Applying Large Language Models to Interpret Qualitative Interviews in Healthcare

Article · August 2024

DOI: 10.3233/SHTI240530

Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240530

Applying Large Language Models to Interpret Qualitative Interviews in Healthcare

Marie WOSNY[a,b] and Janna HASTINGS[a,b,c,1]
[a] *School of Medicine, University of St. Gallen (HSG), Switzerland*
[b] *Institute for Implementation Science in Health Care, University of Zurich*
[c] *Swiss Institute of Bioinformatics, Switzerland*
ORCiD ID: Marie Wosny https://orcid.org/0009-0000-2791-0774
Janna Hastings https://orcid.org/0000-0002-3469-4923

Abstract. To address the persistent challenges in healthcare, it is crucial to incorporate firsthand experiences and perspectives from stakeholders such as patients and healthcare professionals. However, the current process of collecting, analyzing and interpreting qualitative data, such as interviews, is slow and labor-intensive. To expedite this process and enhance efficiency, automated approaches aim to extract meaningful themes and accelerate interpretation, but current approaches such as topic modeling reduce the richness of the raw data. Here, we evaluate whether Large Language Models can be used to support the semi-automated interpretation of qualitative interview data. We compare a novel approach based on LLMs to topic modeling approaches and to manually identified themes across two different qualitative interview datasets. This exploratory study finds that LLMs have the potential to support incorporating human perspectives more widely in the advancement of sustainable healthcare systems.

Keywords. Healthcare, information technology, clinical information systems, healthcare professionals, qualitative research, thematic analysis, topic modeling

1. Introduction

Qualitative healthcare research presents a unique opportunity to harness the firsthand experiences of individuals from their diverse perspectives to inform innovation and transformation in healthcare [1]. Qualitative interviews, for example, yield conceptual and theoretical insights into people's individual views, opinions, emotions, and thoughts, often generating new insights or innovation directions [2]. However, qualitative research is still infrequently used, partly due to the time-constraints and cost associated with annotating qualitative data through coding [3]. As a result, automated approaches have been suggested as a solution to support qualitative analysis [4]. One such approach is topic modeling, a family of algorithms that can be used to automatically identify and categorize similar themes or 'topics' within a dataset of unstructured text, such as interview transcripts [5].

[1] Corresponding Author: Janna Hastings, University of Zurich; E-mail: janna.hastings@uzh.ch.

Recently, large language models (LLMs) have emerged as a breakthrough technology for natural language processing across a range of tasks in health research [6]. They have also been suggested as a potential supportive technology for qualitative research, by facilitating deductive coding of transcripts, providing a systematic and reliable platform for code identification, and helping to avoid analysis misalignment [8]. However, previous studies have focused solely on deductive analysis or were not related to the health sector [7]. In addition, retrospective studies that re-analyze published data have the potential risk that the model has 'seen' the study previously, complicating the evaluation. In the current study, we aim to investigate the potential of LLMs for automating inductive and deductive thematic analysis in qualitative interview studies for the healthcare sector. We compare LLMs to the capabilities of traditional topic modeling and to human annotations, across two different interview datasets: a published dataset of interviews about childhood vaccination [9], and extracts from an interview study we conducted into clinician perspectives on digitalization that explored the lived experiences of healthcare professionals navigating digital tools throughout their daily routines [10], which interviews have not been published and therefore have not been 'seen' by any model previously.

2. Methods

Figure 1 illustrates our overall study design. All analyses were conducted in Python using Jupyter notebooks. Source code is available from [12].

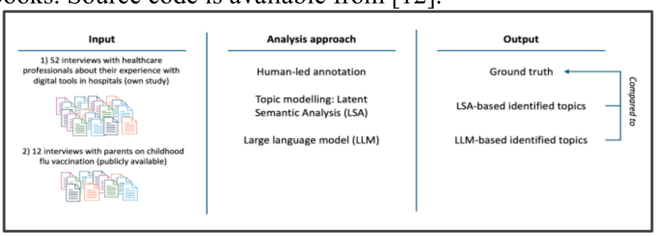

Figure 1. Overview of study design

2.1. Datasets

We used extracts from our own study into digitalization in the clinic, which consisted of 52 interviews conducted with healthcare professionals across Switzerland. Additionally, for a broader analysis and methodological validation, we incorporated a representative interview dataset from a previous study comprising 12 interviews. This dataset, curated by Price et al [9], focused on the barriers and facilitators of childhood flu vaccination, analyzed according to the COM-B model (capability, opportunity, and motivation) [11].

2.2. Topic Modeling - Latent Semantic Analysis (LSA)

Interview transcripts of each dataset were loaded and preprocessed using the spaCy Python library to remove irrelevant content and standardize the text format, and remove stopwords. LSA was performed with the TruncatedSVD implementation from the sklearn.decomposition library. We set the number of topics to 10. Human manual interpretation was necessary to evaluate the final topic categories.

2.3. LLM-based Thematic Analysis

For automating the thematic analysis using LLMs, we used a two-step prompting strategy. We first prompted the model to identify the themes in individual transcripts, and subsequently, we provided all the identified themes back to the model and asked the model to summarize and integrate, which parallels the fashion in which human inductive coders would review and integrate themes across a dataset. For the flu dataset, we used the gpt-3.5-turbo via the API provided by OpenAI, and for our unpublished dataset we used the Mixtral 7x8b model locally to enable us to keep the dataset private.

3. Results

3.1. Topic modeling

The results of LSA applied to each of the datasets returned topics with characteristic words for each topic. Two exemplary topics for each of the datasets are illustrated below (Table 1); the full results are available in online supplementary material at [12].

Table 1. Selected results from Latent Semantic Analysis topic modeling approach of two interview datasets

Dataset	Identified topics
Child-hood flu vaccination (Price et al.)	feel, come, obviously, younger, school, guess, nhs, health, kid, interesting, young, speak, happen, ill, effect, letter, appointment, remember, cold, definitely
	wife, straightforward, pretty, faith, obviously, seek, letter, beneficial, medical, uk, website, alright, important, garden, world, scientist, certainly, science, personally, text
Clinician experience with digital tools (Wosny et al.)	work, patient, think, positive, tool, look, time, course, know, thing, example, good, come, use, like, need, information, lot, quickly, cool
	work, tool, relatively, annoy, difficult, open, expect, trigger, electronically, care, transfer, actually, realize, little, medical, update, technology, grateful, long, anymore

3.2. LLMs

Extracts from the results of applying the LLMs to the two datasets are listed below. The full result transcripts are available in a supplementary file at [12].

3.2.1 Childhood flu vaccination interview dataset

```
1. **Capability:**
  - Understanding and knowledge about the flu, its symptoms, and the flu vaccine.
  - Access to information sources like healthcare professionals, the internet, and official health resources.
  - Informed decision-making process regarding vaccination for children.
2. **Opportunity:**
  - Access to healthcare professionals for vaccination advice and appointments.
 - Challenges in accessing healthcare services and vaccine supply due to personal circumstances and logistical issues.
```

3.2.2. Clinician experience interview dataset

```
Positive experiences with digitalization:
```

* Efficiency: Many participants reported that digital tools helped them save time and increase efficiency in their work. This was achieved through various means such as quick access to patient information, efficient search for ECGs, automatic data transfer from monitors, and time-efficient communication.
* Quality improvement: Digital tools also helped improve the quality of work. For instance, dictation software helped improve letter writing, while digital anesthesia systems helped provide better patient care in critical situations. Digital tools also helped in early recognition of potential issues and avoiding repetition of unsuccessful treatments.
* Convenience: Participants appreciated the convenience offered by digital tools. For example, mobile trolleys with medication and laboratory open, and quick access to personal settings were highly appreciated. Digital tools also allowed for remote access to clinics and work from home, which increased flexibility.
* Modernization: Digital tools helped modernize various aspects of work such as efficient instrument tracking, in-house development of digital tools, and integration of tools into workflow.

4. Discussion

The thematic analysis we conducted on our own clinician interview dataset revealed that the integration of digital tools within the hospital ecosystem and HCPs' resulting experiences were multifaceted but primarily negative, characterized by frustration, annoyance, anger, dissatisfaction and stress. In contrast, themes of positive experiences associated with digital tools include satisfaction, enthusiasm, gratitude, and relief, mainly arising from digital tools' assistance with daily tasks. Although the LSA topic modeling approach successfully identified patterns that could, through manual interpretation of the returned word sets, be determined to be consistent with some of those identified through human coding, the manual interpretation of the topics requires familiarity with the original data and theory. Similarly, for the second dataset on childhood flu vaccination, while the original paper identified a range of barriers and facilitators beneath the three key themes of capability, opportunity and motivation, the LSA approach needed extensive manual interpretation to interpret the returned themes as revealing some of the underlying factors influencing parental perceptions regarding childhood flu vaccination. Notably, the use of the COM-B framework as a structured approach is not accessible within the context of topic modeling. Moreover, the quality of the returned topics is lower the fewer documents there are for analysis, and topic modeling would not typically be used to analyze a single interview. Thus, there is clearly a need for additional approaches that are more suited to the unstructured and personalized nature of interview transcripts, providing motivation for applying LLMs for this purpose as we aimed to do with this study. In contrast to the traditional topic modeling approach, the LLM-based approach was able to deduce a variety of themes directly from provided interview transcripts and group those beneath the requested theoretical frameworks and guiding research questions. For each individual transcript a set of themes corresponding to the prompted interview question was identified, for example, positive experiences with digital tools caused by efficiency and quality improvement in our interview study with clinicians, and capability as well as opportunities that serve as facilitators in the childhood flu vaccination study interview dataset. Following the two-step prompting strategy, the LLM was able to automatically identify themes and categorize the automatically extracted themes beneath structured headers corresponding to the respective theoretical frameworks. It is noteworthy that the LLM approach allows identifying themes even within a single interview or part of an interview, as we tested using only the extracted transcripts from one question in our clinician interview study. However, thematic identification in single interviews was noisy, as different but similar sounding themes were identified in different interviews. The second step of summarizing and integrating provided a single coherent and integrated result in each example we tested, although for the flu dataset, the summary was perhaps a bit too high-level, while the results from the study on clinician experiences were in some cases interpreted or weighted differently as compared to our manually annotated ground truth.

5. Conclusions

LLMs stand poised to transform multiple aspects of healthcare due to their flexible capabilities to support automation of a range of tasks. However, their application should be carefully evaluated for each intended use case. With this brief exploratory study, we found that LLMs are able to offer a user-friendly and promising form of automation for assisting the analysis of qualitative datasets, complementary to the capabilities of existing automation strategies such as topic modeling and more suited to inductive and deductive thematic analysis. However, further research is needed to optimize the method for integration of themes across different interviews and to ensure that the LLM does not alter the interpretation of themes or add any themes that were not present in the original interviews, particularly in the summarization step. In addition, the differential performance across different topics and interview styles should be explored to ensure that the models do not propagate biases or harmful stereotypes, as this has been noted elsewhere as a potential challenge with the technology [13].

In conclusion, the study provides evidence that careful application of this novel technology could help to accelerate the adoption of a wider range of perspectives in support of a more sustainable healthcare transformation.

References

[1] Al-Busaidi ZQ. Qualitative Research and its Uses in Health Care. Sultan Qaboos Univ Med J. 2008 Mar;8(1):11–9.
[2] Tanwir F, Moideen S, Habib R. Interviews in Healthcare: A Phenomenological Approach A Qualitative Research Methodology. JPHI. 2021 Jul 8;4(2):10–5.
[3] Bryman A. Barriers to Integrating Quantitative and Qualitative Research. Journal of Mixed Methods Research. 2007 Jan 1;1(1):8–22.
[4] Lennon R, et al. Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. Family Medicine and Community Health. 2021 Nov 1;9:e001287.
[5] Kherwa P, Bansal P. Topic Modeling: A Comprehensive Review. ICST Transactions on Scalable Information Systems. 2018 Jul 13;0(0):159623.
[6] Clusmann J, et al. The future landscape of large language models in medicine. Communications Medicine. 2023;3(1): 1–8. https://doi.org/10.1038/s43856-023-00370-1
[7] De Paoli, S. (2023). Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. Social Science Computer Review, 08944393231220483. https://doi.org/10.1177/08944393231220483
[8] Tai RH, Bentley LR, Xia X, Sitt JM, Fankhauser SC, Chicas-Mosier AM, et al. Use of Large Language Models to Aid Analysis of Textual Data. bioRxiv 2023.07.17.549361.
[9] Price T, McColl E, & Visram S. Barriers and facilitators of childhood flu vaccination: The views of parents in North East England. Zeitschrift Fur Gesundheitswissenschaften = Journal of Public Health. 2022;30(11): 2619–2626. https://doi.org/10.1007/s10389-022-01695-2
[10] Wosny M, Strasser LM, & Hastings J. The Paradoxes of Digital Tools in Hospitals: A Qualitative Interview Study. JMIR Preprints. 2024 https://doi.org/10.2196/preprints.56095
[11] Michie S, van Stralen MM, & West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Science 2011: IS, 6, 42.
[12] Hastings J, and Wosny M (2024). GitHub. https://github.com/hastingslab-org/topic-modelling
[13] Hastings J Preventing harm from non-conscious bias in medical generative AI. The Lancet Digital Health. 2024;6(1); e2–e3. https://doi.org/10.1016/S2589-7500(23)00246-7