

# Understanding User Experiences of Preparatory Activities in a Digital Smoking Cessation Intervention

Author  
Antonio Florin Lupu  
A.F.Lupu@student.tudelft.nl

Supervisor and Responsible Professor  
Willem-Paul Brinkman

## 3. METHODOLOGY

### Dataset

- Contains around 2300 open-ended responses [1]
- Smokers and vapers answered the question: "How did you approach, do, or experience the preparatory activities suggested by the chatbot?"
- Due to the complexity of the questions (three subquestions combined into one), a lot of responses were irrelevant for our research. Some users only answered one of the three subquestions
- Because of this, we decided to filter the data points so that we continue working only with those that describe how users **experienced** the activity.
  - So any responses that only described **what** the users did (e.g. running, yoga, meditation) were filtered out.
- After filtering, the final dataset we worked with contained around 650 responses.

### Step 1: Manual Thematic Analysis

Apply **thematic analysis** to identify patterns in open-ended user responses

- Selected a random sample of 100 responses from the cleaned dataset
- Developed an initial coding scheme independently
  - A peer independently coded the same data, forming their own coding scheme
  - Compared and discussed both theme sets
  - Merged overlapping ideas to form one final coding scheme
- To ensure some reliability in the coders' application of the final coding scheme, we decided to adopt another peer review approach
  - Two coders independently applied the same final coding scheme to a new random sample of 100 user responses.
  - Inter-coder reliability was tested, with a resulting Cohen's Kappa of around 0.72, which signifies *moderate agreement*
- Following this reliability check, we applied the final coding scheme to the rest of the data points.

### Step 2: AI-Assisted Thematic Analysis

Test whether a local LLM (LLaMA 3.3, 8B) can assist with:

- Reproducing the same themes we got through manual analysis, by analyzing the same dataset.
- Reliably applying a predefined coding scheme to new data.

The general focus was put on prompt engineering rather than performance-testing multiple LLM models, since newer and better models will always be developed. However, a good prompt can always be reused on a different model. To improve reliability, we included role-setting, specific instructions, and temperature control in the prompt.



## 4. RESULTS

### Manual Thematic Analysis Results

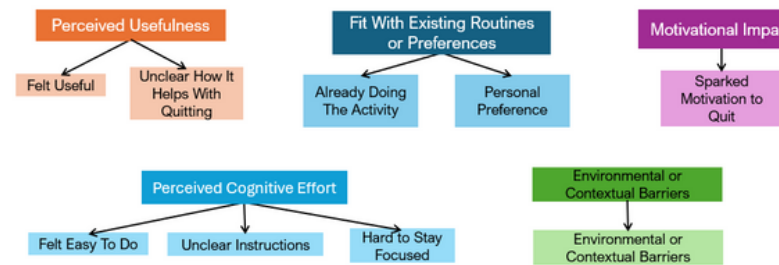


Figure 1. Themes and Subthemes derived from Manual Analysis

### Theme Generation Results

| Human theme                               | AI theme                      |
|---|-------------------------------|
| Perceived Usefulness                      | Activity Helpfulness          |
| Perceived Cognitive Effort                | Task Complexity               |
| Fit with Existing Routines or Preferences | Redundancy in Recommendations |
| Motivational Impact                       | Motivational Impact           |
| Environmental or Contextual Barriers      | External Challenges           |

### Theme Application Results

- When asked to apply the human-created coding scheme to new responses, the LLM performed very poorly
- Measured agreement with human coding was almost nonexistent
- Cohen's Kappa = 0.003, indicating the model struggled with the subjective, complex, or ambiguous responses that were part of the dataset
- As a takeaway, LLMs may assist with exploratory theme generation, but they currently lack the reliability needed for detailed, structured coding in complex qualitative research.

## RELATED LITERATURE

[1] N. Albers, F. Melo, M. Neerincx, O. Kudina, and W.-P. Brinkman, "The impact of human feedback in a chatbot-based smoking cessation intervention: An empirical study into psychological, economic, and ethical factors - Data and analysis code for the PhD thesis chapter," 4TU ResearchData, Dataset, Version 1, 2025. [Online]. Available: <https://doi.org/10.4121/1d9aa8eb-9e63-4bf5-98a3-f359dbc932a4.v1>

## 1. INTRODUCTION

- In a digital intervention designed to help people quit smoking or vaping, a chatbot suggested preparatory activities such as reflecting on motivation, increasing physical exercise, or journaling.
- These tasks aimed to mentally prepare users for quitting — but how did the users actually experience these activities?
- This study explores how users described these activities in their own words, as well as whether a large language model (LLM) can assist in analyzing such qualitative feedback.

## 2. RESEARCH QUESTION

**How do smokers and vapers experience preparatory activities suggested by a chatbot during a smoking cessation intervention?**

### Subquestions:

- What do smokers' and vapers' free-text responses reveal about their experience with chatbot-recommended preparatory activities?
- To what extent can LLMs support qualitative analysis?



## 5. CONCLUSION

**What do smokers' and vapers' free-text responses reveal about their experience with chatbot-recommended preparatory activities?**

- The study shows that users' experiences with chatbot-suggested preparatory activities were mixed and highly individual. Some participants felt the tasks were motivating, relevant, and easy to integrate into their lives. Others found them confusing, irrelevant, or difficult to complete due to personal preferences or life circumstances
- These findings underline that digital health tools like chatbots need to be flexible and personalized, accounting for users' diverse needs, routines, and potential barriers

**To what extent can LLMs support qualitative analysis?**

- In exploring the role of AI, results suggest that LLMs like LLaMA 3.3 (8B) can support the early, exploratory stages of qualitative research by finding broad themes in data. However, they currently lack the precision to reliably apply detailed coding schemes to subjective, nuanced responses
- Human interpretation remains essential, particularly when analyzing open-ended, emotional complex data in behavior change research.

## 6. FUTURE WORK

- Fine-tune LLMs for domain-specific qualitative research
- Use few-shot learning to guide theme application more reliably
- Test LLMs on simplified or better-structured questions

