Harnessing ChatGPT-o1 for Thematic Analysis: A Rapid Leap Forward

Stephanie C C van der Lubbe, V Vien Lee, Lay Hoon Goh, Jose M Valderas

Table of Contents

Harnessing ChatGPT-o1 for Thematic Analysis: A Rapid Leap Forward

Stephanie C C van der Lubbe[1*] PhD; V Vien Lee[1*] PhD; Lay Hoon Goh[1,2] MBBS, MMed, PhD; Jose M Valderas[1,2,3] MD, PhD

[1]Division of Family Medicine Yong Loo Lin School of Medicine National University of Singapore Singapore SG
[2]Department of Family Medicine National University Health System Singapore SG
[3]Centre for Research in Health Systems Performance (CRiHSP) National University of Singapore Singapore SG
[*]these authors contributed equally

Corresponding Author:
Stephanie C C van der Lubbe PhD
Division of Family Medicine
Yong Loo Lin School of Medicine
National University of Singapore
NUHS Tower Block, Level 9, 1E Kent Ridge Road
Singapore
SG

Abstract

No abstract required (Letter to the Editor)

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Letter to the Editor

Harnessing ChatGPT-o1 for Thematic Analysis: A Rapid Leap Forward

Stephanie C. C. van der Lubbe[1,#], PhD; V Vien Lee[1,#], PhD; Lay Hoon Goh[1, 2], MD, PhD; Jose M. Valderas[1,2,3], Prof, MD, PhD

[1]   Division of Family Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[2]   Department of Family Medicine, National University Health System, Singapore, Singapore

[3]   Centre for Research in Health Systems Performance (CRiHSP), National University of Singapore, Singapore

[#]   Authors contributed equally to this work.

Corresponding author:

Stephanie C. C. van der Lubbe, PhD

Email: svdlubbe@nus.edu.sg

We recently published our viewpoint titled "Harnessing ChatGPT for Thematic Analysis: Are We Ready?" in which we explored the potential of using ChatGPT-3.5 and ChatGPT-4 to perform thematic analysis.[1] Our viewpoint highlighted the strengths and limitations of this approach, considering factors like confidentiality and interpretive depth. Since our publication, OpenAI released ChatGPT-o1 in September 2024, a model with enhanced reasoning capabilities suitable for complex tasks.[2] This release presents new opportunities for leveraging ChatGPT in thematic analysis. We reanalyzed the full *Diabetes Discussion – A Diabetes UK Podcast* transcript[3] (used in our original paper) using ChatGPT-o1-preview, yielding new insights that form the basis of this letter.

With the prompt word limit expanded 16 times and the context window over 31 times compared to ChatGPT-3.5,[4] ChatGPT-o1 marks a substantial leap forward in thematic analysis capabilities. Previously, we were limited to uploading transcript segments, and context was partially lost between analysis steps.[1] Now, entire transcripts can be uploaded, and multiple thematic analysis steps can be conducted within a single, continuous conversation. These improvements enable a more cohesive, in-depth analysis, with contextual understanding maintained throughout the process. Specifically, this allows for more precise code and theme assignments by maintaining a clear link between the codes and the original transcript.

A potential workflow leveraging these upgrades might be as follows: First, initial codes are generated from the transcript attached to the prompt. Next, closely related codes are combined to form a structured set of overarching codes with corresponding subcodes by prompting ChatGPT with tailored instructions. Subsequently, the same prompt as in our original paper can generate themes and subthemes based on these codes. Since the entire process occurs within a single conversation, context is continuously retained, provided the length does not exceed 128,000 tokens. Additionally, while previous versions "hallucinated" codes and subcodes, ChatGPT-o1 demonstrated enhanced accuracy, generating themes and codes without any hallucination during our test runs.

Some words of caution are warranted. Although ChatGPT-o1 has significantly advanced beyond versions 3.5 and 4, it should still be seen as an additional member of the analysis team rather than a replacement, with outputs needing to be verified and complemented by human researchers. Furthermore, prompt engineering is essential. For example, the resulting codes lacked clear organizational structure, with codes appearing out of sequence relative to the original transcript. Consequently, detailed instructions on coding and output presentation are required. Moreover, ChatGPT-o1's ability to analyze larger volumes of text inherently increases the number of generated codes and quotes. Researchers should, therefore, establish workflows to ensure clear and efficient management of outputs. We recommend using the GPT API, which supports structured data input and enables customized outputs in formats (e.g., CSV) optimized for data management.

In summary, the new o1 series offers a powerful tool with enhanced contextual understanding for more efficient and in-depth thematic analysis. With these rapid developments in just half a year, it is exciting to consider how ChatGPT's role in qualitative research will continue to evolve, with improvements in context, accuracy, and analytical capabilities on the horizon.

Conflicts of Interest
None declared.

References
[1] VV Lee, SCC van der Lubbe, LH Goh, JM Valderas, Harnessing ChatGPT for Thematic Analysis: Are We Ready? J Med. Internet Res. 2024;26:e54974.
[2] https://openai.com/o1/ (accessed on 12 November 2024)
[3] Diabetes UK. Diabetes discussions: Living with diabetes. 2023; Available from: https://www.diabetes.org.uk/guide-to-diabetes/podcasts/living-with-diabetes.
[4] https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo (accessed on 12 November 2024)

Abbreviations
API: Application Programming Interface
CSV: Comma-separated values
GPT: Generative Pre-trained Transformers