

TRANSPORT FINDINGS

A Comparison of the Results from Artificial Intelligence-based and Human-based Transport-related Thematic Analysis

Thiago Carvalho¹, Hisham Negm¹, Ahmed El-Geneidy¹¹ Urban Planning, McGill University

Keywords: ChatGPT, Thematic analysis, transport, BRT, LLM, Gemini, Large Language Model, OpenAI

<https://doi.org/10.32866/001c.94401>

Findings

Artificial intelligence (AI) tools (in particular Large Language Models) have the potential to reduce the time needed to perform thematic analysis. To better understand their potential in the transportation field, we compare human-based to AI-based outcomes. Our findings indicate that AI tools, such as ChatGPT, could synthesize and summarize major topics present in our dataset regardless of previous user exposure to the subject or not. Nonetheless, caution is required as results might miss the nuance of less frequent themes. These tools could be used to accelerate the process under the supervision of researchers and practitioners given responder consent and the following of ethical practices.

Questions

The application of artificial intelligence (AI) in academic research is currently a contentious topic with many ethical and practical concerns (van Dis et al. 2023), such as biased and inaccurate outputs and the lack of transparency in their use. In this context, researchers have started investigating the potential applications of this technology to a diverse range of tasks (Chang et al. 2023; Ray 2023). If proven reliable, AI applications have the potential to reduce the workload of time-consuming tasks such as thematic analysis. Thematic analysis is a popular technique in social sciences which aims to identify and interpret patterns in qualitative datasets (Braun and Clarke 2022). To further understand the potential of applying AI applications to perform transport-related thematic analysis, we pose the following question: "How do the outputs from artificial intelligence compare to human-generated thematic analysis on a transport-related topic?"

Methods

This study uses data from the 2022 wave of the Montréal Mobility Survey (MMS) conducted in the Montréal metropolitan region (Negm et al. 2023). MMS is a bilingual survey that collects data from over 4000 participants about their travel behavior data and opinions on major transport projects in the region. These projects include the Pie-IX Bus Rapid Transit (BRT), a \$426M Cad system with 20 stops along a 13 km stretch that opened in November 2022 in the East side of the Montréal Island. We specifically use a dataset of 50 open-ended responses to the question *"Is there anything else you would like to share about the anticipated impacts of the Pie-IX BRT? If you do not have any suggestions, you do not need to respond to this question."* Responses are on average 31 (s.d. 29) words long and range from 4 to 150 words. A human-based thematic analysis was conducted based on the 6-step approach suggested by

Nowell et al. (2017), aiming to derive repeated patterns of meaning across the dataset. To better understand the incidence and relevance of the themes derived in the analysis, we also quantify them based on their frequency as proposed in the Applied Thematic Analysis approach (Guest, MacQueen, and Namey 2012). A codebook and peer-debriefing were kept, ensuring the reliability and soundness of the results.

To assess the quality and veracity of the outputs provided by two large language models (LLMs), ChatGPT and Google Gemini, we compare the results found in the first step to the outcomes from the LLM. Both ChatGPT and Google Gemini are generative artificial intelligence chatbots based on pre-trained transformer models. In other words, they generate the most fitting sequence of words as a response to a given prompt based on pre-exposure to large amounts of information and reinforced learning algorithms, which rewards a model based on the output. We conduct three analyses in ChatGPT and one in Google Gemini. In Analysis 1, we used a brand-new account in the free version of ChatGPT, version 3.5. We inserted our full dataset (50 responses) only once and we used the following prompt: “Extract the five main themes from the following survey responses:”. In Analysis 2, also using ChatGPT-3.5, we used an account that has been repeatedly exposed to transport-related queries over six months. We first inserted only a subset of the 50 responses and we asked ChatGPT to find the main themes in these responses. We then employed a series of chain-of-thoughts prompts to dig deeper into the provided response. For example, when ChatGPT returned only negative views, we asked if there were any positive points. Then we asked if there were any less frequent views that are not mentioned by the majority of respondents. Afterwards, we inserted the whole dataset, and we used the same main prompt from the first analysis. We repeated this same step with no changes three times. Each time, ChatGPT provided us with nonidentical but similar five themes. In the last step, we inserted the fifteen themes extracted from the three repetitions and asked ChatGPT to give us a conclusive result with only five themes that cover the topic adequately. This analysis was performed from the same account in July 2023 and repeated in February 2024. The results from both analyses were very similar. In Analysis 3, we used the same account and steps as Analysis 2 but in the paid version of ChatGPT, ChatGPT-4, in February 2024. In Analysis 4, we used a new account in the freely accessible Google Gemini and repeated the same detailed steps as Analysis 2. In comparing human-based and AI-based outputs, we search for the presence of similar topics, themes, subthemes, or wording across the analysis.

Findings

[Table 1](#) reports on human- and artificial intelligence-based thematic analyses. For the human-based version, we display both topic and theme frequency while for the AI iterations, we display their presence among the AI outputs. The theme descriptions provided by ChatGPT and Google Gemini are available in the Supplemental Information file. The human-based thematic

Table 1. Results of human-based and artificial intelligence-based thematic analysis

Theme / Sub-Theme	Human analysis Frequency	Analysis 1 ChatGPT-3.5 Presence	Analysis 2 ChatGPT-3.5 Presence	Analysis 3 ChatGPT-4 Presence	Analysis 4 Google-Gemini Presence
Perception of the project					
Project is outdated/inefficient/inadequate (-)	64.0%	✓	✓	✓	✓
Project is good/efficient (+)	4.0%		✓	✓	✓
Similar project existed before and failed (-)	4.0%				
Regional impacts					
Inadequate choice of technology/mode (-)	36.0%	✓	✓	✓	✓
Construction/planning timeline (-)	20.0%	✓	✓	✓	✓
Ineffective allocation of funds (-)	20.0%	✓	✓	✓	✓
Construct other BRT lines across the city (+)	6.0%		✓	✓	✓
Local impacts					
Nuisance of construction (-)	6.0%	✓	✓	✓	✓
BRT could be expanded to other neighborhoods in the region (+)	6.0%		✓	✓	✓
Road safety concerns (-)	4.0%		✓	✓	✓
Satisfaction with operations					
Satisfaction with service attributes/operations (-)	8.0%			✓	
Operation is limited (i.e., frequency, hours, stops) (-)	6.0%			✓	
Satisfaction with service attributes/operations (+)	4.0%				
Impacts to the environment					
Not favourable to the environment (-)	8.0%	✓	✓	✓	✓

analysis reflects five major themes with their pertaining sub-themes. The first theme relates to positive or negative perceptions of the need for and the benefits from the project. The second focuses on regional (city-wide) impacts of the project while the third considers local nuisances, such as construction and road safety. The fourth theme displays (dis)satisfaction with the BRT operations and the last one focuses on environmental impacts.

Comparing the human-based themes with the ChatGPT results, we find that ChatGPT and Google Gemini are efficiently able to extract almost all the themes and sub-themes. This is specifically true for the trained and multiply prompted ChatGPT-4. The trained ChatGPT-3.5 and Google Gemini were also effective. The only mentionable difference is how LLM classified and named the themes and sub-themes, which is a decision that differ from one researcher to another in human-based analysis. For instance, in Analysis 3, ChatGPT-4 denotes “Project Execution and Community Impact” as a theme with several construction and management aspects to it as sub-themes, while human researchers classified them differently based on regional and local impacts. From the fourteen sub-themes extracted by the researchers, only two were not identified by ChatGPT-4. Both were mentioned only by 4% of the

participants. In Analysis 1, the AI, which was only prompted once without any training, found the main themes that were mentioned by at least 20% of the responses. This is remarkably efficient when we acknowledge that this is the result from the minimum effort from the human side where we just inserted the dataset once and gave one prompt. Altogether, this proves how rigorous LLM can be in thematic analysis when it is well-prompted and multiple repetitions are performed for the same analysis.

In conclusion, ChatGPT-4 was able to identify all frequently mentioned themes and most of the less frequent ones identified by the human-based solution. The less prompted ChatGPT-3.5 performed best only with highly frequent themes (20% present in responses and above). Therefore, it implies that AI tools, such as ChatGPT and Google Gemini, can synthesize and summarize the major topics present in open-ended responses regardless of being previously exposed to the subject by the user. Nonetheless, caution is required as outputs might miss the nuance provided by the less cited themes. Moreover, future updates to the platform might impact the consistency and reliability of the outputs, which makes this an ongoing research topic. Similarly, as LLM usually follow a proprietary source code, it is difficult to know exactly how the tool is understanding prompts and deriving outputs. LLM may lack the necessary rigour and subjectivity to perform more complex analysis involving qualitative data. However, as it stands, it could be used as a preliminary tool to speed up the execution of thematic analyses under the supervision of transport and urban planning researchers and practitioners given the consent from respondents and the following of ethical practices.

Acknowledgments

This research was funded by the Natural Sciences and Engineering Research Council of Canada grant Towards a better understanding of the determinants and satisfaction of travel among different groups in major Canadian Cities (NSERC RGPIN-2023-03852) and the Social Sciences and Humanities Research Council's partnership grant Mobilizing justice: Towards evidence-based transportation equity policy (SSHRC 895-2021-1009).

Submitted: July 26, 2023 AEDT, Accepted: February 27, 2024 AEDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Braun, Virginia, and Victoria Clarke. 2022. "Thematic Analysis: A Practical Guide." *QMiP Bulletin* 1 (33): 46–50. <https://doi.org/10.53841/bpsqmip.2022.1.33.46>.
- Chang, Yupeng, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, and Yidong Wang. 2023. "A Survey on Evaluation of Large Language Models." *ACM Transactions on Intelligent Systems and Technology*.
- Dis, Eva A. M. van, Johan Bollen, Willem Zuidema, Robert van Rooij, and Claudi L. Bockting. 2023. "ChatGPT: Five Priorities for Research." *Nature* 614 (7947): 224–26. <https://doi.org/10.1038/d41586-023-00288-7>.
- Guest, Greg, Kathleen MacQueen, and Emily Namey. 2012. *Applied Thematic Analysis*. California: SAGE Publications, Inc. <https://doi.org/10.4135/9781483384436>.
- Negm, H., P. Redelmeier, A. Soliz, R. Victoriano-Habit, L. Rodrigue, N. Soto, M. James, and A. El-Geneidy. 2023. "Réseau Express Métropolitain (REM) Survey Report: 2019-2022." Montréal, Québec, Canada: Transportation Research at McGill (TRAM), McGill University.
- Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. "Thematic Analysis." *International Journal of Qualitative Methods* 16 (1): 160940691773384. <https://doi.org/10.1177/1609406917733847>.
- Ray, Partha Pratim. 2023. "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope." *Internet of Things and Cyber-Physical Systems* 3: 121–54. <https://doi.org/10.1016/j.iotcps.2023.04.003>.

SUPPLEMENTARY MATERIALS

Analysis results

Download: <https://findingspress.org/article/94401-a-comparison-of-the-results-from-artificial-intelligence-based-and-human-based-transport-related-thematic-analysis/attachment/197811.pdf>
