ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

UDC 004.8:613.2
IRSTI 76.33.35
DOI: 10.53065/kaznmu.2026.78.3.004

Received: 17.08.2026
Accepted for publication: 22.09.2026

AI-ENABLED PERSONALISED NUTRITION FOR CARDIOMETABOLIC HEALTH: A SYSTEMATIC REVIEW OF CLINICAL EFFECTS AND GENERATED DIETARY ADVICE

F. KHOZHAMKUL, A. SHARMAN

Kazakh Academy of Nutrition, Almaty, Kazakhstan

Abstract

Introduction. AI-enabled nutrition includes predictive dietary personalisation, digital lifestyle programmes and generated dietary advice.

Aim. To assess clinical outcomes of AI-enabled personalised nutrition in adults with cardiometabolic risk and the accuracy and safety of generated dietary advice.

Materials and methods. Five bibliographic databases and two trial registries were searched from inception to August 2026, with supplementary searches in Semantic Scholar, medRxiv and selected web sources. Eligible interventions were digital tools in which artificial intelligence, machine learning, a large language model or a conversational agent generated individualised dietary, menu or meal recommendations. Clinical outcomes and generated-advice accuracy were synthesised separately. Risk of bias was assessed using RoB 2, ROBINS-I and an adapted QUADAS-2 instrument; certainty was evaluated with GRADE.

Results. The final working evidence inventory comprised 62 studies (64 reports): 26 in the clinical domain and 36 in the accuracy domain. A machine-learning-guided diet reduced HbA1c by 0.08 percentage points more than a Mediterranean diet at six months (95% CI, −0.14 to −0.02), whereas a separate personalised-diet trial did not improve six-month weight loss over a low-fat diet (difference +1.05 percentage points of baseline weight; 95% CI, −0.40 to 2.50). Personalised LLM prompting improved weight loss compared with manual ChatGPT prompting, while an AI-based sodium programme did not improve salt intake or blood pressure. A 368-participant trial met its prespecified noninferiority criterion for referral to an AI-led rather than human-coached diabetes-prevention programme. Generated-advice studies reported favourable expert ratings alongside nutrient and guideline discrepancies; complex profiles were more challenging in an exploratory chatbot comparison.

Conclusion. Selected interventions improved specific outcomes, but a class-wide benefit remains unestablished. Programme effects do not isolate the AI component, and favourable content ratings do not establish nutritional adequacy or patient safety.

Key words: personalised nutrition, artificial intelligence, large language models, cardiometabolic health, dietary advice, evidence synthesis.

Introduction. AI-enabled nutrition comprises interventions with different mechanisms. A predictive model may rank meals according to expected postprandial glucose responses, whereas a conversational system may generate menus, explanations or behavioural feedback. Predictive personalisation, structured digital lifestyle programmes and open-ended language-model advice should therefore not be interpreted as a single intervention class. Clinical trials already show both an incremental glycaemic benefit and failure to demonstrate incremental weight loss against active dietary comparators [1,2].

The comparator determines what a trial can identify. A comparison of an AI-enabled package with usual care estimates the effect of the package, not necessarily the isolated

© The Author(s), 2026. Licensed under CC BY 4.0

contribution of AI. An active comparator can test the incremental value of personalisation more directly. A noninferiority trial of referral to alternative delivery programmes asks a different question again: whether a more scalable delivery option preserves an acceptable level of benefit [3].

For generated dietary advice, effectiveness is preceded by a content-validity question. Expert acceptability, nutrient adequacy, guideline concordance, reproducibility and observed patient harm are not interchangeable constructs. Expert surveys, simulated-case evaluations and retrieval-augmented benchmarks illustrate how strongly the conclusion can depend on the reference standard used [4–6].

This review therefore separates clinical outcomes from the quality and safety of generated dietary content. It also source-verifies the underlying evidence set and distinguishes what each study design permits the reader to infer. The emphasis is on identifiable contrasts, verified outcomes and explicit limits of inference rather than on a single verdict for all systems labelled "AI".

Materials and methods.

Ethical issues

This review used published literature and publicly available reports. No participants were recruited, and no identifiable individual-level data were collected or analysed; therefore, ethics committee approval was not required.

Review design and reporting framework

The review was registered in PROSPERO as CRD420261480490. Registration was retrospective: searching and screening had already been completed and extraction was under way when the record was submitted. A written protocol specifying the question, eligibility criteria, search strategy and planned synthesis was prepared before screening. Amendments and deviations are reported in the supplementary materials. Reporting follows the PRISMA 2020 statement, while SWiM informed the grouping and presentation of effect estimates when meta-analysis was not appropriate [7,8].

Eligibility, outcomes and technology definition

The clinical population comprised adults with overweight or obesity, prediabetes, type 2 diabetes, metabolic syndrome, hypertension, dyslipidaemia or physician-defined elevated cardiovascular risk. Results from mixed populations were attributed to this scope only when an eligible subgroup was separable. Clinical designs included randomised, non-randomised and prospective single-arm studies. Outcomes of interest were HbA1c, fasting glucose, body weight or BMI, blood pressure and lipids, with dietary behaviour and engagement as secondary outcomes.

The operational technology criterion required a learned model, language model or AI-enabled system to contribute to personalised dietary or nutrition-related advice. An AI component used only for image recognition, logging, risk prediction or data capture did not by itself satisfy the narrow recommendation criterion. Fixed-rule, fuzzy-logic or ontology-only personalisation was excluded under the team clarification. Where the technical role of AI remained uncertain, the record was treated as a technology-boundary case rather than assumed eligible.

The content domain comprised generated dietary or nutrition-management advice assessed against an external clinical, nutrient, algorithmic or expert reference. AI-as-evaluator studies and general health-management evaluations without a separable nutrition endpoint were retained as contextual evidence rather than treated as direct demonstrations of generated-advice accuracy.

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

Information sources and source verification

The documentation reports searches of PubMed, Cochrane CENTRAL, Embase, Scopus and Web of Science, together with ClinicalTrials.gov and WHO ICTRP, before August 2026. It also describes an earlier supplementary search using bibliographic, preprint and web channels. Search strings and source totals were inspected as supplied; the full database searches were not rerun during the separate source-verification audit.

Targeted verification used exact titles, author names, DOI or PubMed identifiers from the working evidence register and extraction sheets. Primary journal pages, indexed abstracts and accessible full reports were checked to resolve outcome-bearing inconsistencies, publication status, technical eligibility and companion reports. Evidence was labelled as full-report verified, abstract/bibliographic verified, extraction-supported only, or unresolved. Lack of access was not converted into a zero effect or an eligibility exclusion.

Data extraction, study linkage and audit procedures

Two reviewers screened the database-derived records and assessed full texts independently, resolving disagreements by discussion. Supplementary records were initially screened in a preliminary pass and subsequently rechecked. Full-text exclusions were assigned one of eight prespecified codes: wrong population, non-AI intervention, AI not generating dietary recommendations, ineligible design, no eligible outcome, unavailable full report, duplicate or companion report, and ineligible publication type.

Data were extracted jointly in a consensus session. Extracted fields included population, setting, intervention, AI type, model name and version, comparator, outcomes, effect estimates, variance, analysis population, attrition, funding and author–sponsor relationships. Accuracy extraction additionally recorded prompts or profiles, reference standards, rater qualifications, blinding, scoring instruments, safety findings and repeated-generation procedures.

The audit checked workbook row counts, study identifiers, outcome definitions, units, copied extraction content and internal cross-references. Numerical corrections were entered only where supported by a primary source or a reproducible calculation from reported summary values; no patient-level observations were reconstructed.

A report and a study were treated as different units. Confirmed companion reports were linked rather than counted as independent replication. The Popp and Kharmats publications were linked through the Personal Diet Study and NCT03336411; confirmed repeated reports from the same commercial cohort were similarly linked. Shared platform, authorship or sponsorship alone was not treated as proof of participant overlap [2,9].

Bias appraisal, prioritisation and synthesis

Risk of bias was assessed at report level. RoB 2 was used for randomised trials, ROBINS-I for non-randomised clinical studies [10], and an instrument adapted from QUADAS-2 for accuracy studies [11], because no validated tool specifically addresses AI-generated dietary content. Two reviewers rated each assessable report and reconciled disagreements. Certainty of clinical evidence was considered using GRADE. No automation tool was used for data extraction or risk-of-bias assessment.

For the main synthesis, priority was given to comparator-based clinical trials and content evaluations with an identifiable primary source. Full-report verification took precedence over abstract-level or extraction-only evidence; lower-access evidence was retained when it materially affected the direction or interpretation of the evidence and was explicitly labelled. Clinical findings were grouped by outcome and comparator: active dietary personalisation, multicomponent programmes, alternative AI configurations and referral-level noninferiority. Content studies were grouped by the construct measured: guideline/nutrient concordance, expert rating, multi-model performance, reproducibility and mitigation.

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

No meta-analysis was performed. The source-qualified subset does not provide a fully reconciled set of independent eligible contrasts, and the trials differ in population, intervention content, comparator, endpoint and follow-up. Pooling selectively retrievable estimates would add a source-availability problem to clinical heterogeneity. No pooled effect, I^2 statistic, funnel-plot test or reconstructed correlation was calculated [8].

Results.

Evidence provenance and study flow

The five stated bibliographic source counts sum to 19,311, whereas 20,442 bibliographic records are reported after removal of 1,173 non-bibliographic items from 21,615 imports. This leaves 1,131 bibliographic records without a source-collection assignment. After removal of 6,418 duplicates, 14,024 records were screened. Subtracting 13,824 title/abstract exclusions yields 200 reports sought for retrieval, consistent with Figure 1.

The full-text screening worksheet contains 200 rows: 69 marked included, 80 excluded and 51 assigned to an ongoing-studies appendix. After re-adjudication and report-level linkage, the final working evidence inventory comprised 62 studies (64 reports): 26 in the clinical-outcome synthesis and 36 in the accuracy-domain synthesis [1–6,9,12–68]. The study-selection pathway is shown in Figure 1.

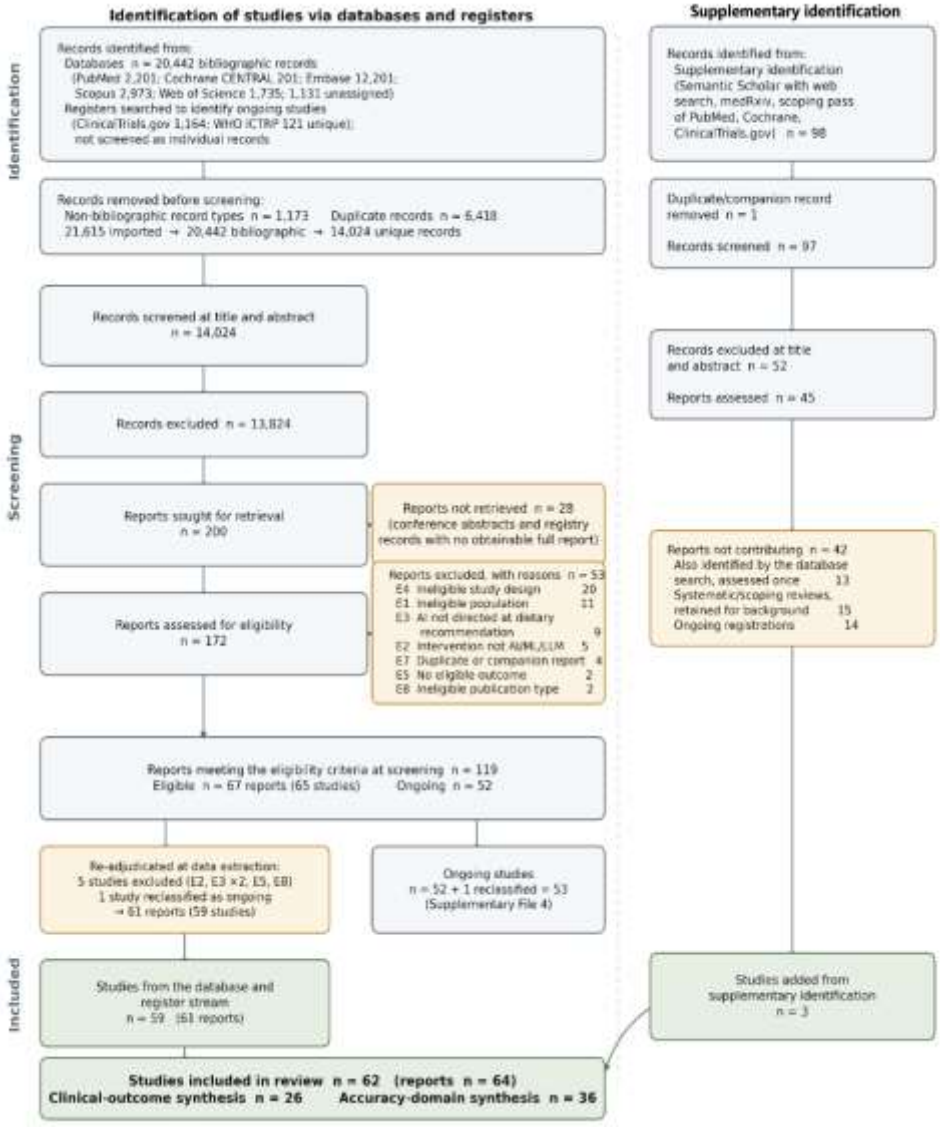


Figure 1. Audited study-selection flow structured according to PRISMA 2020.

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

The review recorded 2 low-risk, 3 some-concerns and 19 high-risk overall judgements in the clinical domain; the accuracy domain recorded 2 low-risk, 17 some-concerns and 7 high-risk judgements. Figures 2 and 3 show the domain-level distribution.

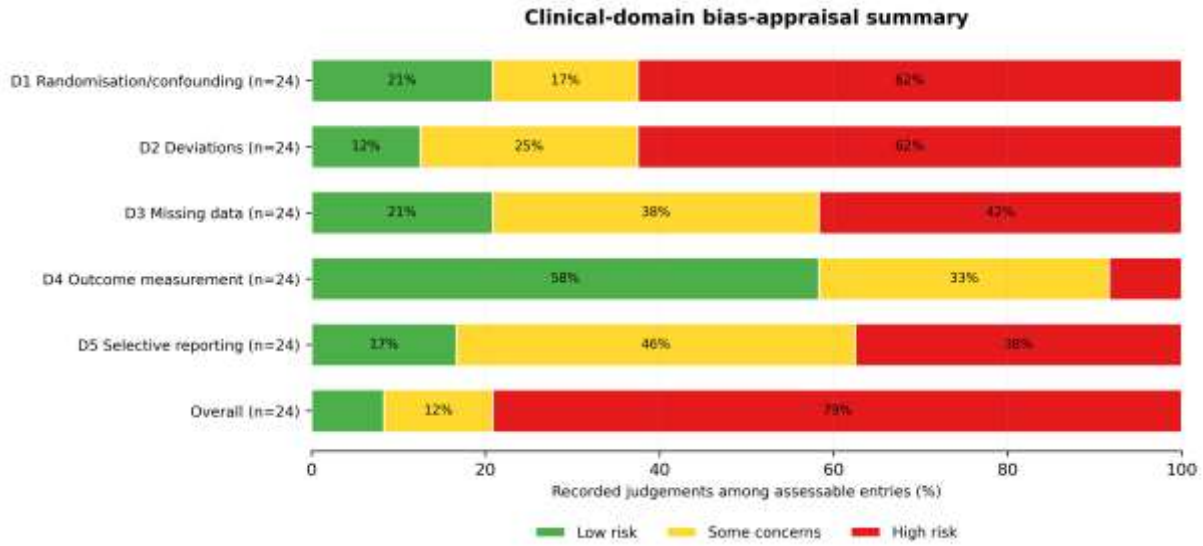


Figure 2. Clinical-domain bias-appraisal summary. Percentages are calculated among assessable recorded judgements in each domain. D1–D5 follow the harmonised author worksheet; because it combines randomised and non-randomised designs, this is an audit summary rather than a new canonical RoB 2 or ROBINS-I assessment.

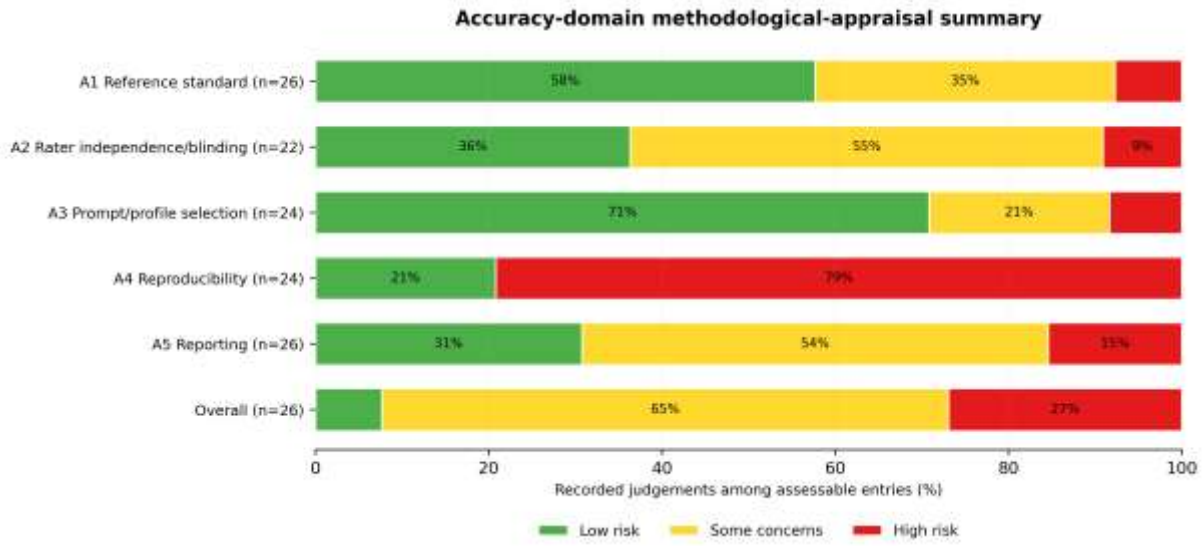


Figure 3. Accuracy-domain methodological-appraisal summary. Percentages are calculated among assessable recorded judgements; not-applicable entries are excluded from each domain denominator. A1–A5 follow the author-developed QUADAS-2-informed instrument, which has not been validated for this application.

The comparative clinical studies and their source-qualified outcome estimates are summarised in Table 1.

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

Table 1. Source-qualified comparative clinical evidence prioritised for synthesis.

Study / design	Comparison	Follow-up	Outcome	Effect estimate
Ben-Yacov 2021 [1]; RCT, n=225	ML-personalised vs Mediterranean diet	6 mo	HbA1c	MD −0.08 pp (95% CI −0.14 to −0.02)
Popp 2022 [2]; RCT, n=204	Personalised vs low-fat diet	6 mo	Weight	MD +1.05% of baseline weight (95% CI −0.40 to 2.50)
Rein 2022 [12]; crossover RCT, n=23	Personalised vs Mediterranean diet	2 wk	Mean glucose	−7.8 mg/dL; CI unavailable
Joshi 2023 [13]; RCT, n=319 eligible	Digital-twin programme vs standard care	1 y	HbA1c	Change −2.9 vs −0.3 pp; p<0.001
Mohd Dan 2026 [14]; RCT, n=160	Personalised prompting + ChatGPT vs manual ChatGPT	24 wk	Weight	Derived descriptive contrast −3.8 kg†
Nakata 2022 [15]; RCT, n=141	CALO mama Plus vs no intervention	3 mo	Weight	Adjusted MD −1.60 kg (95% CI −2.83 to −0.38)
Sato 2025 [16]; three-arm RCT, n=312	Genetic profile + AI app vs control	3 mo	Salt; SBP	Salt −0.2 g/d (−0.7 to 0.3); SBP −2.6 mmHg (−8.7 to 3.4)
Mathioudakis 2025 [3]; NI RCT, n=368	AI-led vs human-coached DPP referral	12 mo	Composite	RD −0.20 pp; lower bound −8.2 pp; NI margin −15 pp

Note: The −3.8 kg contrast is calculated from reported group weight-loss values at 24 weeks; uncertainty for that derived between-group estimate was not available in the review materials. CI – confidence interval; DPP – Diabetes Prevention Program; MD – mean difference; NI – noninferiority; pp – percentage points; RD – risk difference; SBP – systolic blood pressure.

Glycaemic outcomes

In the 225-participant trial by Ben-Yacov and colleagues, a machine-learning-guided postprandial-targeting diet improved HbA1c more than a Mediterranean diet at six months. The between-group difference was approximately −0.08 percentage points, with a published 95% CI of −0.14 to −0.02. Time above the prespecified glucose threshold also improved, whereas the oral glucose tolerance outcome did not show a between-group difference. Both groups received dietary support, so the result supports an incremental effect on selected glycaemic outcomes rather than a general benefit of all forms of personalisation [1].

Rein and colleagues tested two two-week diets in a randomised crossover involving 23 adults with newly diagnosed type 2 diabetes. The personalised diet reduced mean continuously monitored glucose relative to the Mediterranean diet by 7.8 mg/dL in the reported comparison. A subsequent six-month phase involved 16 participants without a concurrent control; its HbA1c change of −0.39 percentage points is therefore an uncontrolled extension result, not the six-month effect of a randomised comparison [12].

The Personal Diet Study provides a counterpoint. The Popp trial did not show superior weight loss, and its linked Kharmats report found no additional improvement in glycaemic

© The Author(s), 2026. Licensed under CC BY 4.0

variability or HbA1c with the personalised diet relative to the standardised low-fat diet. The two papers report one trial and are not independent replication [2,9].

The digital-twin trial by Joshi and colleagues reported much larger glycaemic improvement than standard care, with HbA1c changes of −2.9 and −0.3 percentage points at one year. The programme also incorporated activity, sleep and medication management. The randomised contrast therefore estimates a multicomponent programme effect; it does not isolate the contribution of the AI-generated nutritional component [13].

Weight outcomes and LLM-specific clinical evidence

Popp and colleagues did not demonstrate superior six-month weight loss with a personalised glycaemic-response diet over a low-fat diet when both groups received behavioural counselling. The reported difference was +1.05 percentage points of baseline body weight (95% CI, −0.40 to 2.50), favouring the low-fat diet in direction but not establishing a difference. The finding is specific to the tested intervention, comparator and outcome and is not evidence of equivalence or class-wide ineffectiveness [2].

The NExGEN trial provides directly relevant LLM-based clinical evidence. In 160 adults with overweight or obesity, personalised prompt generation integrated with ChatGPT was compared with structured manual ChatGPT guidance. Reported weight loss was 6.6 versus 3.0 kg at 12 weeks and 5.5 versus 1.7 kg at 24 weeks, corresponding to descriptive signed contrasts of −3.6 and −3.8 kg. Both arms used ChatGPT and the intervention combined dietary and exercise planning, so the trial tests a prompting and personalisation package rather than ChatGPT versus no AI [14].

The CALO mama Plus trial also reported greater weight loss than a no-intervention control. The adjusted between-group difference at three months was −1.60 kg (95% CI, −2.83 to −0.38). Because the app bundles AI-generated advice with self-monitoring and other behavioural functions, the contrast estimates the programme effect rather than the isolated recommendation algorithm [15].

Blood pressure, dietary salt and programme noninferiority

Sato and colleagues randomised 312 participants to a genetic-profile plus AI-app programme, control, or app-only condition. At three months, the treatment group did not reduce salt intake relative to control (MD −0.2 g/day; 95% CI, −0.7 to 0.3) and did not show a systolic blood pressure difference (MD −2.6 mmHg; 95% CI, −8.7 to 3.4). The published outcome table contained 268 observations for salt intake and 128 for blood pressure (85.9% and 41.0% of the randomised sample, respectively). These are outcome-specific analysis denominators rather than a general retention rate [16].

The AI-DPP trial compared referral to an AI-led diabetes-prevention programme with referral to a human-coached programme in 368 adults. Composite success occurred in 58/183 and 59/185 participants, respectively. The unadjusted risk difference was −0.20 percentage points and the published one-sided 95% lower confidence bound was −8.2 percentage points, above the prespecified −15-point noninferiority margin. The result supports noninferiority for the specified composite referral strategy; it does not establish superiority, equal HbA1c control, identical programme exposure or cost-effectiveness [3].

Quality, accuracy and safety of generated dietary advice

Generated-advice studies produced mixed findings that depended on the evaluation construct. Guideline and nutrient assessments identified gaps in weight-loss guidance, energy targets and dietary-pattern concordance, while one comparison reported a lower saturated-fat energy share in AI-generated plans. In simulated diabetes cases, Gemini 1.5 Flash achieved a higher study-specific composite score than GPT-4 (median 4.34 vs 3.79; $p<0.001$). In an exploratory ten-chatbot comparison, no model exceeded 50% on the study accuracy scale in the

© The Author(s), 2026. Licensed under CC BY 4.0

complex profile. These results identify task-specific strengths and failure modes but do not provide a universal accuracy or safety rate [5,17–20].

Expert rating and safety-oriented benchmarking measured different constructs. Kim and colleagues found no statistically significant differences among blinded ratings of a general AI-generated plan and two institutional plans, but the separate personalised-plan assessment was unblinded; nonsignificance therefore does not establish equivalence or nutrient adequacy. Parameswaran and colleagues recorded no harmful responses for a guideline-grounded retrieval-augmented configuration in its finite test set, but this does not establish a zero clinical harm rate or deterministic constraint enforcement [4,6].

The content evaluations, their reference standards and interpretation limits are summarised in Table 2.

Table 2. Source-verified evaluations of generated dietary advice used in the synthesis.

Study	Evaluation / reference	Main finding	Interpretation limit
Naja 2024 [20]	63 prompts; dietitian guideline review	Gaps in weight-loss guidance, energy deficit, nutrients and menu composition	Prompt-level; no patient safety estimate
Kim 2024 [4]	Expert ratings; blinded general-plan comparison	No significant differences among general plans	No equivalence design; personalised assessment unblinded
Ponzo 2024 [5]	10 chatbots; 2 profiles; repeated days	No model exceeded 50% accuracy in the complex profile	Exploratory; no population safety rate
Kenger 2025 [17]	24 simulated hypertension/dyslipidaemia profiles	Weak Mediterranean and DASH adherence	Study-specific profiles and scoring
Ben Lafqih 2025 [19]	52 simulated diabetes cases; multicriteria score	Gemini 4.34 vs GPT-4 3.79; $p<0.001$	Composite score, not universal accuracy
Çelik 2026 [18]	AI vs dietitian nutrient composition	Lower saturated-fat energy share in AI plans	No across-nutrient superiority
Parameswaran 2025 [6]	30 questions; grounded vs off-the-shelf models	No harmful responses in grounded configuration	Finite benchmark; no clinical safety rate

Observational evidence and technology-boundary cases

The observational records mainly describe commercial programmes, engagement-selected users or before-and-after changes without a concurrent counterfactual. They can document feasibility or change among users but cannot separate programme effects from self-selection, concurrent treatment, secular change or regression to the mean. They were not used to derive a pooled treatment effect.

The Lee trial is a clinically informative boundary case. The full report describes an integrated platform in which the AI component performs food recognition to automate dietary logging; it does not establish that a learned model generates individualised dietary recommendations. Under the stated recommendation criterion, the trial is therefore contextual

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 e-ISSN 2524 – 0692

rather than part of the core personalised-recommendation synthesis, despite reporting improved HbA1c with the overall platform [21].

The Educ@dom trial is another boundary case. Its Nutri-Educ software is based on fuzzy arithmetic and heuristic search; under the exclusion of fixed-rule or fuzzy-only personalisation without a learned model, it falls outside the narrow synthesis. This classification is based on technology definition rather than outcome direction [22,23]. Broader management-plan evaluations without a separable nutrition endpoint, including Mondal and colleagues, remain contextual [24].

Certainty of evidence (GRADE)

Certainty was rated per primary outcome across the comparator-based evidence (Table 3).

Table 3. GRADE certainty ratings for primary clinical outcomes.

Outcome	Studies (design)	Certainty	Principal reasons for rating
Weight/BMI vs active or usual care	3 RCTs, 1 pilot RCT	Low	Risk of bias (−1: inactive controls, developer involvement in the positive trials); inconsistency (−1: the largest and cleanest trial is null while smaller, higher-risk trials are positive)
HbA1c vs standard care	2 RCTs, 1 noninferiority RCT, 1 retrospective cohort	Very low	Risk of bias (−2: medication co-intervention in the only strongly positive trial; non-randomised allocation elsewhere); indirectness (−1: intervention bundles)
HbA1c, AI delivery vs human coaching	1 noninferiority RCT	Moderate for noninferiority against the trial's prespecified composite benchmark	Large pragmatic randomised trial with an active, intensity-matched comparator; downgraded once for indirectness (composite benchmark including self-reported components)
Blood pressure / dietary sodium	1 three-arm RCT, 1 uncontrolled pre-post	Moderate for the estimate; the trial found no statistically significant difference	Single well-conducted null RCT with a stated minimum detectable effect; downgraded once for indirectness (workplace population, 90% male)
Lipids, fasting glucose	No adequately controlled primary-endpoint evidence	Very low	Sparse, uncontrolled or secondary reporting only
Dietary quality/adherence	Heterogeneous, mostly single-arm	Very low	Risk of bias, indirectness, no common instrument

No outcome reached high certainty. The only two moderate-certainty conclusions available are not demonstrations of added efficacy: that an AI application for sodium reduction,

© The Author(s), 2026. Licensed under CC BY 4.0

with or without genetic profiling, produced no statistically significant change in dietary sodium or blood pressure in the single trial designed to detect it — a finding of no demonstrated benefit rather than a demonstration of equivalence; and that fully automated AI delivery of a structured lifestyle programme was noninferior, though not superior, to human coaching.

Discussion. The source-verified clinical evidence is heterogeneous rather than uniformly positive or negative. Active-comparator trials support a modest glycaemic benefit in one setting and no incremental weight benefit in another. Multicomponent AI-enabled programmes and an LLM prompting intervention report clinically relevant improvements, while an AI-based sodium-reduction programme was null. The noninferiority trial adds evidence about delivery and scalability rather than a common treatment effect that can be pooled with superiority trials [1–3,12–16].

These results do not show that AI-driven personalisation is necessary for an effective programme, nor do they justify dismissing positive findings solely because the intervention is multicomponent. Findings should be interpreted at the level of the specified programme, comparator, outcome and follow-up. In particular, a package-level effect is not an isolated AI effect, and an inactive comparator cannot separate algorithmic contribution from self-monitoring, contact or other programme components.

Co-interventions have different implications across designs. In a randomised comparison of multicomponent care packages, medication management or coaching may be part of the total intervention whose effect is being estimated. Their presence limits attribution to the AI component but does not automatically invalidate the package contrast. In uncontrolled cohorts, the same changes create a stronger causal-attribution problem because there is no concurrent counterfactual.

Outcome selection also matters. A system designed to alter postprandial glucose responses need not outperform an isocaloric comparator for sustained weight loss, but a plausible mechanism cannot be used to explain away every null result after the fact. Mechanistic interpretation is useful only when it remains anchored to the prespecified clinical contrast and its uncertainty. Noninferiority has an additional requirement: the tolerated loss of benefit is defined by the margin and confidence bound, not by visually similar percentages [1–3,12].

A fluent plan can be rated favourably by experts while still containing a nutrient-level error, and a numerically adequate menu can be impractical, culturally unsuitable or unaffordable. Expert rating and computed guideline concordance therefore measure different constructs. The current studies rarely apply both to the same output set, so the evidence does not support a causal claim that experts systematically substitute fluency for correctness [4–6,17–20].

Reproducibility and clinical differentiation should also be separated. Different menus may be equally suitable, while identical outputs may reproduce the same error. Repeated-generation studies identify instability under the tested prompts and model settings but do not establish a literature-wide failure rate. Model version, query date, prompt, retrieval material, sampling configuration and human review process are part of the identity of the intervention; cross-version generalisation requires new evidence.

Safety is further removed from simple content scoring. A guideline-discordant answer, an expert-identified hazard and an observed patient adverse event are different endpoints. A benchmark can expose a plausible mechanism of risk without estimating its incidence in routine use. Conversely, zero harmful responses in a finite test set cannot establish universal safety, including for retrieval-grounded systems [5,6].

Study limitations

The dominant limitation is incomplete evidence provenance. The search and selection record cannot be fully reconciled, some rows contain copied results or conflicting

© The Author(s), 2026. Licensed under CC BY 4.0

classifications, and the post hoc verification prioritised outcome-bearing discrepancies and accessible primary sources. This creates potential source-availability and verification-priority bias. Accordingly, the reported total of 62 studies represents the review's final working evidence inventory; incomplete provenance and study linkage limit independent verification of that count.

Several effects are supported only by abstracts or extraction, and the clinical bias workbook uses a harmonised template that is not a complete RoB 2 or ROBINS-I implementation. The accuracy tool is an unvalidated QUADAS-2-informed adaptation, and two later-added accuracy candidates are absent from the supplied appraisal worksheet. These limitations are why formal outcome-level GRADE certainty was not reissued.

The English-language restriction and uneven timing of supplementary searches may have missed evidence, and the separate source verification was not a comprehensive search update. In the primary studies, heterogeneous comparators, short follow-up, multicomponent interventions, incomplete outcome measurement and inconsistent safety definitions limit transferability. These constraints preclude a precise estimate of the average benefit or harm of AI nutrition systems as a class.

Clinical trials should state whether they estimate a total programme effect, an AI-component effect or the performance of a referral pathway. Component questions require comparators that hold the remainder of care sufficiently similar; pragmatic package trials need not remove clinically appropriate human input merely to appear more automated. Noninferiority studies require a clinically justified margin and transparent reporting of each composite component.

Content evaluations should combine independent clinical review with reproducible nutrient calculations where menus permit them, report guideline versions and patient constraints, and analyse repeated outputs at the profile or question level. Retrieval grounding should be distinguished from deterministic rule enforcement. Prospective patient outcomes and adverse-event measurement remain necessary before improved benchmark scores are interpreted as clinical safety benefits.

Conclusion. AI-enabled personalised nutrition can improve selected glycaemic or weight outcomes in specific settings, but the evidence does not support blanket superiority or blanket ineffectiveness. The contribution attributable specifically to AI depends on the comparator and intervention architecture, and a noninferiority result for a programme-referral strategy should not be relabelled as equality on individual clinical outcomes.

Generated-advice evaluations identify clinically relevant limitations alongside favourable expert ratings and some favourable nutrient comparisons. Expert acceptability, nutritional adequacy, reproducibility and patient safety require separate measurement. Guideline-grounded retrieval is a promising mitigation strategy supported by limited benchmark evidence, not proof of harm-free deployment. A definitive systematic synthesis requires an auditable selection record, complete study linkage and outcome-specific bias assessment.

Conflict of interest. The authors declare no conflict of interest.

Authors' contribution. Conceptualization, methodology, investigation, data curation, formal analysis, funding acquisition, project administration, writing – original draft, writing – review and editing, A.S.; methodology, investigation, data curation, validation, writing – review and editing, F.K.

Funding. The study was carried out within grant funding project IRN AP26102783, "Health management using NutriSteppe-AI technology with the application of data science and neural networks based on the promotion of healthy nutrition and the modernisation of Kazakh national dishes", under contract No. 352/25-27 dated 2 October 2025 between LLP "Kazakh Academy of Nutrition" and the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

Data availability statement. All quantitative evidence used for the substantive synthesis is reported in the article, including source-qualified clinical contrasts, generated-advice evaluations, the flow audit and appraisal summaries. An additional analytical dataset is available from the corresponding author on reasonable request. No new participant-level data were generated.

© The Author(s), 2026. Licensed under CC BY 4.0

Acknowledgments. The authors thank Zhansaya Isabekova for assistance with full-text retrieval.

LIST OF REFERENCES

1. Ben-Yacov O, Godneva A, Rein M, Shilo S, Kolobkov D, Koren N, et al. Personalized postprandial glucose response-targeting diet versus Mediterranean diet for glycemic control in prediabetes. Diabetes Care. 2021;44(9):1980–91. https://doi.org/10.2337/dc21-0162
2. Popp CJ, Hu L, Kharmats AY, Curran M, Berube L, Wang C, et al. Effect of a personalized diet to reduce postprandial glycemic response vs a low-fat diet on weight loss in adults with abnormal glucose metabolism and obesity: a randomized clinical trial. JAMA Netw Open. 2022;5(9):e2233760. https://doi.org/10.1001/jamanetworkopen.2022.33760
3. Mathioudakis N, Lalani B, Abusamaan MS, Alderfer M, Alver D, Dobs A, et al. An AI-powered lifestyle intervention vs human coaching in the Diabetes Prevention Program: a randomized clinical trial. JAMA. 2025;334(23):2079–89. https://doi.org/10.1001/jama.2025.19563
4. Kim DW, Park JS, Sharma K, Velazquez A, Li L, Ostrominski JW, et al. Qualitative evaluation of artificial intelligence-generated weight management diet plans. Front Nutr. 2024;11:1374834. https://doi.org/10.3389/fnut.2024.1374834
5. Ponzo V, Rosato R, Scigliano MC, Onida M, Cossai S, De Vecchi M, et al. Comparison of the accuracy, completeness, reproducibility, and consistency of different AI chatbots in providing nutritional advice: an exploratory study. J Clin Med. 2024;13(24):7810. https://doi.org/10.3390/jcm13247810
6. Parameswaran V, Bernard J, Bernard A, Deo N, Tsung S, Lyytinen K, et al. Evaluating large language models and retrieval-augmented generation enhancement for delivering guideline-adherent nutrition information for cardiovascular disease prevention: cross-sectional study. J Med Internet Res. 2025;27:e78625. https://doi.org/10.2196/78625
7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71
8. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890. https://doi.org/10.1136/bmj.l6890
9. Kharmats AY, Popp C, Hu L, Berube L, Curran M, Wang C, et al. A randomized clinical trial comparing low-fat with precision nutrition-based diets for weight loss: impact on glycemic variability and HbA1c. Am J Clin Nutr. 2023;118(2):443–51. https://doi.org/10.1016/j.ajcnut.2023.05.026
10. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. https://doi.org/10.1136/bmj.i4919
11. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
12. Rein M, Ben-Yacov O, Godneva A, Shilo S, Zmora N, Kolobkov D, et al. Effects of personalized diets by prediction of glycemic responses on glycemic control and metabolic health in newly diagnosed T2DM: a randomized dietary intervention pilot trial. BMC Med. 2022;20(1):56. https://doi.org/10.1186/s12916-022-02254-y

© The Author(s), 2026. Licensed under CC BY 4.0

13. Joshi S, Shamanna P, Dharmalingam M, Vadavi A, Keshavamurthy A, Shah L, et al. Digital twin-enabled personalized nutrition improves metabolic dysfunction-associated fatty liver disease in type 2 diabetes: results of a 1-year randomized controlled study. Endocr Pract. 2023;29(12):960–70. https://doi.org/10.1016/j.eprac.2023.08.016
14. Mohd Dan AS, Linoby A, Kasim SS, Zaki S, Sazali R, Yusoff Y, et al. Personalized AI prompt generator and ChatGPT for weight loss: randomized controlled trial in adults with overweight and obesity. J Technol Behav Sci. 2026. Epub ahead of print. https://doi.org/10.1007/s41347-026-00618-2
15. Nakata Y, Sasai H, Gosho M, Kobayashi H, Shi Y, Ohigashi T, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients. 2022;14(21):4608. https://doi.org/10.3390/nu14214608
16. Sato K, Inoue K, Yamaguchi T, Yagi M, Ebisawa M, Mizuno S, et al. Sodium reduction program incorporating genetic profile and an AI-based app: a randomized clinical trial. JAMA Netw Open. 2025;8(10):e2537540. https://doi.org/10.1001/jamanetworkopen.2025.37540
17. Kenger EB, Özlü Karahan T. Artificial intelligence-generated diet plans for hypertension and dyslipidemia: adherence and nutritional insights. Iran J Public Health. 2025;54(6):1243–51. https://doi.org/10.18502/ijph.v54i6.18902
18. Çelik MN, Karademir E, Gürbüz M. The integration of artificial intelligence-generated diets into the field of dietetics: synergy or threat? Nutr Res. 2026;145:25–36. https://doi.org/10.1016/j.nutres.2025.11.007
19. Ben Lafqih M, Ait Lahbib F, Boutahar K, Azargui H, Elfatimy R, Motaib I. Assessing the quality and guideline-concordance of AI-powered chatbots in nutritional management of diabetes. Diabetes Metab Syndr. 2025;19(12):103355. https://doi.org/10.1016/j.dsx.2025.103355
20. Naja F, Taktouk M, Matbouli D, Khaleel S, Maher A, Uzun B, et al. Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. Eur J Clin Nutr. 2024;78(10):887–96. https://doi.org/10.1038/s41430-024-01476-y
21. Lee YB, Kim G, Jun JE, Park H, Lee WJ, Hwang YC, et al. An integrated digital health care platform for diabetes management with AI-based dietary management: 48-week results from a randomized controlled trial. Diabetes Care. 2023;46(5):959–66. https://doi.org/10.2337/dc22-1929
22. Turnin MC, Gourdy P, Martini J, Buisson JC, Chauchard MC, Delaunay J, et al. Impact of a remote monitoring programme including lifestyle education software in type 2 diabetes: results of the Educ@dom randomised multicentre study. Diabetes Ther. 2021;12:2059–75. https://doi.org/10.1007/s13300-021-01095-x
23. Buisson JC. Nutri-Educ, a nutrition software application for balancing meals, using fuzzy arithmetic and heuristic search algorithms. Artif Intell Med. 2008;42(3):213–27. https://doi.org/10.1016/j.artmed.2007.12.001
24. Mondal A, Naskar A, Roy Choudhury B, Chakraborty S, Biswas T, Sinha S, et al. Evaluating the performance and safety of large language models in generating type 2 diabetes mellitus management plans: a comparative study with physicians using real patient records. Cureus. 2025;17(3):e80737. https://doi.org/10.7759/cureus.80737
25. Aenuginti CS, Mangalagiri T, Gurujala NSG, Polisetty VK, Sharma P, Guntupalli Y, et al. A cross-sectional study evaluating the quality of AI-generated patient education guides on diet and exercise for diabetes, hypertension, and obesity using ChatGPT-4o, Google Gemini 1.5, Claude Sonnet 4, Perplexity, and Grok. Cureus. 2026;18(3):e106221. https://doi.org/10.7759/cureus.106221

© The Author(s), 2026. Licensed under CC BY 4.0

26. Agne I, Gedrich K. Personalized dietary recommendations for obese individuals – a comparison of ChatGPT and the Food4Me algorithm. Clin Nutr Open Sci. 2024;56:192–201. https://doi.org/10.1016/j.nutos.2024.06.001
27. Akduman G, Cebeci A. Guideline-based evaluation of AI-generated vs dietitian-designed meal plans in cardiovascular disease. J Food Nutr Res. 2026;65(2):133. https://doi.org/10.64122/fwdf3373
28. Ataguba G, Orji R. Exploring large language models for personalized recipe generation and weight-loss management. ACM Trans Comput Healthc. 2025;6(2):1–57. https://doi.org/10.1145/3712709
29. Bayram HM, Çelik ZM, Barcın Güzeldere HK. Can artificial intelligence (AI) chatbot tools be used effectively for nutritional management in obesity? Nutr Health. 2025;31(4):1825–34. https://doi.org/10.1177/02601060251329070
30. Bayram HM, Arslan S, Ozturkcan A. Evaluating AI-generated meal plans for simulated diabetes profiles: a guideline-based comparison of three language models. J Eval Clin Pract. 2025;31(7):e70295. https://doi.org/10.1111/jep.70295
31. Chiu E, Young C. 2767-LB: Evaluation of generative AI dietary advice for older, low-income Chinese population with type 2 diabetes in the U.S. Diabetes. 2026;75(Suppl 1):2767-LB. https://doi.org/10.2337/db26-2767-lb
32. Dai R, Cui L, Hu K, An J, Deng N. Using large language models to generate dietary feedback similar to human experts in weight management: experiments on real-world scenario data. Bioengineering (Basel). 2026;13(4):468. https://doi.org/10.3390/bioengineering13040468
33. Suraya Mohd Dan A, Linoby A, Kasim SS, Zaki S, Sazali R, Yusoff Y, et al. Assessing the quality of a personalized prompt generator and AI-chatbot (ChatGPT) for dietary and exercise planning in obese adults using the fuzzy Delphi method. openRxiv. 2025. https://doi.org/10.1101/2025.08.27.25334528
34. Fappa E, Micheli M, Panaretos D, Skordis M, Tsirpanli P, Panoutsopoulos GI. Human vs. AI: assessing the quality of weight loss dietary information published on the web. Information. 2025;16(7):526. https://doi.org/10.3390/info16070526
35. Hart K, Wilson-Barnes S, Stefanidis K, Tsatsou D, Gymnopoulos L, Dimitropoulos K, et al. The suitability of dietary recommendations suggested by artificial intelligence technology via a novel personalised nutrition mobile application. Proc Nutr Soc. 2022;81(OCE1):E37. https://doi.org/10.1017/s0029665122000374
36. Karakas PE, Calik A, Bilen AB, Kandemir K, Alphan ME. Large language models as clinical nutrition decision tools: quantitative bias and guideline deviation in type 2 diabetes meal planning. Healthcare (Basel). 2026;14(6):739. https://doi.org/10.3390/healthcare14060739
37. Katsigiannis G, Panoutsopoulos G, Perrea A, Detopoulou P. Comparison of ChatGPT and dietitians in formulating diet plans and recommendations for patients with cardiometabolic diseases. J Nutr. 2026;156(8):101667. https://doi.org/10.1016/j.tjnut.2026.101667
38. Kaya Kaçar H, Kaçar ÖF, Avery A. Diet quality and caloric accuracy in AI-generated diet plans: a comparative study across chatbots. Nutrients. 2025;17(2):206. https://doi.org/10.3390/nu17020206
39. Onay T, Bekar D, Çoban E, Doğan N, Günşen U. Artificial intelligence in clinical nutrition: a descriptive comparison of ChatGPT- and dietitian-planned diets for chronic disease scenarios. J Hum Nutr Diet. 2025;38(5):e70135. https://doi.org/10.1111/jhn.70135

© The Author(s), 2026. Licensed under CC BY 4.0

40. Ozlu Karahan T, Kenger EB, Yilmaz Y. Artificial intelligence-based diets: a role in the nutritional treatment of metabolic dysfunction-associated steatotic liver disease? J Hum Nutr Diet. 2025;38(2):e70033. https://doi.org/10.1111/jhn.70033
41. Ozlu Karahan T, Kenger EB. ChatGPT-4o for weight management: comparison of different diet models. Food Sci Nutr. 2025;13(7):e70639. https://doi.org/10.1002/fsn3.70639
42. Pan E, Wu G, Lee H, Sidhu S, Sidhu A, Ashok A, et al. SUN-808 ADCES7 guidelines and AI chatbots: do they help our patients? J Endocr Soc. 2025;9(Suppl 1):bvaf149.1314. https://doi.org/10.1210/jendso/bvaf149.1314
43. Papastratis I, Stergioulas A, Konstantinidis D, Daras P, Dimitropoulos K. Can ChatGPT provide appropriate meal plans for NCD patients? Nutrition. 2024;121:112291. https://doi.org/10.1016/j.nut.2023.112291
44. Ponzo V, Goitre I, Favaro E, Merlo FD, Mancino MV, Riso S, et al. Is ChatGPT an effective tool for providing dietary advice? Nutrients. 2024;16(4):469. https://doi.org/10.3390/nu16040469
45. Sloma Krzeslak M, Kowalski O. Evaluation of the compliance of diet plans and nutritional advice generated by artificial intelligence with guidelines for cardiac patients. Eur J Cardiovasc Nurs. 2024;23(Suppl 1):zvae098.120. https://doi.org/10.1093/eurjcn/zvae098.120
46. Sun H, Zhang K, Lan W, Gu Q, Jiang G, Yang X, et al. An AI dietitian for type 2 diabetes mellitus management based on large language and image recognition models: preclinical concept validation study. J Med Internet Res. 2023;25:e51300. https://doi.org/10.2196/51300
47. Suraya Mohd Dan A, Linoby A, Shahlan Kasim S, Zaki S, Sazali R, Yusoff Y, et al. Validation of a personalized AI prompt generator (NExGEN-ChatGPT) for obesity management using fuzzy Delphi method. Biol Methods Protoc. 2025;10(1):bpaf085. https://doi.org/10.1093/biomethods/bpaf085
48. Talay L, Lagesen L, Yip A, Vickers M, Ahuja N. ChatGPT-4o and 4o1 preview as dietary support tools in a real-world medicated obesity program: a prospective comparative analysis. Healthcare (Basel). 2025;13(6):647. https://doi.org/10.3390/healthcare13060647
49. Yang E, Garcia T, Williams HG, Kumar B, Ramé M, Rivera E, et al. A behavioral science-informed agentic workflow for personalized nutrition coaching: development and validation study. JMIR Form Res. 2025;9:e75421. https://doi.org/10.2196/75421
50. Bolat Yilmaz A, Kenger EB, Ozlu Karahan T, Saglam D, Bas M. Guideline compliance of artificial intelligence–generated diet plans after bariatric surgery: a cross-sectional simulation comparing ChatGPT-4o, DeepSeek and Grok-3. Nutrients. 2025;17(24):3957. https://doi.org/10.3390/nu17243957
51. Zhang Y, Liu XJ, Hu Q, Galaviz KI, Casanova IG, Colditz J, et al. Response consistency of ChatGPT-4o for type 2 diabetes nutrition and physical-activity recommendations: a pilot NLP-based assessment of GPT outputs. openRxiv. 2026. https://doi.org/10.64898/2026.06.23.26356399
52. Zhou S, Wu H, Zhang Y, Zhong Q, Diao Y, Fang H, et al. Automated quality assessment of dietary weight loss videos: a likelihood-based large language model framework. SSRN. 2025. https://doi.org/10.2139/ssrn.5312906
53. Forman EM, Hagerman CJ, Hong AE, Huang Z, McCausland HC, Sun JH, et al. Feasibility study of an artificial intelligence chatbot-powered cognitive-behavioral lifestyle modification program. Ann Behav Med. 2026;60(1):kaag038. https://doi.org/10.1093/abm/kaag038

© The Author(s), 2026. Licensed under CC BY 4.0

54. Graham SA, Pitter V, Hori JH, Stein N, Branch OH. Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. Digit Health. 2022;8:20552076221130619. https://doi.org/10.1177/20552076221130619
55. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating glycemic control in patients of South Asian origin with type 2 diabetes using a digital therapeutic platform: analysis of real-world data. J Med Internet Res. 2021;23(3):e17908. https://doi.org/10.2196/17908
56. Kumbara AB, Iyer AK, Green CR, Jepson LH, Leone K, Layne JE, et al. Impact of a combined continuous glucose monitoring-digital health solution on glucose metrics and self-management behavior for adults with type 2 diabetes: real-world, observational study. JMIR Diabetes. 2023;8:e47638. https://doi.org/10.2196/47638
57. Lockwood KG, Kulkarni PR, Paruthi J, Buch LS, Chaffard M, Schitter EC, et al. Evaluating a new digital app-based program for heart health: feasibility and acceptability pilot study. JMIR Form Res. 2024;8:e50446. https://doi.org/10.2196/50446
58. Sakane N, Suganuma A, Domichi M, Sukino S, Abe K, Fujisaki A, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth. 2023;11:e43236. https://doi.org/10.2196/43236
59. Shamanna P, Dharmalingam M, Sahay R, Mohammed J, Mohamed M, Poon T, et al. Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the Digital Twin Precision Treatment Program. Sci Rep. 2021;11(1):14892. https://doi.org/10.1038/s41598-021-94339-6
60. Shamanna P, Erukulapati RS, Shukla A, Shah L, Willis B, Thajudeen M, et al. One-year outcomes of a digital twin intervention for type 2 diabetes: a retrospective real-world study. Sci Rep. 2024;14(1):25478. https://doi.org/10.1038/s41598-024-76584-7
61. Shamanna P, Joshi S, Shah L, Dharmalingam M, Saboo B, Mohammed J, et al. Type 2 diabetes reversal with digital twin technology-enabled precision nutrition and staging of reversal: a retrospective cohort study. Clin Diabetes Endocrinol. 2021;7(1):21. https://doi.org/10.1186/s40842-021-00134-7
62. Shamanna P, Saboo B, Damodharan S, Mohammed J, Mohamed M, Poon T, et al. Reducing HbA1c in type 2 diabetes using digital twin technology-enabled precision nutrition: a retrospective analysis. Diabetes Ther. 2020;11(11):2703–14. https://doi.org/10.1007/s13300-020-00931-w
63. Sinha R, Kachru D, Ricchetti RR, Singh-Rambiritch S, Muthukumar KM, Singaravel V, et al. Leveraging genomic associations in precision digital care for weight loss: cohort study. J Med Internet Res. 2021;23(5):e25401. https://doi.org/10.2196/25401
64. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes. 2017;2(2):e28. https://doi.org/10.2196/diabetes.8590
65. Swamy P. Personalized nutrition prescription generated by AI platform improves cardiometabolic health. Int J Life Sci Biotechnol Pharma Res. 2025;14:1554–8.
66. Veluvali A, Dehghani Zahedani A, Hosseinian A, Aghaeepour N, McLaughlin T, Woodward M, et al. Impact of digital health interventions on glycemic control and weight management. NPJ Digit Med. 2025;8(1):20. https://doi.org/10.1038/s41746-025-01430-7
67. Wang J, Faruqui SHA, Alaeddini A, Du Y, Li S, Yang Y, et al. Human-in-the-loop AI predictive digital twin to extend virtual precision diabetes care between visits. NPJ Health Syst. 2026;3(1). https://doi.org/10.1038/s44401-026-00118-8

© The Author(s), 2026. Licensed under CC BY 4.0

68. Xin M, Yao Y, Huang P, Li Q. Application study of an artificial intelligence and big data-based personalized chronic disease management model for diabetes patients. Front Public Health. 2026;14:1735295. https://doi.org/10.3389/fpubh.2026.1735295

Information about authors
@Khozhamkul Fariza, Researcher, Kazakh Academy of Nutrition, Almaty, Kazakhstan, khozhamkul.fariza@gmail.com, https://orcid.org/0000-0003-4630-319X.
Sharman Almaz, PhD, President of Kazakh Academy of Nutrition, Almaty, Kazakhstan, almaz@sharman.pro, https://orcid.org/0000-0003-3977-3636.
Авторлар туралы мәліметтер
@Хожамқұл Фариза, ғылыми қызметкер, Қазақ тағамтану академиясы, Алматы, Қазақстан, khozhamkul.fariza@gmail.com, https://orcid.org/0000-0003-4630-319X.
Шарман Алмаз Төрегелдіұлы, PhD, Қазақ тағамтану академиясы президенті, Алматы, Қазақстан, almaz@sharman.pro, https://orcid.org/0000-0003-3977-3636.
Сведения об авторах
@Хожамқұл Фариза, научный сотрудник, Казахская академия питания, Алматы, Казахстан, khozhamkul.fariza@gmail.com, https://orcid.org/0000-0003-4630-319X.
Шарман Алмаз Торегельдиевич, PhD, президент Казахской академии питания, Алматы, Казахстан, almaz@sharman.pro, https://orcid.org/0000-0003-3977-3636.

КАРДИОМЕТАБОЛИКАЛЫҚ ДЕНСАУЛЫҚ ҮШІН ЖАСАНДЫ ИНТЕЛЛЕКТ КӨМЕГІМЕН ДЕРБЕСТЕНДІРІЛГЕН ТАМАҚТАНУ: КЛИНИКАЛЫҚ ӘСЕРЛЕР МЕН ЖАСАЛҒАН ТАМАҚТАНУ ҰСЫНЫМДАРЫНА ЖҮЙЕЛІ ШОЛУ

Ф. ХОЖАМҚҰЛ, А.Т. ШАРМАН

Қазақ тағамтану академиясы, Алматы, Қазақстан

Түйіндеме

Кіріспе. Жасанды интеллект (ЖИ) қолданылатын тамақтану диетаны болжамды дербестендіруді, өмір салтын басқарудың цифрлық бағдарламаларын және генерацияланған тамақтану ұсынымдарын қамтиды.

Мақсаты. Кардиометаболикалық тәуекелі бар ересектерде ЖИ қолданылатын дербестендірілген тамақтанудың клиникалық нәтижелерін, сондай-ақ генерацияланған ұсынымдардың дәлдігі мен қауіпсіздігін бағалау.

Материалдар мен әдістер. Бес библиографиялық дерекқор мен екі клиникалық зерттеулер тізілімінде олардың құрылған кезінен 2026 жылғы тамызға дейін іздеу жүргізіліп, Semantic Scholar, medRxiv және таңдалған веб-дереккөздерде қосымша іздеу орындалды. Жарамды араласуларда ЖИ, машиналық оқыту, үлкен тілдік модель немесе диалогтық агент жеке диета, мәзір немесе тағам ұсынымдарын қалыптастырды. Клиникалық нәтижелер мен генерацияланған ұсынымдардың дәлдігі бөлек синтезделді. Жүйелі қателік тәуекелі RoB 2, ROBINS-I және QUADAS-2 негізінде бейімделген құралмен бағаланды; дәлелдемелердің сенімділігі GRADE арқылы анықталды.

Нәтижелер. Соңғы жұмыс дәлелдемелер жинағына 62 зерттеу (64 жарияланым) кірді: 26-сы клиникалық, 36-сы дәлдік доменінде. Машиналық оқытуға негізделген диета алты айдан кейін жерорта теңізі диетасымен салыстырғанда HbA1c деңгейін 0,08 пайыздық тармаққа көбірек төмендетті (95% СИ −0,14-тен −0,02-ге дейін), ал басқа зерттеуде дербестендірілген диета майы аз диетамен салыстырғанда алты айлық салмақ

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

жоғалтуды жақсартпады (бастапқы салмақтан айырмашылық +1,05 пайыздық тармақ; 95% СИ −0,40-тан 2,50-ге дейін). Дербестендірілген LLM промптингі қолмен жасалған ChatGPT промптингіне қарағанда салмақ жоғалтуды жақсартты, ал натрийді азайтуға арналған ЖИ бағдарламасы тұз тұтыну мен артериялық қысымды жақсартпады. 368 қатысушысы бар зерттеу ЖИ басқаратын және адам-жаттықтырушы жүргізетін диабеттің алдын алу бағдарламаларын салыстырғанда алдын ала белгіленген кем түспеушілік критерийіне жетті. Генерацияланған ұсынымдар сарапшылардан оң баға алғанымен, нутриенттер мен нұсқаулықтарға сәйкессіздіктер анықталды, күрделі профильдер қиынырақ болды.

Қорытынды. Кейбір араласулар жекелеген нәтижелерді жақсартты, бірақ ЖИ қолданылатын тамақтанудың жалпы пайдасы дәлелденбеді. Бағдарламалық әсерлер ЖИ компонентінің жеке үлесін ажыратпайды, ал оң мазмұндық бағалар нутритивтік жеткіліктілік пен пациент қауіпсіздігін дәлелдемейді.

Түйінді сөздер: дербестендірілген тамақтану, жасанды интеллект, үлкен тілдік модельдер, кардиометаболикалық денсаулық, тамақтану ұсынымдары, дәлелдемелерді синтездеу.

ПЕРСОНАЛИЗИРОВАННОЕ ПИТАНИЕ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ КАРДИОМЕТАБОЛИЧЕСКОГО ЗДОРОВЬЯ: СИСТЕМАТИЧЕСКИЙ ОБЗОР КЛИНИЧЕСКИХ ЭФФЕКТОВ И ГЕНЕРИРУЕМЫХ РЕКОМЕНДАЦИЙ ПО ПИТАНИЮ

Ф. ХОЖАМҚҰЛ, А.Т. ШАРМАН

Казахская академия питания, Алматы, Казахстан

Аннотация

Введение. Питание с искусственным интеллектом (ИИ) включает предиктивную персонализацию рациона, цифровые программы управления образом жизни и сгенерированные рекомендации по питанию.

Цель. Оценить клинические результаты персонализированного питания с ИИ у взрослых с кардиометаболическим риском, точность и безопасность сгенерированных рекомендаций.

Материалы и методы. Поиск проводили в пяти библиографических базах данных и двух реестрах клинических исследований с момента создания до августа 2026 года, дополнив его поиском в Semantic Scholar, medRxiv и веб-источниках. Включали цифровые инструменты, в которых ИИ, машинное обучение, большая языковая модель или диалоговый агент формировали индивидуальные рекомендации по рациону, меню или приемам пищи. Клинические результаты и точность сгенерированных рекомендаций синтезировали раздельно. Риск систематической ошибки оценивали по RoB 2, ROBINS-I и адаптированному инструменту QUADAS-2, а достоверность доказательств по GRADE.

Результаты. Итоговый рабочий массив доказательств включал 62 исследования (64 публикации): 26 в клиническом домене и 36 в домене точности. Через шесть месяцев диета под управлением машинного обучения снижала HbA1c на 0,08 процентного пункта больше, чем средиземноморская диета (95% ДИ от −0,14 до −0,02), тогда как в другом исследовании персонализированная диета не улучшила снижение веса по сравнению с диетой с низким содержанием жиров (разница +1,05 процентного пункта от исходного веса; 95% ДИ от −0,40 до 2,50). Персонализированный промптинг для LLM улучшил

© The Author(s), 2026. Licensed under CC BY 4.0

ISSN 2524 – 0684 **e-ISSN 2524 – 0692**

снижение веса по сравнению с ручным промптингом ChatGPT, а программа снижения потребления натрия на основе ИИ не улучшила потребление соли или артериальное давление. Исследование с участием 368 человек достигло заранее установленного критерия неменьшей эффективности при сравнении направления в программу профилактики диабета под управлением ИИ и человека-инструктора. Исследования сгенерированных рекомендаций выявили благоприятные экспертные оценки наряду с несоответствиями по нутриентам и руководствам, сложные профили оказались более трудными.

Заключение. Отдельные вмешательства улучшали конкретные результаты, но польза питания с ИИ как класса не установлена. Эффекты программ не изолируют вклад ИИ, а положительные оценки содержания не подтверждают нутритивную адекватность или безопасность пациентов.

Ключевые слова: персонализированное питание, искусственный интеллект, большие языковые модели, кардиометаболическое здоровье, рекомендации по питанию, синтез доказательств.

© The Author(s), 2026. Licensed under CC BY 4.0