequacy, ultra-processed food burden, cost, and carbon footprint

Esra Tansu Sarıyer, Tuğçe Özlü Karahan, Bedriye Ural and Elvan Yılmaz Akyüz*

Department of Nutrition and Dietetics, Hamidiye Faculty of Health Science, University of Health Sciences Türkiye, Istanbul, Türkiye

Introduction: While the use of generative AI tools in meal planning is becoming increasingly widespread, the alignment of the menus they generate with national dietary guidelines, the level of food processing, economic accessibility, and environmental impacts have not yet been sufficiently clarified. This study aimed to compare daily menus generated by ChatGPT and Gemini for different life stages and physiological conditions with the Türkiye Dietary Guidelines (TÜBER) in terms of nutrient adequacy, ultra-processed food (UPFs) content, economic cost, carbon footprint, and short-term temporal consistency.

Methods: Using ChatGPT-4o and Gemini Pro (Google), three menus were generated at 2-week intervals for eight population profiles: preschool children, school-age children, adolescents, adult women, adult men, older adults, pregnant women, and lactating women. A total of 48 AI-generated menus (24 from ChatGPT and 24 from Gemini) were compared with 8 profile-specific reference menus included in TÜBER. Energy and nutrient content were analyzed using BeBiS 9; NOVA 4 content, daily cost, and carbon footprint were calculated.

Results: Energy adequacy was 95.2%, 95.1%, and 93.4% in the ChatGPT, Gemini, and TÜBER menus, respectively. A significant difference in carbohydrate adequacy was observed among the menu sources (Holm-adjusted omnibus $p = 0.018$), with the difference specifically identified between the Gemini and TÜBER menus (Holm-adjusted pairwise $p = 0.024$, $r = 0.89$). Dietary fiber adequacy differed significantly among menu sources (Holm-adjusted omnibus $p = 0.040$), with post-hoc analysis showing higher adequacy in ChatGPT than in Gemini menus (Holm-adjusted pairwise $p = 0.024$, $r = 0.89$). The proportion of menus containing at least one NOVA 4 component was 16.7% for ChatGPT and 41.7% for Gemini. The median daily costs for ChatGPT, Gemini, and TÜBER menus were 485.6, 449.8, and 405.2 ₺, respectively; the carbon footprints were 6.4, 7.3, and 6.7 kg CO_2-eq/day, respectively, and the analyses provided insufficient evidence of differences among menu sources for either cost or carbon footprint.

Discussion: In the analysis of daily menus, although generative AI models generally produced energy values close to profile-specific references, limitations were observed in terms of nutritional balance, UPFs content, and output consistency. In nutrition planning, the dietitian must be the primary decision-maker, and AI outputs must be validated and individualized in accordance with evidence-based national guidelines such as TÜBER.

10.3389/fsufs.2026.1984101

KEYWORDS
artificial intelligence, carbon footprint, economic accessibility, nutrition planning, Türkiye Dietary Guidelines, ultra-processed foods

1 Introduction

Nutrition practices in public health have a broad perspective that addresses ecological, social, cultural, and economic dimensions together. In this context, sustainable healthy eating aims to support individuals' health and wellbeing while minimizing environmental impacts; at the same time, it refers to dietary patterns that are accessible, affordable, safe, fair, and culturally acceptable. Nutritional recommendations are an important tool in promoting public health at the regional and national levels. In this context, food-based dietary guidelines aim to transform the extensive evidence base on the relationships between foods, dietary patterns, and health into culturally appropriate, feasible, and understandable recommendations. The aforementioned guidelines aim to guide consumer behavior and provide a scientific basis for national food, nutrition, and health policies and programs (Herforth et al., 2019; Springmann et al., 2020; Erkkola et al., 2026). In Türkiye, based on existing nutritional issues and needs, the Türkiye Dietary Guidelines (TÜBER) has been prepared using national scientific evidence (Republic of Türkiye Ministry of Health, 2022). Sustainable nutrition, a central theme in global nutrition policies in recent years, emphasizes economically accessible and culturally acceptable dietary patterns that ensure nutritional sufficiency while minimizing environmental impact (Food and Agriculture Organization of the United Nations and World Health Organization, 2019). In this context, TÜBER has adopted the NOVA classification and has made reducing the consumption of UPFs one of its top priorities; it encourages the consumption of seasonal and locally sourced foods and offers environmentally conscious recommendations regarding the frequency of consumption of animal-based foods (Republic of Türkiye Ministry of Health, 2022). Current studies highlight the negative health effects of UPFs and demonstrate that these foods constitute a significant risk area for public health. Therefore, it is critically important to include restrictions and warnings regarding the consumption of UPFs in national nutrition guidelines (Anastasiou et al., 2022; U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2026). The NOVA classification system evaluates foods based on their level of processing and defines UPFs as formulations that are produced primarily from food-derived ingredients, additives, and industrial processes and that typically contain little whole or minimally processed food content (Monteiro et al., 2017; Petrus et al., 2021). TÜBER's inclusion of the NOVA classification in its guidelines emphasizes that dietary quality should be assessed not only by nutrient content but also by the level of food processing. However, it is observed that the nutritional patterns of Turkish society are increasingly approaching Western-style diets. According to the Türkiye Nutrition and Health Survey, the average daily energy intake is 1912.56 ± 736.36 kcal, of which 605.30 ± 422.69 kcal (30.64%) comes from UPFs (Republic of Türkiye Ministry of Health, 2019). A recent analysis of products available on the Turkish market found that 61.6% of the 1,365 UPFs across 13 categories have an unhealthy nutritional profile; milk and dairy products, sauces, frozen meals, processed meat products, and salty and sugary snacks were reported to fall predominantly into the unhealthy category (Şimşek et al., 2025).

The integration of artificial intelligence (AI) into various aspects of daily life has brought about significant transformations in many sectors, including healthcare, education, and nutrition. The increasing use of AI-powered applications is also raising interest in evaluating the effectiveness, reliability, and potential limitations of these tools (Hee Lee and Yoon, 2021; Alowais et al., 2023). Simultaneously with ongoing public health challenges, large language models such as GPT (OpenAI) and Gemini (Google) have become readily accessible, informal resources for health and nutrition counseling. While traditional nutrition and diet planning relies on assessments by health professionals and evidence-based guidelines, recent technological advancements have brought to the forefront the use of chatbots capable of generating individualized diet plans tailored to specific energy requirements and health goals. While these AI-powered tools are promising, significant questions remain regarding the accuracy, nutritional quality, level of individualization, and clinical safety of the diet plans they generate (Ponzo et al., 2024a; Kaya Kaçar et al., 2025). A study evaluated the alignment of ChatGPT's recommendations for the management of diabetes and metabolic syndrome with the guidelines of the Academy of Nutrition and Dietetics; reported that some recommendations could be potentially harmful or fail to adequately account for individual clinical complexities (Naja et al., 2024). Similarly, a recent study compared the quality of diet plans created for different gender and energy requirement profiles using standardized prompts from Gemini, Microsoft Copilot, and ChatGPT 4.0; it showed significant variability between models and within the outputs of the same model (Kaya Kaçar et al., 2025). A recent systematic review on generative AI in nutritional sciences found high methodological heterogeneity and a lack of reproducibility assessments in studies analyzing LLM-based dietary recommendations and recommended the GenAI in Nutritional Sciences (GAINS) reporting framework (Hieronimus et al., 2025) to improve rigor, transparency, and comparability in future research. Notably, the focus of the ongoing investigation contrasts with that of related earlier analyzes. Kaya Kaçar et al. (2025) evaluated the nutritional quality and the caloric accuracy of the weight-loss diet plans generated by various chatbots according to specified energy levels and gender demographics. Bayram et al. (2026) examined the performance of various LLMs in classifying foods based on the NOVA classification system. Conversely, the current research compares AI-generated menus with profile-specific reference menus developed from the TÜBER in a common framework that includes nutrient sufficiency, NOVA 4 food composition, economic expenditure, carbon emissions, and short-term output fluctuations.

The novelty of this study is that it provides a comprehensive assessment of AI-generated menus against a national dietary guideline reference across a range of outcomes, including nutrition, food processing, economics, environment, and short-term variability, all within a single comparative framework. Accordingly, the aim of this study is to systematically assess the alignment of AI-generated meal plans with TÜBER recommendations. The study aims to examine AI outputs not only in terms of energy and nutrient adequacy but also in terms of compliance with recommended intake levels for food groups, ultra-processed food load, carbon footprint, and economic

accessibility. Thus, the study aims to identify the potential strengths and limitations of AI-based dietary recommendations in terms of public health nutrition, sustainable nutrition, and compliance with national guidelines.

2 Materials and methods

2.1 Study design

This study is a comparative, iterative, output-based menu modeling study in which daily menus generated by generative artificial intelligence are compared with reference menus based on the national nutrition guideline. Instead of collecting food consumption data from individuals, the menu modeling approach models daily menus for population profiles created based on life stage and physiological status categories defined in TÜBER 2022; the nutritional, food processing level, environmental, and economic characteristics of these menus are then evaluated.

In the study, daily menus representing eight different population profiles were generated using two widely used AI models, GPT-4o (OpenAI) and Gemini Pro (Google), and compared with the profile-specific reference menus included in TÜBER 2022, Appendix 5. To assess short-term variability in the AI models' outputs, menu generation for each population profile was repeated three times at 2-week intervals. Based on previous nutrition-focused studies (Ponzo et al., 2024b; Ozlu Karahan and Kenger, 2025) that examined the reproducibility or temporal reliability of nutrition results generated by artificial intelligence through three-replication assessments, we conducted three consecutive replication studies. This resulted in 48 menus generated by AI (24 by ChatGPT and 24 by Gemini) and 8 reference menus from TÜBER specific for the profile, thus a total of 56 menus for the comparison. The menus were evaluated in terms of profile-specific energy and nutrient adequacy, NOVA 4 content, daily economic cost, and carbon footprint. Additionally, the short-term consistency of repeated outputs from GPT-4o and Gemini Pro, was examined using the coefficients of variation for the outcome variables. Therefore, the intended purpose was to identify short-run output fluctuations during the study period, rather than a full description of all possible menus generated by the models for the future generations.

2.2 Reference menus (TÜBER 2022)

The sample daily menus included in TÜBER 2022, Appendix 5, were used as a benchmark in this study. These menus were prepared in accordance with the energy requirements and food group recommendations established for different life stages and physiological conditions and consist of meal patterns that reflect Türkiye's culinary culture. Eight profiles representing the different life stage and physiological condition categories defined in TÜBER were included in the study: preschool-aged children, school-aged children, adolescents, adult women, adult men, older adults, pregnant women, and breastfeeding women (Republic of Türkiye Ministry of Health, 2022). For each profile, the relevant reference category in TÜBER was identified, and standardized profile definitions, based on that same category, were created to be fed into the artificial intelligence models. This allowed for a comparison between the menus generated by the artificial intelligence and the sample TÜBER menus within the same life stage or physiological condition framework. Appendix 5 provides a model daily menu for each selected reference category, with one TÜBER menus per demographic profile serving as the guideline-based standard; consequently, these menus are sample reference menus, not consistent examples of guideline-adherent menus. The eight profiles and their corresponding TÜBER reference groups are summarized in Table 1.

2.3 AI menus production

For the menus generated by artificial intelligence, two widely used generative AI models, GPT-4o (OpenAI) and Gemini Pro (Google), were used. The model selection was made considering the global and Turkish prevalence of generative AI tools (Artificial Intelligence Policies Association (AIPA), 2025; Liu and Wang, 2026).

The prompts presented to the artificial intelligence models were prepared in Turkish. The original Turkish-language prompts used for all eight population profiles are provided in Supplementary Table S1. The basic structure of the prompt remained constant across all population profiles; the model was asked to create a daily nutrition plan for the defined individual and specify each food item in grams or in portion sizes (household measurements). Individual definitions were created according to the life stage and physiological status categories in TÜBER 2022 and were structured to include profile-specific daily life characteristics related to age, gender, and living status in Türkiye.

TABLE 1 Population profiles and corresponding TÜBER reference categories used for AI-generated menus.

Population profile	TÜBER reference category	Reference age/ physiological status	Profile specified in AI prompt
Preschool child	Male child	4 years	4-year-old boy
School-age child	Female child	10 years	10-year-old girl
Adolescent	Male adolescent	16 years	16-year-old boy
Adult woman	Adult woman	19–50 years	42-year-old woman
Adult man	Adult man	19–50 years	36-year-old man
Older adult	Older man	≥65 years	68-year-old man
Pregnant woman	Pregnant woman	Third trimester	25-year-old woman, 8 months pregnant
Lactating woman	Lactating woman	0–6 months postpartum	30-year-old woman, 3 months postpartum

TÜBER, Türkiye Dietary Guidelines.

For example, the prompt used for the preschool child profile is as follows:

> "*You are a nutrition assistant. Create a daily meal plan for the person described below. Specify each food in grams or household portion sizes. Person description: A 4-year-old boy living in Türkiye; eats breakfast at home and attends preschool.*"

No energy, macronutrient, or micronutrient targets were specified in the prompt. Thus, the models were allowed to determine the menu components and portion sizes themselves, and then the extent to which the resulting menus met the profile-specific TÜBER reference values was independently evaluated. Both ChatGPT-4o and Gemini Pro were accessed through their standard web-based conversational interfaces rather than through an application programming interface (API). Therefore, no user-defined sampling parameters, such as temperature, top-p, or random seed, were specified.

To assess the output variability and short-term temporal consistency that may arise from the generative AI models, menu generation was performed three times at 2-week intervals for each population profile. The prompt language was the same for each demographic profile across the three generations and did not include any date or timestamp information within the prompts. Each generation was done in a separate independent chat session, to avoid any effect from the previous conversational context. The three sequential generations were used as repeated snapshots of model output over the duration of the study and not intended to represent the entire range of possible outputs from each generative AI model. It was stated that a 2-week period would ensure consistency in the evaluation by anticipating short-term fluctuations and the risk of change in the responses due to the frequent updating of the AI tools (Bayram and Ozturkcan, 2024). A total of 48 menus generated by AI were obtained, including 24 menus from ChatGPT and 24 from Gemini, across 8 demographic profiles (8 profiles x 3 iterations x 2 models). These 48 AI-generated menus were compared with 8 reference menus of TÜBER adapted to specific profiles. The repeated runs were also used to assess the short-term variability of the models in terms of energy, nutrient content, economic cost, and carbon footprint. All menu generations were conducted between June and August 2026.

2.4 Nutrient analysis

The energy and nutrient content of all daily menus from ChatGPT, Gemini, and TÜBER were analyzed using BeBiS (Nutrition Information System, version 9). The AI-generated menus also included the food quantities in grams as well as in household portion sizes. Therefore, the gram amounts as listed were used directly in the nutrient analysis without needing to convert from household measures to grams. The household portion measurements for the TÜBER menus were adjusted by using the standardized portion sizes and amounts in TÜBER Appendix 2 ("Standard Portion Measures and Amounts of Foods According to Food Groups for Türkiye"). The two researchers did the conversions and cross checked each other's entries before the nutritional analysis.

The amounts of energy (kcal), protein (g), carbohydrates (g), total fat (g), dietary fiber (g), calcium (mg), iron (mg), folate (μg), and vitamin B12 (μg) were calculated for each menu. The obtained values were compared with the relevant energy and nutrient reference values defined in TÜBER 2022 for each population profile.

To evaluate the performance of the menus against the reference requirements for different age groups and physiological conditions on a common scale, a reference fulfillment percentage was calculated for each nutrient. This value was obtained by dividing the amount of energy or nutrient provided by the menu by the TÜBER reference value specific to the relevant profile and multiplying the result by 100. Thus, menus for different population profiles were compared based on their levels of compliance with their profile-specific reference values, rather than on absolute amounts.

In the profile-based visual assessments, reference compliance levels were operationally classified as below the < 90% reference level, within the 90%–110% reference range, and above the >110% reference level. In the profile-based visual assessments, reference compliance levels were operationally classified as below 90% of the reference value, within the 90%–110% reference range, or above 110% of the reference value. The ± 10% interval was selected as a pragmatic descriptive margin to distinguish values that were broadly consistent with the reference from those showing a noticeable deviation. This approach was informed by previous nutritional studies that used a 90%–110% range when evaluating intake relative to reference or requirement values (Herrera-Cuenca et al., 2023; Ostrowska et al., 2025). However, this interval was defined for descriptive purposes in the present study and should not be interpreted as a universally established threshold for nutritional adequacy or safety. In particular, values above 110% do not necessarily indicate excessive or harmful intake, just as values below 90% do not automatically indicate nutritional inadequacy. The interpretation of these values depends on the nutrient being assessed, the type and purpose of the reference value, and whether a tolerable upper intake level has been established. Therefore, these categories were used to describe deviations from the reference values rather than to make clinical judgments about nutrient adequacy or safety.

2.5 NOVA classification

The foods listed in the ChatGPT, Gemini, and TÜBER menus were classified by the researchers according to their processing levels using the NOVA food classification system. The AI models were not used to assign NOVA categories. Developed by the University of São Paulo in Brazil, the NOVA classification is recognized by the Food and Agriculture Organization of the United Nations (FAO) as a valid tool for public health research and policy (Sadler et al., 2021). The NOVA system classifies foods into four groups based on the nature, extent, and purpose of industrial processing: unprocessed or minimally processed foods (NOVA 1), processed culinary ingredients (NOVA 2), processed foods (NOVA 3), and UPFs (NOVA 4).

The foods included in each daily menu were classified according to the NOVA groups, and the analyses focused specifically on UPFs classified as NOVA 4. Two indicators were used to assess the UPFs burden: (i) whether the daily menu included at least one NOVA 4 food item and (ii) the percentage of total daily menu energy contributed by NOVA 4 foods. The energy contribution of NOVA 4 foods was calculated by dividing the total energy provided by foods classified as NOVA 4 in the relevant menu by the menu's total energy content and multiplying the result by 100. The distribution of the energy contribution from NOVA 4 foods was assessed based on menus containing at least one NOVA 4 food. The NOVA classification was carried out by two researchers and subsequently reviewed by two additional researchers, all of whom are among the authors of the study. The two primary researchers independently

performed the NOVA classification. There were no differences in their evaluations. The categorizations were later reviewed by two additional researchers. A formal inter-rater reliability coefficient was not calculated. Researchers were aware of the source of the menu (ChatGPT, Gemini or TÜBER) during the classification and rating process, so the NOVA classification was not blindly performed considering the source of the menu.

2.6 Carbon footprint estimation

Since there is no comprehensive and standardized national database containing life-cycle-based carbon footprint values for foods specific to Türkiye, the SU-EATABLE LIFE Environmental Impact Database (SEL) was used as the basis for carbon footprint calculations (Petersson et al., 2021). The SEL database provides emission factors at the food item and food group levels by harmonizing carbon footprint values obtained from various life cycle assessment studies under standardized system boundaries and functional units. For foods for which no appropriate entry was found in the database, life cycle-based emission values reported by Poore and Nemecek (2018) were used. Therefore, the resulting carbon footprint estimates should be regarded as indicative rather than a definitive assessment of the environmental impact of menus in Türkiye. The emission factors used are not specific to food production and supply chains in Türkiye.

For carbon footprint calculations, the gram quantities reported directly in the AI-generated menus were used. For TÜBER menus, household portion measures were converted to gram amounts using the standardized portion sizes and amounts provided in TÜBER Appendix 2 ("Standard Portion Measures and Amounts of Foods According to Food Groups for Türkiye").

Each food item listed in the menus has been matched with the most appropriate food or food group in the data source. The quantities of the foods in the menu have been converted from grams to kilograms and multiplied by the emission factor for greenhouse gases per unit quantity of the relevant food. The carbon footprint of each food item has been calculated as follows:

$$\text{Carbon footprint of food (kg } CO_2\text{-eq)} = \text{Amount of food on the menu (kg)} \times \text{Emission factor (kg } CO_2\text{-eq/kg of food)}$$

The total carbon footprint of the menu was calculated by summing the calculated emission values of all foods included in the one-day menu, and the results are expressed in kg CO_2-eq/day.

2.7 Economic cost analysis

The economic cost of the ChatGPT, Gemini, and TÜBER menus was calculated based on the portion sizes of all foods included in the daily menus. Food prices were primarily determined using consumer price statistics from the Turkish Statistical Institute (TÜİK); for products not included in official price data or for which current price information was unavailable, prices were obtained from the Market Price[1] platform developed by Scientific and Technological Research Council of Türkiye Informatics and Information Security Research Center Product prices from chain supermarkets listed on the platform were recorded during the specified period (June–August 2026), and the average price for each food item during this period was calculated (Market Fiyatı, 2025). The pricing interval was chosen as June to August so that it matched the time when the AI menus were created. Food prices in Türkiye stayed under pressure from persistent inflation, but food prices fluctuated relatively little from month to month over the study period, relative to the larger swings seen in some earlier months of 2026. The main goal was to assess menu sources in a common price framework, not to observe changes in purchasing power over time, so the average nominal price for each food item during this period was used uniformly across all menu sources. No other adjustment for inflation was made.

For AI-generated menus, the gram quantities reported directly by the models were used for cost calculations wiithout conversion from household portion measures. For TÜBER menus, household portion measures were converted to gram or milliliter equivalents using the standardized portion sizes and amounts provided in TÜBER Appendix 2 ("Standard Portion Measures and Amounts of Foods According to Food Groups for Türkiye").

The cost of each ingredient was calculated by dividing the amount used in the menu by the product's standard unit of sale on the market. Portions expressed by the AI in household measurements (such as a glass of water, a tablespoon, a slice, or a piece) were converted to standard gram or milliliter equivalents before the cost calculation. The cost of the ingredients was calculated using the following formula:

$$\text{Food cost (₺)} = [\text{Amount used in the menu / Sales unit quantity of the product}] \times \text{Sales unit price}$$

The total daily menu cost (₺/day) was calculated by summing the costs of all food items included in the daily menu. The daily costs, calculated in Turkish lira, were also converted to euros using the exchange rate corresponding to the period during which pricing was determined and expressed as €/day. The same pricing and unit conversion approach was applied to all menu sources to ensure the comparability of economic costs.

2.8 Statistical analysis

IBM SPSS Statistics for Windows, Version 30.0 (IBM Corp., Armonk, NY, USA) was used for the statistical analysis of the data (George and Mallery, 2010). Menus generated by ChatGPT and Gemini, as well as profile-specific sample menus from TÜBER, were evaluated in terms of nutrient content. The nutrient content of each menu was standardized according to the TÜBER reference values corresponding to the relevant population profile. The percentage of nutrient reference fulfillment was calculated using the formula: (nutrient content of the relevant menu/profile-specific TÜBER reference value) × 100. In the profile-specific visual evaluation, reference fulfillment levels were operationally classified as <90% "below the reference range", 90%–110% "within the reference range," and >110% "above the reference range" for descriptive visualization purposes, as defined in Section 2.4. In the ultra-processed food (NOVA 4) evaluation, the frequency of menus containing at least one NOVA 4 component was determined as both a number and a percentage. In menus containing NOVA 4, the contribution of energy from ultra-processed foods to the total menu energy was calculated as a percentage, and these values were summarized using the median (Q1–Q3). The results for daily menu cost and carbon footprint are presented as the median

1 marketfiyati.org.tr/

(Q1–Q3). To assess the short-term temporal variability of ChatGPT and Gemini outputs, the coefficient of variation (CV) was calculated for each population profile based on three menu generations conducted at 2-week intervals. CV was determined using the formula (standard deviation/mean) × 100; the distribution of CV values for the eight profiles for each outcome variable is summarized as the median (Q1–Q3). Lower CV values indicate higher temporal consistency, while higher CV values indicate greater variability between runs. Since there were no replicate runs, TÜBER menus were not included in the CV analysis. When making inferential comparisons among the three menu sources, the median of three replicate runs for each population profile was first calculated for ChatGPT and Gemini, thereby obtaining observations at the eight-profile level for each model. For TÜBER, a single reference menu corresponding to each of the eight population profiles was used. Differences among the three menu sources were assessed using the exact Friedman test for paired samples, and Kendall's W was reported as the effect size measure. To control for the probability of Type I errors arising from multiple tests in the nutrient adequacy results, Holm's correction was applied to the p-values. For the two secondary inferential outcomes, daily menu cost and carbon footprint, Holm's correction was additionally applied to control for multiplicity within this outcome family. Following the significant omnibus tests, pairwise comparisons were performed using the exact two-sided Wilcoxon signed-rank test, and the Holm correction was also applied to these comparisons. For post-hoc Wilcoxon signed-rank comparisons, effect size was calculated as $r = Z/\sqrt{N}$, where Z represents the standardized Wilcoxon test statistic and N represents the number of non-zero paired observations. Given that the inferential analyses were based on only eight matched population profiles, these analyses were considered exploratory rather than confirmatory. Accordingly, statistical significance or non-significance was interpreted cautiously, and non-significant findings were not considered evidence of equivalence between menu sources. The frequency of menus containing NOVA 4, the energy contribution from NOVA 4 foods, and short-term temporal variability assessed by the coefficient of variation were analyzed descriptively.

3 Results

In this study, a total of 8 population profiles representing different life stages and physiological states were evaluated. These profiles consist of preschool children, school-age children, adolescents, adult women, adult men, older adults, pregnant women, and breastfeeding women (Table 1). Each profile was matched with the corresponding reference category in TÜBER; in the AI prompts, more specific characteristics related to age and physiological state were defined.

According to profile-specific TÜBER reference values, energy adequacy in the ChatGPT, Gemini, and TÜBER menus was generally found to be close to the reference value, with median values determined to be 95.2% (Q1–Q3: 92.0%–105.8%), 95.1% (87.9%–109.4%), and 93.4% (84.3%–112.5%), respectively (Table 2). However, the analysis did not provide sufficient evidence of differences in energy adequacy among the menu sources (Holm-adjusted $p = 0.967$; Kendall's W = 0.016), with Kendall's W indicating a negligible effect size. This non-significant finding should not be interpreted as evidence of statistical equivalence among the menu sources. While carbohydrate adequacy was below 90% of the reference value in all menu sources, protein and total fat adequacy exceeded 110% of the reference value. A significant difference in carbohydrate adequacy was observed among the menu sources (Holm-adjusted omnibus $p = 0.018$; Kendall's W = 0.672), and post-hoc analysis showed that the difference was between the Gemini and TÜBER menus (Holm-adjusted pairwise $p = 0.024$, $r = 0.89$). Dietary fiber adequacy also differed significantly among menu sources (Holm-adjusted omnibus $p = 0.040$; Kendall's W = 0.609); post-hoc analysis showed higher dietary fiber adequacy in ChatGPT than in Gemini menus (Holm-adjusted pairwise $p = 0.024$, $r = 0.89$). Vitamin B12 adequacy was substantially above the reference value in all menu sources, with median values of 313.6% (241.6%–568.1%) in ChatGPT, 249.2% (227.5%–352.1%) in Gemini, and 300.0% (116.6%–355.4%) in TÜBER menus; however, the difference between the groups was not statistically significant ($p = 0.570$) (Table 2).

When nutrient reference levels were examined by profile, a variable pattern was observed across menu sources (Figure 1). In ChatGPT menus, a large proportion of nutrients specific to

TABLE 2 Nutrient adequacy of ChatGPT-, Gemini-, and TÜBER-based menus relative to profile-specific reference values.

Nutrient	ChatGPT, % of reference	Gemini, % of reference	TÜBER menus, % of reference	p[a]	Kendall's W
Energy	95.2 (92.0–105.8)	95.1 (87.9–109.4)	93.4 (84.3–112.5)	0.967	0.016
Protein	225.0 (162.6–293.8)	202.0 (151.6–261.2)	169.4 (148.1–220.3)	0.480	0.297
Carbohydrate	55.5 (52.1–62.6)	42.9 (39.7–62.9)	81.5 (69.6–86.6)	0.018[b]	0.672
Total fat	151.5 (137.7–170.1)	164.8 (141.0–199.9)	130.0 (105.6–144.6)	0.126	0.484
Dietary fiber	135.2 (121.1–147.6)	107.3 (88.5–123.9)	135.5 (112.0–143.4)	0.040[c]	0.609
Calcium	123.3 (104.0–127.2)	103.6 (87.2–113.9)	97.2 (83.4–117.1)	0.180	0.438
Iron	169.4 (117.6–202.8)	141.2 (108.9–191.1)	138.0 (121.5–158.3)	0.480	0.250
Folate	116.9 (111.7–134.6)	92.8 (73.3–126.2)	101.9 (85.5–117.2)	0.180	0.438
Vitamin B12	313.6 (241.6–568.1)	249.2 (227.5–352.1)	300.0 (116.6–355.4)	0.570	0.188

Values are presented as median (Q1–Q3). For ChatGPT and Gemini, the median of three repeated menu generations was first calculated for each population profile. TÜBER values represent the corresponding single reference menu for each of the eight profiles. TÜBER, Türkiye Dietary Guidelines. [a]p values were obtained from Friedman tests and adjusted for multiple testing across nutrient outcomes using the Holm method. [b]Post-hoc pairwise comparison: Gemini vs TÜBER, Holm-adjusted p =0.024, $r = 0.89$. [c]Post-hoc pairwise comparison: ChatGPT vs Gemini, Holm-adjusted $p = 0.024$, $r = 0.89$.

10.3389/fsufs.2026.1984101

the adolescent profile were above the reference values, while some nutrients in the pregnant women's profile remained below the reference levels. A more heterogeneous distribution was observed in Gemini menus, and in the elderly profile in particular, multiple nutrients were found to be below the reference levels. In TÜBER menus as well, uniform compliance with reference levels was not observed across all nutrients; while some nutrients fell within the reference range, others were either below or above it (Figure 1).

The economic cost and carbon footprint metrics for the menus are presented in Table 3. The median daily menu cost was 485.6 ₺ for ChatGPT (395.4–525.2), 449.8 ₺ for Gemini (384.5–559.6), and 405.2 ₺ for TÜBER menus (288.8–490.4). The median carbon footprint for the ChatGPT, Gemini, and TÜBER menus was 6.4 (5.3–7.9), 7.3 (5.1–8.9), and 6.7 (5.3–7.5) kg CO_2-eq/day, respectively. Descriptively, the median daily cost was higher in the ChatGPT menus, whereas the median carbon footprint was higher in the Gemini menus. The analyses did not provide sufficient evidence of differences among menu sources in daily menu cost (Holm-adjusted $p = 0.472$; Kendall's W = 0.141) or carbon footprint (Holm-adjusted $p = 0.472$; Kendall's W = 0.203). Given the limited number of profile-level observations, these findings should not be interpreted as evidence of equivalence between menu sources (Table 3). Inferential testing was not repeated for euro-denominated costs because these values were direct currency conversions of the costs expressed in Turkish lira and therefore did not constitute an independent outcome.

In the profile-specific analysis, it was observed that both the daily menu cost and the carbon footprint varied according to both the menu source and the population profile (Figure 2). The highest values for daily cost were observed particularly in the adolescent, pregnant, and breastfeeding women profiles, while the ranking of menu sources differed across profiles. For example, the Gemini menu had the highest cost in the breastfeeding women profile, while the ChatGPT menu had a higher cost in the adult women profile. Similarly, no single menu source consistently exhibited a higher or lower carbon footprint across all profiles. The most notably high carbon footprint was observed in the Gemini menu for the adolescent profile, while the ChatGPT menu showed a lower carbon footprint in the preschool-aged child profile (Figure 2).

The frequency of UPFs on menus and their contribution to total energy intake were also evaluated. The proportion of menus containing at least one NOVA 4 component was found to be 41.7% (10/24) in Gemini and 16.7% (4/24) in ChatGPT, while no NOVA 4 components were detected in any of the TÜBER menus (0/8) (Figure 3). Among the menus containing NOVA 4, the median contribution of these foods to total energy was determined to be 2.93% (2.6–3.1) in ChatGPT and 6.43% (3.9–9.6) in Gemini. It was observed that the energy contribution from NOVA 4 showed a wider distribution in the Gemini menus, and the highest individual value was observed in this group (Figure 3). Individual NOVA 4 energy-contribution values for all menus containing at least one NOVA 4 component are provided in Supplementary Table S2.

Short-term temporal variability was assessed using the CV, which reflects the difference among the three production runs (Figure 4). The median CV for energy was 10.3% in ChatGPT (Q1–Q3: 9.3%–13.2%) and 13.5% in Gemini (5.9%–16.6%), with relatively limited variability observed in both models. In Gemini, the median CV values for protein [20.4% (15.6%–28.3%)], carbohydrates [22.9% (19.0%–26.3%)], iron [15.6% (10.0%–30.4%)], and folate [16.6% (10.7%–23.1%)] were higher than those of ChatGPT. In contrast, the most pronounced variability was observed in vitamin B12, with the median CV reaching 32.4% (Q1–Q3: 23.7%–63.8%) in ChatGPT; the corresponding value in Gemini was 22.7% (15.3%–40.2%). The profile-specific vitamin B12 CV values for both models are presented in Supplementary Table S3. The median CV for daily menu cost was 12.9% for ChatGPT and 17.6% for Gemini, while for carbon footprint, these values were 16.7% and 12.8%, respectively (Figure 4).

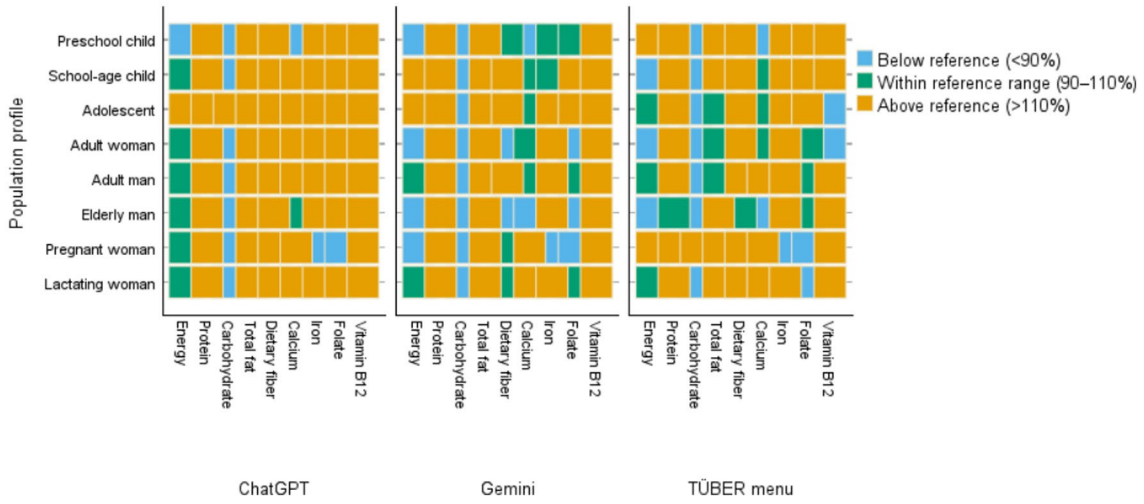


FIGURE 1
Profile-specific nutrient reference attainment of ChatGPT-, Gemini-, and TÜBER-based menus. Nutrient values were expressed as percentages of the corresponding profile-specific TÜBER reference values. For ChatGPT and Gemini, the median of three repeated menu generations was used for each population profile. Reference attainment was grouped functionally into < 90%, 90%–110%, and >110% of the relevant profile-specific reference value to enable descriptive visualization; these groups do not correspond to thresholds of clinical adequacy or safety. TÜBER, Türkiye Dietary Guidelines.

TABLE 3 Economic cost, and carbon footprint of ChatGPT-, Gemini-, and TÜBER-based menus.

Outcome	ChatGPT	Gemini	TÜBER menus	p^a	Kendall's W
Economic indicators					
Daily menu cost, ₺/day	485.6 (395.4-525.2)	449.8 (384.5-559.6)	405.2 (288.8-490.4)	0.472	0.141
Daily menu cost, €/day	8.8 (7.2-9.6)	8.2 (6.9-10.2)	7.4 (5.2-8.9)	–	–
Environmental indicators					
Carbon footprint, kg CO_2-eq/day	6.4 (5.3-7.9)	7.3 (5.1-8.9)	6.7 (5.3-7.5)	0.472	0.203

Values are presented as median (Q1–Q3). For ChatGPT and Gemini, the median of three repeated menu generations was first calculated for each population profile. TÜBER values represent the corresponding single reference menu for each of the eight profiles. [a]Exact *p* values were obtained from Friedman tests and adjusted for multiple testing across the two secondary inferential outcomes (daily menu cost and carbon footprint) using Holm's method. Inferential testing was not repeated for euro-denominated costs because these values were direct currency conversions of costs expressed in Turkish lira. TÜBER, Türkiye Dietary Guidelines; CO_2-eq, carbon dioxide equivalent.

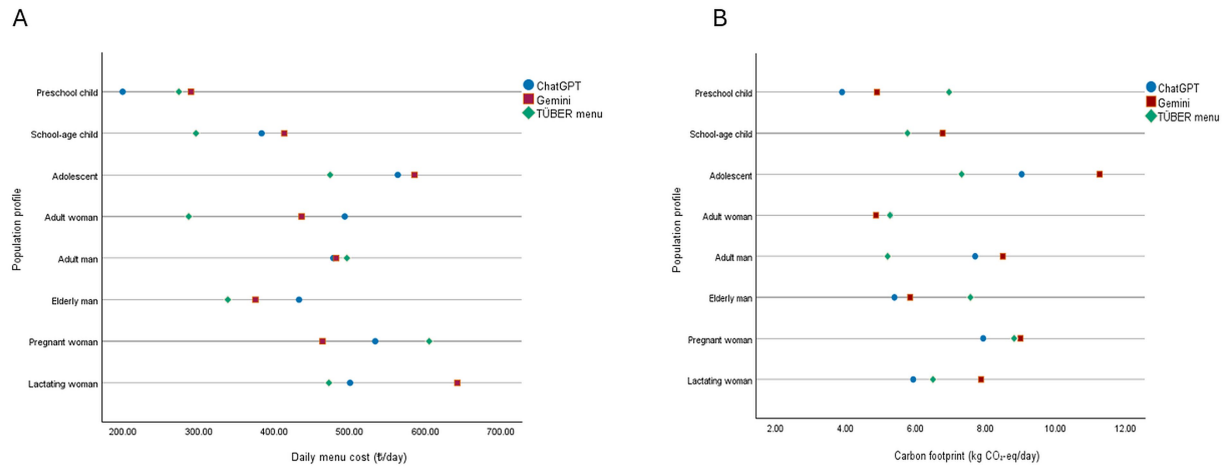


FIGURE 2
Profile-specific daily menu cost and carbon footprint of ChatGPT-, Gemini-, and TÜBER-based menus. Panel **A** shows the daily menu cost in Turkish lira (₺/day), and Panel **B** shows the daily carbon footprint in kg CO_2-eq/day. For ChatGPT and Gemini, each point represents the median of three repeated menu generations for the corresponding population profile. For TÜBER, each point represents the respective profile-specific reference menu. CO_2-eq, carbon dioxide equivalent; TÜBER, Türkiye Dietary Guidelines.

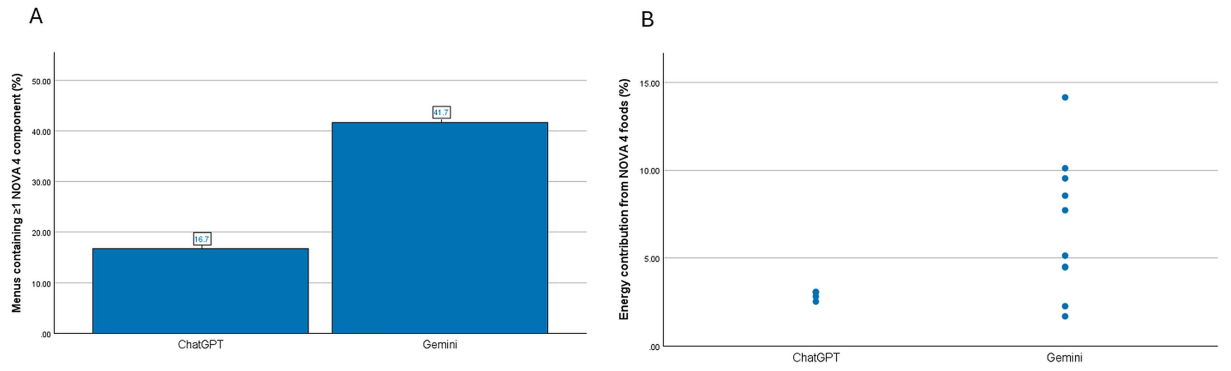


FIGURE 3
Occurrence and energy contribution of NOVA 4 foods in ChatGPT- and Gemini-based menus. Panel **A** shows the percentage of menus containing at least one NOVA 4 component. Panel **B** shows the percentage contribution of NOVA 4 foods to total menu energy only among menus containing at least one NOVA 4 component (ChatGPT, *n* = 4; Gemini, *n* = 10; TÜBER menus, *n* = 0). Each point in Panel **B** represents an individual menu. NOVA 4, ultra-processed foods; TÜBER, Türkiye Dietary Guidelines.

10.3389/fsufs.2026.1984101

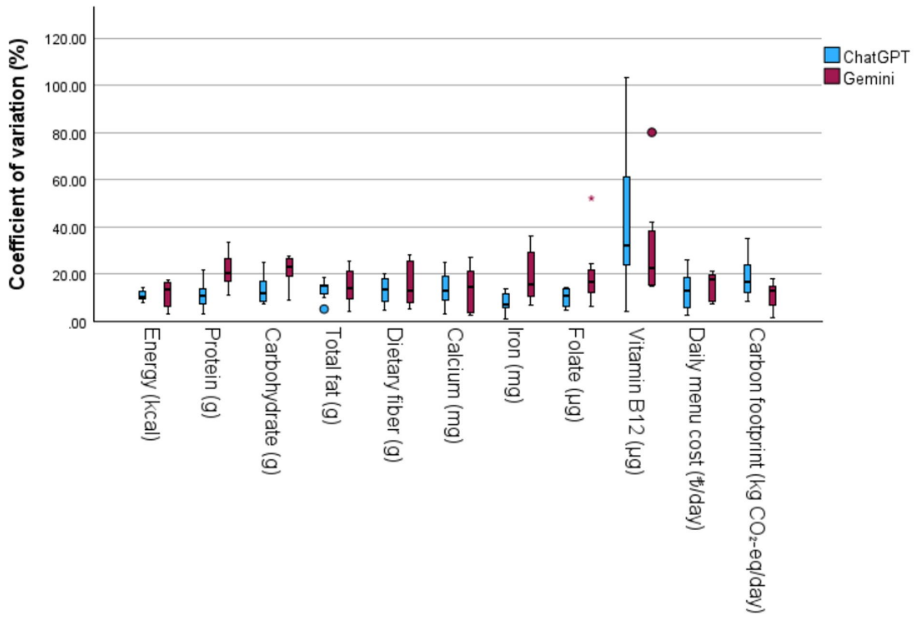


FIGURE 4
Short-term temporal variability in ChatGPT- and Gemini-generated menu outputs across three generations at 2-week intervals Short-term temporal variability was assessed using the coefficient of variation (CV) calculated separately for each population profile from three menu generations performed at 2-week intervals. CV was calculated as the standard deviation divided by the mean and multiplied by 100. Boxes represent the interquartile range (Q1–Q3), and the horizontal line within each box represents the median CV across the eight population profiles. Lower CV values indicate greater temporal consistency, whereas higher values indicate greater variability across repeated generations. Circles indicate outliers, and asterisks indicate extreme outliers. TÜBER menus were excluded because repeated menu generations were not suitable. CV, coefficient of variation; TÜBER, Türkiye Dietary Guidelines.

4 Discussion

In this study, daily menus generated by ChatGPT and Gemini for different life stages and physiological conditions were compared with TÜBER-based reference menus in terms of nutrient adequacy, UPFs content, economic cost, carbon footprint, and short-term temporal consistency. Importantly, the TÜBER comparator was a unique, guideline-based menu for each demographic profile and should therefore be considered a reference example rather than a comprehensive representation of all menus that may align with TÜBER recommendations. While the energy adequacy of the AI-generated menus was generally close to the TÜBER references, significant differences were observed in the distribution of macronutrients. In particular, carbohydrate adequacy was found to be low, while protein and total fat adequacy were high; profile-specific evaluations showed that nutrient alignment can vary according to life stages. While UPFs were not found in TÜBER menus, they appeared more frequently, particularly in Gemini menus. Conversely, although economic costs were higher in the AI-generated menus and the carbon footprint was descriptively higher, particularly in the Gemini menus, differences among menu sources were not found to be statistically significant; furthermore, short-term variability was observed for certain nutrients in repeated AI outputs.

In our study, while the energy adequacy of the menus generated by ChatGPT and Gemini was generally close to the profile-specific TÜBER reference values, this alignment in energy content was not reflected in the macronutrient distribution. Across all menu data sources, carbohydrate values were below 95% of the corresponding reference values, whereas protein and total fat values exceeded 110% of their respective reference values. These operational thresholds were used descriptively and do not, by themselves, indicate clinical inadequacy or excessive intake. This pattern was observed across all menu sources and should therefore not be considered an AI-specific finding. From a dietary perspective, maintaining a relatively lower carbohydrate and higher protein and fat distribution over time could alter the overall balance of macronutrient intake. However, the present study evaluated single-day menus and used the <90%, 90%–110%, and >110% categories as pragmatic descriptive classifications. Therefore, these findings should be interpreted with caution and should not be considered evidence of inadequate carbohydrate intake. Similarly, values exceeding 110% do not constitute evidence of harmful or excessive consumption and should be interpreted in the context of the specific nutrient and the relevant nutritional reference framework. Similarly, Hieronimus et al. (2024) reported that in menus generated by ChatGPT and Bard for different dietary models, energy and carbohydrate intake were generally below targets, while protein intake was mostly above recommendations (Hieronimus et al., 2024). In another study, menus created using ChatGPT-4 and ChatGPT-4o with a target of 1,800 kcal were also found to be deficient in energy, fat, and carbohydrates (Bayram et al., 2025). Other studies comparing different AI models and dietary patterns have also reported that energy and macronutrient contents can vary significantly among the models (Bayram and Arslan, 2025; Gürbüz, 2025). A recent study highlighted the importance of a multidimensional approach when evaluating AI-generated nutrition plans. In this study, nutrition plans created by seven different AI tools for eight client profiles were assessed using a multi-criteria framework that considers nutritional adequacy, feasibility, personalization, reproducibility, and other clinical and practical criteria (Erik et al., 2026). Similarly, Papastratis et al. (2024) examined nutrition plans generated by two large language models for patients with obesity, cardiovascular disease, and type 2 diabetes, concluding

that further oversight by nutritionists or knowledge-based systems is required to ensure meal suitability (Papastratis et al., 2024). When these findings are evaluated together, the fact that a menu generated by AI is close to the target or reference value in terms of total energy does not necessarily mean that the macronutrient distribution is adequate and balanced. In particular, it should be noted that even if a specific energy target is provided to the AI, the actual energy and nutrient content of the generated menu may not align with that target. Furthermore, profile-specific assessments have shown that general median values may mask nutrient deficiencies during certain life stages. This is particularly important during life stages when nutritional needs must be met with greater care. In our study, the menu generated by ChatGPT for the elderly profile fell below the reference value for carbohydrates while exceeding the reference values for iron, folate, and B12; in the Gemini menu, however, energy, carbohydrates, dietary fiber, and calcium were below the reference values. In a study evaluating ChatGPT menus for individuals aged 70 and older living in nursing homes, it was reported that levels of protein, carbohydrates, dietary fiber, vitamin A, vitamin E, vitamin B2, and selenium were significantly higher than the DRI values. Similarly, model-specific discrepancies were observed in the pregnancy profile as well (Altundağ, 2025). While the ChatGPT menu was within the energy reference range, carbohydrates, iron, and folate were below the reference range; similarly, the Gemini menu showed that energy, carbohydrates, iron, and folate were also below the reference range. These discrepancies suggest that even in AI-generated menus for the same population group, nutrient adequacy can vary depending on the model used, the request structure, and the menu creation conditions. NutriSteppe-AI has introduced an alternative approach based on a combination of a structured nutrient database with Large Language Model (LLM)-supported interaction, fixed dietary constraints, multi-objective optimization via linear programming, and processing-level scoring based on the NOVA classification. The system was found to fully meet constraints in 96.8% of 10,000 simulated user profiles (Salkhanova et al., 2026). Due to significant differences in system design and evaluation methodology, these results cannot be directly compared with the current study. However, the findings indicate that integrating LLMs with specific dietary constraints and systematic optimization processes could enhance adherence to established nutritional goals.

In addition to nutrient adequacy, the degree of processing of the foods included in the menus is also an important factor in assessing diet quality. In our study, while none of the TÜBER reference menus contained UPFs classified in the NOVA 4 group, the proportion of menus containing at least one NOVA 4 component was found to be 16.7% in ChatGPT and 41.7% in Gemini; the energy contribution from these foods in menus containing NOVA 4 was also higher and more variable in Gemini. Recent studies directly addressing AI and the NOVA classification also indicate that differences may exist between models. Cattem et al. (2026) reported that in the NOVA classification of 1,168 foods, ChatGPT-o1 demonstrated higher accuracy and sensitivity than DeepSeek-R1 in identifying UPFs (Cattem et al., 2026). Similarly, a recent study evaluating 2920 packaged products showed that ChatGPT, Gemini, and Grok performed differently from the reference NOVA classification and that all models underestimated the prevalence of UPFs under a standard initial prompt; furthermore, performance was significantly sensitive to prompt structure and model version (Bayram et al., 2026). These studies examined the capability of LLMs to classify foods according to NOVA and not the NOVA 4 components of AI generated menus, but they do show the limitations of using only AI models for food processing classifications. In this study, the authors used NOVA categories, not the AI models; therefore, differences in the NOVA 4 content represent differences in the foods present in the generated menus and not differences in the models' ability to classify NOVA. Also, an AI-generated menu that appears adequate in terms of energy and nutrient content may not necessarily represent a healthy dietary pattern with respect to the degree of food processing.

Economic accessibility and environmental sustainability are important components of the feasibility of healthy diets, as well as nutritional adequacy. Although daily cost was descriptively higher in the AI-generated menus and the highest median carbon footprint was observed for Gemini, the analyses did not provide sufficient evidence of differences among menu sources for either outcome. Given the limited number of profile-level observations, these non-significant findings should not be interpreted as demonstrating equivalence. Furthermore, the observed small differences in median carbon footprints across ChatGPT, Gemini and TÜBER menus should be interpreted with caution, as emission factors were mainly derived from global life-cycle assessment databases, which are associated with uncertainties due to differences in production systems, supply chains and food-specific assessments. Therefore, these descriptive differences may be within the ambiguity of the basic emission factors and should not be interpreted as actual environmental differences between sources of menu items. Similarly, a recent study reported that the carbon footprints of gluten-free menus created by ChatGPT and Gemini were not significantly different (Duman et al., 2026). Köktürk and Yardımcı (2026) showed that while sustainable diets created by Gemini, Copilot, ChatGPT, and Grok generally have a low ecological footprint, their environmental performance varies according to the model and menu composition, and the carbon footprint is particularly related to protein and some nutrients (Köktürk and Yardımcı, 2026). Economically, studies have reported that ChatGPT offers lower costs and better cost-effectiveness compared to DeepSeek-R1 in menus designed for pregnant women (Gao et al., 2026). These findings suggest that the economic and environmental performance of AI-generated menus may vary depending on the AI model used and the composition of the menus.

Another important point in evaluating AI-generated menus is the repeatability of the outputs produced under the same conditions. While short-term variability in energy was relatively limited in our study, more significant differences were observed in some macro and micronutrients; particularly high coefficients of variation were detected in vitamin B12. The comparatively high variability observed for vitamin B12 may reflect the strong influence of specific B12-rich foods on the nutrient content of a single-day menu. Relatively small differences between repeated productions in the inclusion or portion size of foods such as meat, fish, eggs, and dairy products may therefore result in substantial proportional changes in total vitamin B12 content. Menu cost and carbon footprint also varied between repeated productions. Similarly, ChatGPT-4o reported that energy and macronutrient composition, in particular, can show inconsistencies over time in menus generated over repeated periods for different dietary models (Ozlu Karahan and Kenger, 2025). Ponzo et al. (2024b) also showed that the nutritional recommendations of AI tools are not fully repeatable in repeated queries. These findings suggest that a single AI output may be insufficient to represent the overall performance of the model. Moreover, generative AI models' performance may differ with the updates to the models themselves. In this regard, the results generated from a specific model version and for a particular timeframe may not be fully reproducible in the future. Nutritional results

generated by AI are constantly changing, so they need to be constantly validated, not just once.

In contrast, a recent study used EvoMeal, an AI-based tool employing a multi-objective optimization methodology, to redesign existing nursing home dinner menus based on nutritional, menu-planning, and ecological criteria. The optimized 4-week meal plans showed a reduction in carbon footprint of approximately 36%–42%, while a 40% reduction was observed in the 12-week menu. Notably, the composition of the existing meals remained unchanged; instead, the arrangement of these meals was improved through a deliberate optimization process (Aytekin-Sahin et al., 2026). When compared to the generative approach used in the present study, this methodological difference suggests that environmental benefits and sustainability criteria may depend on their explicit integration into the menu planning or optimization process.

4.1 Limitations

This study has several limitations that should be considered when interpreting the findings. The analysis included only two generative AI models, eight population profiles, and three menu generations for each profile. Although these repeated generations provided some information on menu variability, three iterations may not fully capture the range of menus that the models could generate. This may be particularly relevant for nutrients with greater variation, such as vitamin B12. The comparison with TÜBER was also based on one reference menu for each profile, while repeated TÜBER menus were not available. Thus, the variability observed across AI-generated menus could not be directly compared with variability in guideline-based menus. In addition, the inferential analyses were based on only eight matched profile-level observations, which limited statistical power. Accordingly, non-significant findings, particularly those related to cost and carbon footprint, should be interpreted as indicating insufficient evidence of a difference rather than evidence of equivalence. Likewise, although the post-hoc comparisons yielded relatively large effect-size estimates, these should be interpreted cautiously given the small number of matched population profiles on which they were based. Future studies could address these limitations by evaluating more AI-generated menus and including multiple guideline-based reference menus for each population profile. Another limitation is that the analysis focused on single-day menus. Such menus may not reflect usual dietary intake or the range of foods consumed over longer periods, and they do not provide information on long-term adherence. The carbon footprint estimates were also based on international life-cycle assessment data and may not fully reflect food production and supply conditions in Türkiye. In addition, uncertainty in the underlying emission factors may affect the magnitude of the estimated differences among menu sources. These estimates should therefore be considered indicative rather than country-specific estimates. In addition, the study evaluated modeled menus rather than actual dietary practices. Consequently, factors such as cultural acceptability, feasibility of food preparation, individual preferences, and adherence in real-life settings could not be assessed. The NOVA classification was not performed blinded to menu origin, which may have introduced some potential for classification bias, although the classification and subsequent review were conducted by multiple researchers. Finally, the AI models were accessed through web-based interfaces, and exact backend build or version identifiers were not available. Changes to the models over time may therefore affect the reproducibility of the findings.

5 Conclusion

This study has demonstrated that the one-day menus generated by the evaluated AI models generally showed energy values close to the profile-specific TÜBER references; however, they may exhibit limitations in terms of nutrient balance, UPFs content, and output consistency. Given the exploratory nature and methodological limitations of the study, the results should be considered as hypothesis-generating and not hypothesis-testing. Although the analyses provided insufficient evidence of differences among menu sources in economic cost and carbon footprint, descriptive variations across models and population profiles were observed. These findings indicate that AI-based menus should not be viewed as independent tools that replace dietitians in nutrition counseling but rather as supportive tools that can be used under the guidance of a dietitian's expert assessment and decision-making process. In particular, dietitians should remain at the center of nutrition planning for different life stages and physiological conditions; AI outputs should be validated in accordance with evidence-based national nutrition guidelines such as those from TÜBER and individualized as needed. Future research should evaluate extended multi-day meal plans to determine if these findings are sustained over longer periods and more closely approximate dietary diversity and habitual intake.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ES: Data curation, Resources, Visualization, Conceptualization, Project administration, Methodology, Investigation, Writing – original draft. TÖ: Data curation, Conceptualization, Project administration, Resources, Writing – original draft, Formal analysis, Visualization, Investigation. BU: Writing – original draft, Data curation, Visualization, Investigation, Project administration, Conceptualization. EY: Investigation, Methodology, Writing – review & editing, Supervision, Resources, Visualization, Conceptualization, Data curation, Project administration.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Anthropic Claude (Claude Sonnet 4.6, Anthropic PBC, San Francisco, CA, USA; https://claude.ai/) was used to assist with language editing and improving sentence structure. The authors have checked the factual accuracy of all content generated with the assistance of generative AI and take full responsibility for the final content of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsufs.2026.1984101/full#supplementary-material

References

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., et al. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med. Educ.* 23:689. doi: 10.1186/S12909-023-04698-z.

Altundağ, Ö. Ö. (2025). Menu planning and nutrient analysis with contemporary AI technologies: a ChatGPT-based approach (Günümüz yapay zeka teknolojileri ile menü planlaması ve besin öğeleri analizi: chatgpt tabanlı bir yaklaşım). *Çukurova Üniv. Sos. Bilim. Enst. Derg.* 34, 16–35. doi: 10.35379/CUSOSBIL.1642390

Anastasiou, K., Ribeiro De Melo, P., Slater, S., Hendrie, G. A., Hadjikakou, M., Baker, P. K., et al. (2022). From harmful nutrients to ultra-processed foods: exploring shifts in 'foods to limit' terminology used in national food-based dietary guidelines. *Public Health Nutr.* 26:2539. doi: 10.1017/S1368980022002580.

Artificial Intelligence Policies Association (AIPA) (2025). *Toplumda Yapay Zeka Algısı [Perception of Artificial Intelligence in Society]*. Available online at: AIPA report – Toplumda Yapay Zeka Algısı. Available online at: https://aipaturkey.org/media/pdfs/yapay-zeka-politikalari-aipa-dernegi-gelecek-arastirmasi-toplumda-yapay-zeka-algisi_WLM77TU.pdf (Accessed July 7, 2026).

Aytekin-Sahin, G., Sahin, O., and Henjum, S. (2026). Optimizing nursing home menus in Norway from a sustainability and nutritional perspective. *Front. Nutr.* 13:1776523. doi: 10.3389/fnut.2026.1776523.

Bayram, H. M., and Arslan, S. (2025). Nutritional analysis of AI-generated diet plans based on popular online diet trends. *J. Food Compos. Anal.* 145:107850. doi: 10.1016/J.JFCA.2025.107850

Bayram, H. M., Arslan, S., and Ozturkcan, A. (2026). Can machines detect ultra-processed foods? A head-to-head evaluation of large language models using NOVA classification. *Int. J. Food Sci. Technol.* 61:vvag142. doi: 10.1093/IJFOOD/VVAG142

Bayram, H. M., Çelik, Z. M., and Barcın Güzeldere, H. K. (2025). Can artificial intelligence (AI) chatbot tools be used effectively for nutritional management in obesity?. *Nutr. Health* 31, 1825–1834. doi: 10.1177/02601060251329070.

Bayram, H. M., 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. *Br. Food J.* 126, 3335–3346. doi: 10.1108/BFJ-02-2024-0158

Cattem, M. V. d. O., Machado, F. O., da Silva, M. E. G., Canella, D. S., and Koury, J. C. (2026). Comparing ChatGPT and DeepSeek for ultra-processed food classification: AI models for nutritional research and dietary assessment. *Nutrition* 144:113066. doi: 10.1016/J.NUT.2025.113066.

Duman, E., Songür Bozdağ, A. N., Memiç İnan, C., and Canbolat, E. (2026). Gluten-free diet plans generated by ChatGPT and Gemini differ in nutrient adequacy, while diet quality and carbon footprint are comparable. *Nutrition* 150:113323. doi: 10.1016/J.NUT.2026.113323.

Erik, A., Hamidy, S. M., Karamancıoğlu, H., and Küçük Kırtıklı, B. N. (2026). AI in nutrition: multi-criteria analysis of diet plans across diverse client profiles. *Nut. Res.* 146, 53–67. doi: 10.1016/j.nutres.2025.12.006

Erkkola, M., Berg, O., Brugård Konde, Å., Gunnarsdottir, S., Dahl Lassen, A., Kaartinen, N. E., et al. (2026). From evidence to action: implementing the Nordic Nutrition Recommendations in national policy. *Proc. Nutr. Soc.* 85, 241–252. doi: 10.1017/S0029665125100682.

Food and Agriculture Organization of the United Nations and World Health Organization (2019). *Sustainable Healthy Diets Guiding Principles*. Rome: Food and Agriculture Organization of the United Nations (FAO) and World Health Organization (WHO).

Gao, Z. J., Li, J., and Fang, W. Y. (2026). Dietary guidance for pregnant women using DeepSeek-R1 and ChatGPT-4.0: a comparative analysis. *Front. Public Health* 14:1728954. doi: 10.3389/fpubh.2026.1728954.

George, D., and Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update*. 10 th. Boston: Pearson.

Gürbüz, M. (2025). Artificial intelligence-generated diets in comparison: a nutrient profiling approach of popular diet models (Yapay zeka destekli diyetlerin karşılaştırılması: popüler diyetlere besin ögesi profilleme yaklaşımı). *J. Health Sci. Manag.* 1, 1–9. doi: 10.29228/JOHESAM.49

Hee Lee, D., and Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. *Int. J. Environ. Res. Public Health* 18, 1–18. doi: 10.3390/IJERPH18010271.

Herforth, A., Arimond, M., Álvarez-Sánchez, C., Coates, J., Christianson, K., and Muehlhoff, E. (2019). A global review of food-based dietary guidelines. *Adv. Nutr.* 10:590. doi: 10.1093/ADVANCES/NMY130.

Herrera-Cuenca, M., Yépez García, M. C., Cortés Sanabria, L. Y., Hernández, P., Sifontes, Y., Ramírez, G., et al. (2023). Contribution of proteins to the Latin American diet: results of the ELANS study. *Nutrients* 15:669. doi: 10.3390/nu15030669.

Hieronimus, B., Hammann, S., Hammann, S., and Podszun, M. C. (2024). Can the AI tools ChatGPT and Bard generate energy, macro- and micro-nutrient sufficient meal plans for different dietary patterns? *Nut. Res.* 128, 105–114. doi: 10.1016/J.nutres.2024.07.002

Hieronimus, B., Lopez-Aguirre, M. L., Birringer, M., and Podszun, M. (2025). GenAI in nutritional sciences (GAINS): a systematic review and reporting framework for future research. *Nut. Res.* 143, 66–77. doi: 10.1016/J.nutres.2025.09.011

Kaya Kaçar, H., Kaçar, Ö. F., and Avery, A. (2025). Diet quality and caloric accuracy in AI-generated diet plans: a comparative study across chatbots. *Nutrients* 17:206. doi: 10.3390/nu17020206/S1

Köktürk, S. N., and Yardımcı, H. (2026). Ecological and nutritional analysis of artificial intelligence-based sustainable dietary models. *J. Food Compos. Anal.* 149:108814. doi: 10.1016/J.JFCA.2025.108814

Liu, Y., and Wang, H. (2026). Who on Earth is using generative AI?. *World Dev.* 199:107260. doi: 10.1016/J.WORLDDEV.2025.107260

Market Fiyatı (2025). Available online at: https://marketfiyati.org.tr/ (Accessed June 01, 2026).

Monteiro, C. A., Cannon, G., Moubarac, J. C., Levy, R. B., Louzada, M. L. C., and Jaime, P. C. (2017). The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutr.* 21, 5–17. doi: 10.1017/S1368980017000234.

Naja, F., Taktouk, M., Matbouli, D., Khaleel, S., Maher, A., Uzun, B., et al. (2024). Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. *Eur. J. Clin. Nutr.* 78, 887–896. doi: 10.1038/S41430-024-01476-y.

Ostrowska, M., Komoń-Janczara, E., Mikoluc, B., Iłowiecka, K., Jarczak, J., Zagórska, J., et al. (2025). The gut mycobiome and nutritional status in paediatric phenylketonuria: a cross-sectional pilot study. *Nutrients* 17:2405. doi: 10.3390/nu17152405.

Ozlu Karahan, T., and Kenger, E. B. (2025). ChatGPT-4o for weight management: comparison of different diet models. *Food Sci. Nutr.* 13:e70639. doi: 10.1002/FSN3.70639.

Papastratis, I., Stergioulas, A., Konstantinidis, D., Daras, P., and Dimitropoulos, K. (2024). Can ChatGPT provide appropriate meal plans for NCD patients?. *Nutrition* 121:112291. doi: 10.1016/J.NUT.2023.112291.

Petersson, T., Secondi, L., Magnani, A., Antonelli, M., Dembska, K., and Valentini, R. (2021)et al. A multilevel carbon and water footprint dataset of food commodities. *Sci. Data* 8:127. doi: 10.1038/s41597-021-00909-8

Petrus, R. R., do Amaral Sobral, P. J., Tadini, C. C., and Gonçalves, C. B. (2021). The NOVA classification system: a critical perspective in food science. *Trends Food Sci. Technol.* 116, 603–608. doi: 10.1016/J.TIFS.2021.08.010

Ponzo, V., Goitre, I., Favaro, E., Merlo, F. D., Mancino, M. V., Riso, S., et al. (2024a). Is ChatGPT an effective tool for providing dietary advice?. *Nutrients* 16:469. doi: 10.3390/NU16040469/S1

Ponzo, V., Rosato, R., Scigliano, M. C., Onida, M., Cossai, S., De Vecchi, M., et al. (2024b). Comparison of the accuracy, completeness, reproducibility, and consistency of different AI chatbots in providing nutritional advice: an exploratory study. *J. Clin. Med.* 13:7810. doi: 10.3390/JCM13247810.

Poore, J., and Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science (1979)*. 360, 987–992. doi: 10.1126/science.aaq0216

Republic of Türkiye Ministry of Health (2019). *Türkiye Nutrition and Health Survey 2019*. Ankara: Republic of Türkiye Ministry of Health. Available online at: https://ekutuphane.saglik.gov.tr/Yayin/738 (Accessed July 7, 2026).

Republic of Türkiye Ministry of Health (2022). *Türkiye Dietary Guidelines (TÜBER) 2022*. Ankara: Republic of Türkiye Ministry of Health.

Sadler, C. R., Grassby, T., Hart, K., Raats, M., Sokolović, M., and Timotijevic, L. (2021). Processed food classification: conceptualisation and challenges. *Trends Food Sci. Technol.* 112, 149–162. doi: 10.1016/J.TIFS.2021.02.059

Salkhanova, A., Nabigazinova, E., Kaldybay, A., Omirbekova, A., Sabit, M., Baikonsova, L., et al. (2026). NutriSteppe-AI: development, architecture, and explainable design of a large language model–driven chatbot for personalized health menu generation. *Nutrients* 18:2228. doi: 10.3390/NU18142228

Şimşek, H., Rajabi, A., Öztürk, B., and Uçar, A. (2025). Ultra-processed foods marketed in Türkiye: an analysis of nutritional quality and packaging sustainability. *Nutr. Bull.* 50, 302–310. doi: 10.1111/nbu.70005.

Springmann, M., Spajic, L., Clark, M. A., Poore, J., Herforth, A., Webb, P., et al. (2020). The healthiness and sustainability of national and global food based dietary guidelines: modelling study. *BMJ* 370:m2322. doi: 10.1136/BMJ.M2322.

U.S. Department of Health and Human Services, U.S. Department of Agriculture (2026). *Dietary Guidelines for Americans, 2025–2030*. 10th ed. Washington, DC: U.S. Department of Health and Human Services and U.S. Department of Agriculture. Available online at: RealFood.gov (Accessed August 31, 2026).