Potential and Pitfalls of ChatGPT and Natural-Language Artificial Intelligence Models for Diabetes Education

Gerald Gui Ren Sng,[1] *Joshua Yi Min Tung,*[2] *Daniel Yan Zheng Lim,*[3] *and Yong Mong Bee*[1]

Diabetes Care 2023;46:e103–e105 | https://doi.org/10.2337/dc23-0197

Diabetes self-management and education (DSME) is an integral part of diabetes care and has been shown to improve glycemic control, reduce complications, and increase quality of life (1). The traditional model in which clinicians and diabetes educators share responsibility for patient education faces challenges such as reduced access to care during the pandemic and a shortage of trained educators.

Artificial intelligence (AI) solutions are increasingly recognized to have a strong use case in DSME (2). Smart conversational agents ("chatbots") have shown potential as tools for direct patient engagement and education (3). While previous generations of chatbots delivered structured output based on preset queries and responses, modern natural-language AI models are designed to accept unstructured or nonstandardized inputs and provide human-like responses. These models draw on a large repository of human-generated textual content to produce responses statistically likely to match the query. While chatbots may be able to augment patient care by providing on-demand answers to patient questions, they are based on language patterns rather than objective databases and may provide patients with authoritative-sounding information that is inaccurate.

ChatGPT is a chatbot developed by OpenAI based on the GPT3 large language model. It is readily accessible by the general public and has gained popular traction. It has recently been shown to be able to pass the U.S. Medical Licensing Examination (4). We aimed to explore the quality and veracity of DSME advice provided by ChatGPT.

We instructed ChatGPT to answer a series of questions regarding diabetes self-management. Thereafter, we presented it with common questions on four domains of DSME: first, diet and exercise; second, hypoglycemia and hyperglycemia education; third, insulin storage; and last, insulin administration. The interaction was repeated in an unstructured manner to simulate natural real-world patient communication, with both versions evaluated for consistency and reliability.

ChatGPT was able to answer all the questions posed. It demonstrated a systematic approach to answering questions, often providing instructions in clear point form. In general, the responses were concise and well-organized and used layperson's terms. Jargon, where present, was explained in parentheses. Almost all responses recommended "consult[ing] your health care team." The content of responses was consistent between iterations. Strengths and limitations noted from these simulated interactions with ChatGPT, together with a selected excerpt of the relevant prompts, are summarized in Table 1.

Certain inaccuracies were noted. ChatGPT did not recognize that insulin analogs should be stored at room temperature once opened (with the unprompted exception of insulin lispro) (5). ChatGPT was inflexible in certain scenarios, such as recommending diet plans. It occasionally required additional prompts to generate a full list of instructions for insulin administration.

ChatGPT generally performed well in generating easily understood and accurate responses to questions about diabetes care. Theoretically, stochasticity can be an issue with language models, as using statistical probability to generate replies can randomly lead to inconsistencies across iterations. However, this was not observed here. Implementing large language models may be a way to offload some burdens of basic diabetes patient education, freeing trained providers to take on higher-complexity DSME duties.

However, it is important to recognize that these models are limited by the data sets they are trained on. In this case, ChatGPT is trained on a general information database and not a database specific to medical information, which may explain the lack of nuance in differentiating types of insulin or recognizing that blood glucose values may have different units. Moreover, it is trained on information from before 2021, and responses may not capture new evidence that has been published since then.

Potential factual inaccuracies pose a strong safety concern. ChatGPT does not source information in the traditional sense and cannot evaluate its output's reliability. The model generates output based on the likelihood of phrases and sentences going together rather than by reference to specific knowledge bases. It is therefore prone to "hallucination," presenting inaccurate or untruthful information in a

[1]*Department of Endocrinology, Singapore General Hospital, Singapore*
[2]*Department of Urology, Singapore General Hospital, Singapore*
[3]*Health Services Research Unit, Singapore General Hospital, Singapore*

Corresponding author: Gerald Gui Ren Sng, gerald.sng@mohh.com.sg

Received 7 February 2023 and accepted 15 February 2023

Table 1—Summary of strengths and limitations identified in interaction with ChatGPT

Selected prompts	Strengths	Limitations
Domain 1: diet and exercise		
"What kind of diet is best for diabetes?"	Offered systematic and well-considered lists of recommended diets based on American Diabetes Association guidelines	
"Could you give me an example of a meal plan?"	Gave a sample meal plan including multiple options for each meal	Meal structure is generic and suggests a meal plan with snacks between meals
"Is it ok to snack?" "Is snacking ok if I am on insulin as well?"		Unable to differentiate between patients on basal or premixed insulin regimens (in which snacking may be better tolerated) and patients on multiple-daily-injection regimens (in which snacking is more likely to lead to hyperglycemia due to uncovered carbohydrate intake)
"I have heard a lot about the ketogenic diet. Is it safe for me if I have diabetes?"	Appropriately advised caution with ketogenic diets in patients with type 1 diabetes mellitus	
"You suggested a diet plan earlier for diabetes. It does seem quite catered to a Western diet. Might you have any alternative suggestions for an East Asian setting?"	With prompting, provided modified sample meal plan contextualized to East Asia	
"How much exercise should I do a day?"	Gave ordered point-form list of instructions and precautions when exercising	
"Can I exercise while on insulin?"	With prompting, gave tailored instructions to suit a patient on insulin therapy	
"Can exercise increase my blood sugar?"	With prompting, described differences in glycemic changes with different forms of exercise	
Domain 2: hypoglycemia and hyperglycemia education		
"Can I find out more about hypoglycemia?"	Delivered concise advice about hypoglycemia, including causes, symptoms, prevention, treatment, and long-term management	
"My blood sugar is 25, what should I do?"	Provided simple list of steps for hypoglycemia correction	Assumed that blood glucose readings were in mg/dL when presented with integer readings without units; hypoglycemia correction was advised in a scenario with a submitted reading of "25," which could worsen hyperglycemia if the unit of measurement was mmol/L instead of mg/dL
"I am experiencing some of these symptoms like sweating and shaking but when I check, my blood sugar is not that low (around 5 mmol/L). Why is that?"		Wrongly classified pseudohypoglycemia as hypoglycemia unawareness: "If you are experiencing symptoms of hypoglycemia (sweating and shaking) but your blood sugar level is not that low (around 5 mmol/L), it's possible that you are experiencing a condition called hypoglycemic unawareness."
"What if my blood sugar is low (3 mmol/L) but I feel fine. Is that ok?"	Still suggested that hypoglycemia be corrected with simple list of steps	
"How high does my blood sugar need to be before I go to the hospital?"	Suggested that user review symptoms in conjunction with blood glucose reading to decide if escalation of care is required	
"If my blood sugar level is high, should I jab more insulin?"	Advised that increases in insulin doses in response to hyperglycemia be made in consultation with health care providers	
Domain 3: insulin storage		
"Can I ask more about insulin storage?"	Provided concise list of guidelines for insulin storage	
"Can I bring my vial out with me when I am outside?"	Emphasized that insulin should be stored at 2–8°C and suggested use of insulated bag with cooling or heating pads. Also suggested avoiding direct sunlight.	

Continued on p. e105

Table 1—Continued

Selected prompts	Strengths	Limitations
"Should I keep my insulin pens in the fridge too, even after I open them?"		Did not differentiate between storage of regular insulin and insulin analog pens, although manufacturers for insulin analog pens suggest that room temperature storage is required once pens are open: "Room temperature storage is not recommended for all types of insulin."
"Can I bring insulin with me when I travel?"	Gave clear list of instructions to follow when traveling with insulin	
"My insulin is not expired but looks a little cloudy. Can I still use it?"	Suggested consultation with health care provider and to err on side of caution and open a new vial when faced with a cloudy vial of insulin	
Domain 4: insulin administration		
"What are the steps to using an insulin pen?" "Do I need to prime the pen?"	Provided an ordered simple list of steps in point form	Did not include priming in list of steps required to use an insulin pen until prompted to do so
"I injected my insulin 1 hour before my meal and now I feel giddy. What should I do?"	Accurately diagnosed hypoglycemia as the most likely cause of giddiness postinjection and suggested steps to correct hypoglycemia	
"If I missed a dose of insulin, should I double the next dose?" "What if I am having a bigger meal, can I increase my dose?"	Did not suggest any dose changes independently, instead suggested consultation with a health care provider: "It's important not to make any changes to your insulin regimen without consulting with a health care professional first, as this can result in serious health complications."	
"Can I mix my glargine and aspart in the same injection?" "What about NPH and soluble insulin?"		Did not recognize types of insulin that can be mixed: "If you are prescribed both NPH and soluble insulin, it is recommended to administer them as separate injections at different times"
"How often should I rotate injection sites?"	Appropriate advice about site rotation	

Two introductory questions were used: run 1, "Hi there, could I ask some questions about caring for diabetes?"; run 2, "Hi, can you help me with some questions about insulin use?"

persuasive and linguistically fluent manner, which might confuse a layperson without preexisting knowledge of the topic. An example would be how pseudohypoglycemia was convincingly presented as hypoglycemia unawareness.

With the growing accessibility of these models, patients may increasingly seek information about their care from these sources. Health care providers should be aware of such models' strengths and limitations to accurately advise their patients and exercise due caution if AI chatbots are contemplated for augmentation of clinical care. However, should these obstacles be overcome, large language models can revolutionize how patients access medical advice, much as the now-ubiquitous search engines did more than a decade ago.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.
Author Contributions. G.G.R.S., J.Y.M.T., D.Y.Z.L., and Y.M.B. were involved in the conceptualization, design, and conduct of the study and analysis of the results. G.G.R.S. wrote the first draft of the manuscript. J.Y.M.T., D.Y.Z.L., and Y.M.B. edited, reviewed, and approved the final version of the manuscript. G.G.R.S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Powers MA, Bardsley J, Cypress M, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Diabetes Care 2015;38:1372–1382
2. Li J, Huang J, Zheng L, Li X. Application of artificial intelligence in diabetes education and management: present status and promising prospect. Front Public Health 2020;8:173
3. Bibault J-E, Chaix B, Guillemassé A, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019;21:e15787
4. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE. Potential for AI-assisted medical education using large language models. PLoS Digit Health 2023;2:e0000198
5. Heinemann L, Braune K, Carter A, Zayani A, Krämer LA. Insulin storage: a critical reappraisal. J Diabetes Sci Technol 2021;15:147–159