

How do smokers perceive preparatory activities proposed by a chatbot to help them quit smoking?

Keshav Nair

K.Nair@student.tudelft.nl

Supervisor: Willem-Paul Brinkman

Introduction

Given that smoking remains a leading cause of preventable death, causing over 8 million deaths worldwide per year, supporting cessation of smoking is crucial. This is not just important for smokers, as this issue affects non-smokers too, with about 1.2 million deaths being caused by second-hand smoking.

Although there is plenty of research on the topic of using chatbots to support the cessation of smoking, this specific topic is under-researched. To tackle this, we employ a thematic analysis of an existing dataset of smokers' written responses to a survey regarding the preparatory activities provided by the chatbot.

Research Questions

Main research question:
- How do smokers experience preparatory activities suggested by a chatbot?

Sub research questions:
- What key themes emerge from smokers' descriptions of their experiences with preparatory activities in the intervention?
- How do the themes generated by LLMs in thematic analysis compare to those identified through manual thematic analysis?
- How do LLMs perform in applying a predefined coding scheme?

Methodology

Thematic Analysis is a structured method to analyse qualitative data, such as text responses, to identify underlying themes in these responses. Thematic analysis will be used on text responses by smokers to answer the RQ.

- Familiarizing with Data:** Read and re-read your data (e.g., transcripts) to become immersed and note initial ideas.
- Generating Initial Codes:** Create codes to label interesting features of the data (words, phrases, sentences).
- Searching for Themes:** Group codes into potential themes, looking for patterns of meaning.
- Reviewing Themes:** Refine themes; ensure they are coherent and distinct, and review them against the data.
- Defining and Naming Themes:** Clearly define each theme and give it a concise, informative name.
- Producing the Report:** Write up the analysis, using examples to illustrate the themes.

After the manual thematic analysis, it will be used as a baseline to compare against an LLM based thematic analysis. This will be conducted using the Top-Down Structured Prompting (TDSP) technique. This consists of the following steps:

- Generation of Overarching Themes and Sub-themes:** The LLM is provided with the research question and the complete sanitized dataset to generate initial high-level themes and their corresponding sub-themes.
- Categorization of data by sub-theme**
- Extraction of key insights within each sub-theme**
- Synthesis of evaluation**

To compare the performance of the various thematic analyses, Cohen's kappa will be used to calculate agreeability. For manual thematic analysis, it will be calculated on a peer coding, so another researcher will perform the same manual thematic analysis and the agreement will be calculated, theme generation is also done using peer coding. This will be done with the LLMs instead of the other researcher, when evaluating the coding scheme application of LLMs.

Results

Theme Name	Cohen's Kappa (κ)
Building Self-Awareness	0.446
Coping Strategies	0.896
Motivation and Goal Reinforcement	0.675
Perceived Usefulness	0.819
Personal Obstacles	0.661
Skepticism	0.736
Stress Management	0.852
Willingness to Change	0.755
Average Kappa Score	0.730

Table 1: High Inter-Rater Reliability in Manual Analysis

Manual Analysis	Qwen3	Mistral Nemo
Motivation and Goal Reinforcement	Motivational Techniques and Future Identity	Motivation and Commitment
Coping Strategies	Distraction Through Activities	—
Willingness to Change	—	—
Skepticism	Skepticism and Individual Differences	Concerns and Skepticism
Building Self-Awareness	Awareness and Monitoring	Monitoring and Awareness
Stress Management	Behavioral Strategies and Self-Discipline	—
Personal Obstacles	Skepticism and Individual Differences	Concerns and Skepticism
Perceived Usefulness	—	Activity Engagement

Table 2: LLMs Identify Broad Themes but Miss Human Nuance

Model Name	Average Cohen's κ
Qwen3 30B A3B	0.135
DeepSeek R1 distill Qwen 7B	0.054
Mistral Nemo Instruct 2407	0.326
Manual Inter-Rater (Human Baseline from Table 1)	0.730

Table 3: LLM Application Accuracy is Substantially Lower than Human Baseline

Table 1 takeaway: A strong Kappa score of 0.730 demonstrates the human-developed coding scheme is robust and reliable

Table 2 takeaway: LLMs failed to generate the 'Willingness to Change' theme and conflated 'Personal Obstacles' with 'Skepticism'

Table 3 takeaway: The best LLM only achieved 'fair' agreement ($\kappa=0.326$), far below the 'substantial' human agreement ($\kappa=0.730$)

Key Findings

The manual analysis uncovered eight key themes capturing the smokers' journey.

Users were motivated when activities aligned with their personal health goals.

Users frequently expressed doubts about whether the proposed activities would be effective against a physical addiction.

Willingness to change was a powerful underlying commitment to quit, overriding specific doubts about an activity's effectiveness.

Now for the LLM based analysis. LLMs successfully identified broad topics like 'Motivation' and 'Skepticism'.

However they failed to identify 'Willingness to change', and found it challenging to separate personal obstacles from skepticism.

The LLMs struggled to apply the predefined coding scheme consistently.

The best performing LLM achieved an average Kappa score of 0.326. This is significantly lower than the human-to-human reliability score of 0.730

Conclusion

To be effective, cessation aids must be designed to personalize activities to avoid personal obstacles and user skepticism. Additionally inherent motivation plays a key role in cessation and how users perceive the preparatory activities.

LLMs show potential as an exploratory aid for generating high-level themes. They are not yet a viable tool for applying a predefined, nuanced coding scheme where accuracy is critical. Human analytical oversight remains essential for ensuring the depth and validity of qualitative research