Muhamad Noor Mohamed . Raja Nurul Jannat Raja Hussain .
Mardiana Mazaulan . Noor Azila Azreen Md Radzi .
Nurul Ain Abu Kasim . Nur Hani Syazwani Bakri .
Ummi Khaltum Mohd Mokhtar . Mohd Aizzat Adnan .
Editors

Proceedings of the 1st International Summit Conference on Exercise Science, Sports Management, Outdoor Recreation, and Physical Education, ExSPORT 2024, 28th - 29th August, Malaysia

Exporting Research Insights to Practical Applications in Sports
Turning Challenges into Opportunities

ORGANIZED BY

IN COLLABORATION

SUPPORTED BY

Assessing the Quality of Diet and Exercise Plans Generated by AI Chatbots: A Preliminary Study Using the NExGEN Prompt Generator System

Azwa Suraya Mohd Dan, Adam Linoby*, Sazzli Shahlan Kasim, Siti Aida Lamat, Sufyan Zaki, and Razif Sazali.

Abstract | Artificial intelligence (AI) chatbots like ChatGPT are increasingly used in obesity research to track diets, activity, and energy expenditure. However, its effectiveness in diet and exercise planning depends on the precision and completeness of user inputs. This study evaluates the quality of ChatGPT output when combined with the newly developed diet and exercise prompt generator system, NExGEN. A cohort of obese participants (n = 18) was enlisted to contribute interpersonal data for the NExGEN prompt generator. Utilising ChatGPT-4, this data informed the creation of 36he36ious36s36d weekly dietary and exercise plans. Accredited professionals (n = 16) conducted a blind evaluation of these plans by grading the quality and validity of the NExGEN-ChatGPT responses using the DISCERN and content validity index (CVI), respectively. The evaluators graded the NExGEN-ChatGPT responses as bottom tier 2.2% of the time, middle tier 16.3% of the time, and top tier 81.5% of the time. The CVIs score was ≥ 80% with a correlation coefficient between 0.89 – 0.99, and overall Cronbach's alpha score at 0.798. This study demonstrates that integrating ChatGPT with the NExGEN system effectively generates high-quality diet and exercise plans for obese individuals, as evidenced by favourable quality and validity assessments by professionals.

Keywords: *Artificial intelligence, obesity, physical activity, weight loss, weight management, nutrition.*

A.S., Mohd Dan, A., Linoby* (✉), S.A., Lamat, and R., Sazali.
Faculty of Sports Science and Recreation, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, Malaysia.
**Corresponding author: linoby@uitm.edu.my*

S.S., Shahlan.
Faculty of Medicine, Universiti Teknologi MARA, Cawangan Sungai Buloh, Selangor Campus, Selangor, Malaysia.

S., Zaki.
Faculty of Sports Science and Recreation, Universiti Teknologi MARA, Shah Alam Campus, Selangor, Malaysia.

I. Introduction

The release of ChatGPT, an artificial intelligence (AI) chatbot by OpenAI, has impacted both the public and healthcare professionals [1]. The chatbot's effectiveness depends on the precision of user prompts [2]. This study evaluated the quality of personalized diet and exercise plans generated through the combined use of ChatGPT and a novel prompt generator tool called NExGEN.

II. Methods

A cohort of 48 obese participants was recruited to evaluate the personalized exercise and dietary plans based on NExGEN and ChatGPT. Twelve accredited professionals in nutrition and exercise blindly assessed the plans using the DISCERN tool [3] to ensure an unbiased evaluation of the AI-generated plans.

III. Results and Discussion

In this study, the internal consistency of the DISCERN tool was satisfactory, with a Cronbach's α of 0.89. The inter-rater agreement for DISCERN items varied, ranging from $\kappa = 0.182$ for item 14 to $\kappa = 0.90$ for item 5. The index of agreement for the overall quality rating (item 16) was $\kappa = 0.71$. The intraclass correlation coefficient for total DISCERN scores was 0.89 ($p < 0.001$), indicating a high level of agreement. Higher agreement levels were observed for more objective items, such as "Is the information in the plan consistent with current knowledge in exercise and nutrition science?" (item 2, $\kappa = 0.83$) and "Is the plan updated to reflect current guidelines and recommendations?" (item 4, $\kappa = 0.84$), which aligns with previous research findings [4]. Conversely, lower agreement levels were noted for more subjective items, such as "Does the plan disclose any potential biases or conflicts of interest?" (item 8, $\kappa = 0.34$) and "Does the plan include clear markers for assessing progress?" (item 14, $\kappa = 0.18$) (see Table 1).

TABLE 1
SUMMARY OF AGREEMENT BETWEEN RATERS FOR EACH ITEM OF THE DISCERN

Item		Weighted K	95% CI	Level Of Agreement
1	Does the plan specify the sources of information used to create it?	0.398	(0.364, 0.432)	Fair
2	Is the information in the plan consistent with current knowledge in exercise and nutrition science?	0.827	(0.775, 0.882)	Almost Perfect
3	Does the plan provide both positive and negative aspects of the recommended regimen?	0.583	(0.547, 0.619)	Moderate
4	Is the plan updated to reflect current guidelines and recommendations?	0.841	(0.782, 0.884)	Almost Perfect
5	Does the plan consider the subject's personalized information?	0.902	(0.878, 0.926)	Almost Perfect
6	Is the plan comprehensive, covering all important aspects of exercise and nutrition?	0.775	(0.749, 0.801)	Substantial
7	Does the plan provide detailed and specific instructions for each recommended activity and meal plan?	0.738	(0.704, 0.772)	Substantial
8	Does the plan disclose any potential biases or conflicts of interest?	0.342	(0.308, 0.376)	Fair
9	Are the exercise and dietary goals realistically achievable for the given participant?	0.778	(0.729, 0.818)	Substantial
10	Is the plan tailored to the participant's specific needs and preferences?	0.799	(0.738, 0.845)	Substantial
11	Does the plan include instructions to avoid injuries and adverse health effects?	0.454	(0.420, 0.488)	Moderate
12	Is the information within the plan clearly presented and understandable, making it easy for someone to follow without additional explanations?	0.709	(0.675, 0.743)	Substantial
13	Is the plan viable in terms of resources (time, equipment, and food availability)?	0.452	((0.418, 0.486)	Moderate
14	Does the plan include clear markers for assessing progress?	0.182	(0.148, 0.216)	Slight
15	Does the plan include easy-to-understand guidance or annotations that help explain how to implement and follow the recommended routines effectively?	0.757	(0.703, 0.806)	Substantial
16	How would you rate the overall quality and reliability of this exercise and dietary plan?	0.708	(0.658, 0.763)	Substantial

*Level of agreement as indicated by Landis and Koch.

IV. CONCLUSIONS

This study is the first to evaluate the quality of information generated by combining a prompt generator (NExGEN) with ChatGPT for weight management in obese individuals. Using the DISCERN tool, the research demonstrated satisfactory internal consistency in the information generated by the NExGEN-ChatGPT framework. Inter-rater agreement varied across DISCERN items, with higher agreement for objective items such as consistency with current knowledge in exercise and nutrition science and inclusion of updated guidelines. Lower agreement was noted for subjective items like potential biases and progress markers. These findings highlight the potential of combining prompt generators like NExGEN with AI systems to produce high-quality health information while emphasizing the need for improvement in subjective evaluations. The study also revealed high overall quality ratings and a robust intraclass correlation coefficient, indicating reliable content. Further research is necessary to broaden the applicability of these findings across various populations and health contexts.

ACKNOWLEDGMENTS: We would like to thank the Faculty of Sports Science & Recreation, Universiti Teknologi MARA Cawangan Negeri Sembilan, Seremban Campus, for their support in this study.

REFERENCES

[1] E. Loh, "ChatGPT and generative AI chatbots: challenges and opportunities for science, medicine, and medical leaders," Bmj Leader, leader-2023, 2023. [Online]. Available: https://doi.org/10.1136/leader-2023-000797.

[2] J. M. Perkel, "Six tips for better coding with ChatGPT," Nature, vol. 618, no. 7964, pp. 422-423, 2023. Doi: 10.1038/d41586-023-01833-0.

[3] R. M. McKearney, R. C. Mackinnon, M. C. Smith, and R. J. Baker, "Tinnitus information online – does it ring true?" The Journal of Laryngology & Otology, vol. 132, no. 11, pp. 984-989, 2018. Doi: 10.1017/s0022215118001792.

[4] F. Sadeghi-Ghotbabadi, E. Shakibazadeh, N. Omidvar, F. Mortazavi, and F. Kolahdooz, "Consensus development on the essential competencies for Iranian public health nutritionists," Public Health Nutrition, vol. 18, no. 4, pp. 752-758, 2014. Doi: 10.1017/s1368980014000895.