



Nutrition Society Congress 2024, 2–5 July 2024

Using artificial intelligence to estimate the nutritional content of meal photos: an evaluation of ChatGPT-4

C. O'Hara¹, G. Kent¹, A.C. Flynn¹, E.R. Gibney^{2,3} and C.M. Timon¹

¹School of Population Health, Royal College of Surgeons in Ireland, Dublin, Ireland

²UCD Institute of Food and Health, University College Dublin, Dublin, Ireland

³School of Agriculture and Food Science, University College Dublin, Dublin, Ireland

Dietary intake assessment is an essential part of nutrition research and practice, with the use of digital technology now well established⁽¹⁾ and artificial intelligence (AI) in the form of image recognition readily available in research and commercial settings⁽²⁾. Recent advances in large language models (LLMs), such as ChatGPT, allow computers to converse in a human-like way providing text responses to typed queries. No studies, however, have utilised both the LLM and image recognition components of ChatGPT-4 to evaluate its accuracy to estimate nutritional content of meals.

The aim of this study was to evaluate the accuracy of the ChatGPT-4 LLM and image recognition model in estimating the nutritional content of meals.

Thirty-eight meal photographs with known nutritional content (from McCance and Widdowson's Composition of Foods) were uploaded to ChatGPT, and it was asked to provide point estimates for each of the meals for each of the following: energy (kcal), protein (g), total carbohydrate (g), dietary fibre (g), total sugar (g), total fat (g), saturated fat (g), monounsaturated fat (g), polyunsaturated fat (g), calcium (mg), iron (mg), sodium (mg), potassium (mg), vitamin D (mcg), folate (mcg), and vitamin C (mg). Comparisons were made between ChatGPT estimates and those from McCance and Widdowson using the Wilcoxon signed rank test, percent difference, Spearman's correlation, and cross-classification of quartiles. Interpretation of statistical measures was based on Lombard et al.⁽³⁾.

For estimating the content of meals, differences ($p < 0.05$) existed between the methods for 11 of the 16 nutrients, and 12 nutrients had a percent difference of $>10\%$, indicating poor agreement for most nutrients. ChatGPT underestimated 15 of the 16 nutrients. Conversely, when considering the ranking of meals, all nutrients had correlation coefficients which indicated good ($r_s \geq 0.50$)(11 of 16) or acceptable ($0.20 < r_s < 0.49$)(5 of 16) agreement. In the cross-classification of quartiles, $\geq 50\%$ of meals were classified into the same quartile by both methods for 9 nutrients and 10% of meals were classified into opposite quartiles for 14 nutrients, indicating good agreement. ChatGPT also provided caveats regarding its estimations such as "the caloric estimate assumes the butter is spread thinly" and "cornflakes can often be fortified with vitamins and minerals [...] and exact content could also vary based on the brand of cornflakes".

ChatGPT showed poor agreement for estimating the nutritional composition of meal photographs for most nutrients, but the correlation and cross-classification of quartiles indicate good ability to rank meal photographs according to nutritional composition. Further research is required with a wider range of meals and in real-world settings. Future work should consider the training of language models using high-quality nutrition data to improve accuracy and maximise the potential for their use in dietary assessment.

Acknowledgments

The authors thank Dr. Laura Kirwan of Nutritics for her assistance with the Nutritics software for dietary assessment.

References

1. Eldridge AL, Piernas C, Illner AK *et al.* (2018) *Nutrients* **11**(1), 55.
2. Boushey CJ, Spoden M, Zhu FM *et al.* (2017) *Proc Nutr Soc* **76**(3), 283–94.
3. Lombard MJ, Steyn NP, Charlton KE & Senekal M. (2015) *Nutr J* **14**(1), 40.