Towards Effective Smoking Cessation: Understanding the Needs of Daily Smokers from eHealth Chatbot Interactions

Vlad-Gabriel Iftimescu

Supervisor(s): Nele Albers, Willem-Paul Brinkman

EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
February 2, 2024

Name of the student: Vlad-Gabriel Iftimescu
Final project course: CSE3000 Research Project
Thesis committee: Nele Albers, Willem-Paul Brinkman, Huijuan Wang

Abstract

Smoking has been one of the great threats to health in recent years, being strongly correlated with multiple negative health consequences, including lung cancer. Recent research suggests that artificial intelligence chatbots can be effective in persuading healthy behavior change. However, these chatbots usually rely on persuasive techniques to achieve their goal. Such techniques depend on identifying and meeting users' needs to be effective. To help improve understanding of the domain of healthy behavior change, we proposed a study which analysed the needs of daily smokers as they emerged from their interactions with a chatbot specifically designed to help them quit. The study performed a thematic analysis on users' free-text feedback, from which a set of 8 themes that directly correspond to 8 different needs were observed. The user needs were correlated with their genders, ages and highest completed education levels. While most of the results indicate that there is no significant correlation between the needs and user characteristics, which suggests that user needs are evenly distributed, certain correlations were highlighted for further analysis.

1 Introduction

In the light of recent research by He et al. (2022), society confronts the persistent burdens of tobacco-related diseases. Consequently, delving into the complex domain of smoking cessation now reveals a nuanced landscape, which requires a strategic and evidence-based approach for effective applications in real-life situations. A recent study by Dai et al. (2022) describes the correlation between exposure to smoking and 36 different health effects. The study concluded that smoking has a strong-to-very-strong correlation with 8 of those 36 effects, including lung cancer, and a weak-to-moderate correlation with 21 other effects. Given the harmful nature of smoking, it comes as no surprise that one of the most popular areas of research in the past decades has been that of healthy behavioral changes, especially in the case of smoking cessation.

Modifying unhealthy behaviors such as smoking can be one of the defining aspects for one's health (Mokdad, 2004). However, research such as that of Cooper et al. (2010) showed that smoking cessation is difficult without proper help. New theories and methods for giving up smoking have been proposed over the years, and the domain has evolved alongside technology. As such, one of the newer means of encouraging healthy behavioral changes is that of eHealth applications, which also have use cases in the context of smoking cessation (Fang et al., 2023; Meijer et al., 2021). These applications typically try to persuade people into quitting smoking by providing them with persuasive messages, knowledge on the harmful effects of smoking as well as general well-being strategies (Lentferink et al., 2017).

As per the study of Albers et al. (2022), eHealth applications are currently faced with the problem of low effectiveness on the employed persuasive attempts. One of the emerging technologies that can aid researchers in persuading healthy behavioral changes is that of artificial intelligence (AI) chatbots. Previous research suggests that such chatbots have the potential to solve the persuasion ineffectiveness problem by tailoring chatbot responses to user characteristics and interaction style (Albers et al., 2023). Chatbots can provide people with advice, personalized tasks to improve various aspects related to healthy lifestyle, as well as mental and emotional support (Aggarwal et al., 2023; Zhang et al., 2020). However, such AI programs need to be designed with real-life use cases in mind, which requires a large dataset of actual interactions with people trying to quit smoking, from which their needs can be deduced. Such a dataset has already been made available by the Perfect Fit study of Albers and Brinkman (2023). The study gathered data from 671 daily smokers' interactions with the chatbot Mel, which was designed to help them stop smoking. This dataset encompasses users' demographic information, effort values, as well as their feedback on their sessions with Mel.

The research of Albers et al. (2022) pointed that current smoking cessation applications lack a thorough understanding of users' needs. If a user's needs are not met by the smoking cessation strategy of their choice, they will lack a reason or incentive to continue with their self-improvement journey. As such, studies have been previously conducted to understand how to design eHealth applications that meet needs regarding smoking cessation or weight management (Albers et al., 2023; Asbjørnsen et al., 2020). A number of papers that were found for the purpose of the current research suggested that there is still much more to understand and improve when it comes to understanding user needs in the context of healthy behavioral changes.

The purpose of this paper is to provide an analysis on the needs of participants from the study by Albers and Brinkman (2023) regarding their preparatory activities. Preparatory activities are small tasks that the chatbot would suggest to participants so that they are better prepared for quitting smoking. Examples of such activities include creating motivational quotes for quitting smoking and focusing on past successes to become more physically active. The free-text feedback that participants provided for their preparatory activities, together with their personal characteristics, can be analysed to better understand people's needs when trying to quit smoking. As such, the research question that this paper aims to answer is:

What are users' needs for doing their preparatory activities?

Since this research question is fairly complex, it has been divided into 3 sub-questions, which will be referred to over the rest of this study. The 3 sub-questions are as follows:

1. *What are the key themes that arise from analysing users' experiences with preparatory activities?*
2. *What user needs can be identified from the key themes?*

3. *How do user characteristics influence the identified needs?*

The paper is structured as follows: Chapter 2 presents the methods that were followed to answer the research question. Chapter 3 delves into the ethics and reproducibility of the employed methods. Chapter 4 discusses the results, as well as the limitations of the analysis. Chapter 5 contains the conclusions and potential improvements to the current study, and Chapter 6 acknowledges the contributions of separate parties to this paper.

2 Methods

To answer the sub-questions and ultimately the main research question, the dataset from Albers and Brinkman (2023) was subjected to a thematic analysis that adhered to the steps outlined by Braun and Clarke (2006). More specifically, users' free-text responses have been encoded according to various particularities so as to uncover themes. The themes were then investigated to determine a set of user needs. Lastly, the influence of user characteristics on the needs was explored.

The study of Albers and Brinkman (2023) gathered data from the interactions of 671 daily smokers with the chatbot Mel between 21 July and 27 August 2023. The Human Research Ethics Committee of Delft University of Technology granted ethical approval for the research (Letter of Approval number: 2939). Every user interacted with Mel in at most 5 different sessions, each 2 sessions being at least about 2 days apart. During each session, Mel assigned the user an activity meant to reduce smoking or increase physical activity. Each activity was classified as being either a preparatory or a persuasive activity. Our interest resides in the preparatory activities, which were activities designed to help the user become more prepared to quit smoking by increasing self-reflection about their habits, as well as by helping them think about the consequences of their habits. The participants provided feedback for each activity during their next session with the chatbot. Each participant's age, gender, and highest completed education level were recorded.

2.1 Pre-processing the dataset

The dataset from Albers and Brinkman (2023) was initially pre-processed by the means of a Python script so that it was more aligned with our thematic analysis needs. Since our research is concerned with preparatory activities specifically, we removed the persuasive activities that were also part of the initial data set. Moreover, a large number of responses that concerned the current paper were left empty or the user did not provide enough information for proper analysis. These responses were given a specific code during the thematic analysis process, which showed that the respective answers were irrelevant to the study. By not completely removing such responses from the dataset, we ensured that statistical data regarding the meaningful response rate of certain activities could still be easily gathered in the future if such data is considered of interest.

2.2 Thematic analysis

The main aspect of the current research is the thematic analysis that was performed on the dataset. The purpose of such an analysis is to discover the themes or trends that reside within the data. This analysis was performed in ATLAS.ti Windows 23.4.0, which is specifically designed for qualitative analysis (ATLAS.ti Scientific Software Development GmbH., 2023). After importing the dataset into ATLAS.ti 23.4.0, the coding scheme was created. After creating an initial coding scheme, we used the findings of previous healthy behavior change studies as inspiration for improving it. Papers such as those by Asbjørnsen et al. (2020) and Albers et al. (2022) were especially insightful into how a coding scheme should be tailored specifically for understanding smoking cessation needs. As such, the current thematic analysis was performed in an inductive manner by creating our own codes, to which we added certain codes inspired by the aforementioned 2 papers. In the following 2 paragraphs we will outline how these papers were used as inspiration.

The results of Asbjørnsen et al. (2020), even though related to weight management eHealth applications specifically, can be fully applied to the context of smoking cessation. The paper discovered 8 user values that are key to developing optimal persuasive messages. These user values focus more on empowering the user and maintaining their self-regulation and well-being with values such as *Autonomy* and *Motivation*. As such, our final coding scheme included codes that deal with these values, such as the *Motivation* and *Positive attitude* codes, as well as the entirety of the *Personal empowerment* category.

Similarly, the paper by Albers et al. (2022) describes 14 themes grouped into 4 categories that determine users' efforts in the context of a smoking cessation eHealth application. Themes such as *Motivation to change* and *Getting help, advice or tips and learning* can be translated to themes also used in Asbjørnsen et al. (2020), which solidified the importance of including codes related to motivation and learning. Furthermore, its coding scheme includes codes for describing user behavior and mood, which inspired us to add the *Emotions* category, as well as codes like *Lifestyle changes* or *Superficial analysis* to our final coding scheme. The study of Albers et al. (2022) also underlined the importance of environment in determining user participation, thus giving it a separate category. To take into account aspects of users' environments, we included codes such as *External support*, *Impact on others* and *Obstacles*.

Taking into consideration the information that was gathered from the previously-mentioned papers, the final coding scheme of this study contains 71 codes, which were grouped into 11 categories (see Appendix A). Codes were grouped into categories according to their relations to other codes. For example, the codes *Cravings* and *Desire for change* both refer to different instances of the concept of desire, and were thus both grouped under the category *Desire*.

2.3 Double coding

Before beginning to search for themes, the results of the thematic analysis needed to be verified. The main option for reviewing qualitative analysis results, which have a high potential for researcher subjectivity, is some form of inter-rater reliability check. To this end, we employed the use of double coding, which sources such as Burla et al. (2008) deem to be a reliable tool for determining inter-coder reliability. This method required training another person (called the second coder) to re-code the dataset using our own scheme. As such, we trained another researcher with a background in Computer Science and Engineering, which then performed their thematic analysis completely independently from us. This method helped maintain the integrity and realism of a third-party coder that might want to perform an analysis using our coding scheme.

2.4 Quantitative analysis

The results of the second coder's thematic analysis were compared with our initial analysis using Cohen's Kappa measure of inter-rater reliability (Cohen, 1992). To calculate the value of Cohen's Kappa for the 2 thematic analyses, the information from both files were provided as input to a Python script. The script extracted the data from each of the 2 files and translated them into coding matrices. A coding matrix contains the free-text responses on its rows and the codes on its columns, with 0 and 1 values within the matrix indicating the absence or presence of a certain code within a particular response. After computing the 2 coding matrices, the script takes each pair of columns corresponding to the same code as 2 separate arrays, on which it computes the value of the *cohen_kappa_score* method from the *sklearn* library in Python.

The Kappa values computed in this manner were mostly high. Only 3 codes in our scheme yielded Cohen's Kappa values below 0.41: *Behavior change: Compulsive behaviors*, *General factors: Superficial analysis* and *Well-being: Time constraints*. A few codes yielded values between 0.41 and 0.60 (such as the case of the code *Behavior change: Lifestyle changes* and *Emotions: Ineffectiveness*) and a few received values between 0.61 and 0.80 (such as for the codes *Health: Health awareness* or *Personal empowerment: Positive attitude*). However, a surprising number of codes within the scheme yielded values for Cohen's Kappa that were between 0.81 and 1.00, which indicated that their reliability on this dataset was especially strong. For a lot of these codes, the result can be attributed to their low frequency within the dataset. An example would be the code *Financial: Costs of smoking*, which yielded a Kappa value of 1.00, but only appeared in 9 responses over the entire dataset. However, the Cohen's Kappa value was surprisingly high even for certain high-frequency codes, such as for the code *Well-being: Physical activity*, which appeared in 128 responses and yielded a Kappa value of 0.89.

The original interpretation of the Kappa measure, performed by Landis and Koch (1977), deemed values between 0.41 and 0.60 to represent moderate agreement between the 2 coders, values between 0.61 and 0.80 to indicate substantial agreement and values between 0.81 and 1.00 to suggest near perfect agreement. These interpretations are supported by the meta-analysis of Sun (2011) and the study of Warrens (2015). Taking into account these interpretations, we can safely say that the 2 thematic analyses agreed to a substantial degree. With this in mind, we started analysing our own thematic analysis for possible themes. The codes which yielded Kappa values below 0.41 were not taken into consideration for this process, however.

2.5 Finding the themes

Initially, we decided to group related codes into groups which corresponded to possible themes. While this was a good way of understanding how the codes were tied to the user experiences, a lot of the coded responses did not necessarily provide insights specifically into users' needs. To remedy this, a filtering of the responses was performed, removing all the quotations that did not contain insights into what users disliked. During this filtering process, the dataset was reduced from 685 responses to 131 meaningful ones. After analysing the filtered dataset, we decided that it would be more reliable to look for themes within pairs of codes that had a high correlation within the filtered dataset.

More specifically, we performed a co-occurrence analysis using the in-built tool from ATLAS.ti Scientific Software Development GmbH. (2023). Co-occurrence analysis involves computing the frequencies to which a code appears together with certain other codes within the same quotations (which are named co-occurrence values). The result of this analysis is a table which contains co-occurrence values for each pair of codes within the scheme. With the help of this in-built tool we analysed the quotations that were attributed pairs of highly co-occurring codes. For example, while analysing the quotations that were coded with both *Emotions: Disinterest* and *General factors: Lack of involvement* (which yielded a high co-occurrence value), we discovered a large percentage described how the repetitiveness or perceived ineffectiveness of the activities led to users not performing said activities. These responses could all be categorised by a general trend of users who request, directly or indirectly, more varied, engaging and meaningful activities. This in turn transformed into the **More interesting activities** theme.

After performing similar analyses for other highly co-occurring codes, we reached a set of 9 potential themes. These categories were then subjected to an analysis by both us and the second coder from the double coding process. The second coder analysed the themes and the types of quotations that were part of each theme, and after a brainstorming session provided us with suggestions on how to tailor the themes to the problem even further, as well as how to find potential solutions for meeting users' needs in each of the themes.

During this process, one of the 9 initial themes, named *Easier/simpler activities*, was removed. This potential theme included quotations such as *"I really tried to focus a few* *times doing the activity but my body and home life made it impossible to do"* and *"i tried to visualise it but found it quite difficult"*. Preparatory activities did not require users to perform any demanding actions. They required thinking about smoking and physical activity habits, about the desired self, as well as watching or reading educational materials related to smoking cessation. After a short discussion with the second coder, we reached a decision that these responses only reflected a vague disinterest in the activities, without providing insights as to why that is the case. Hence, we scrapped this potential theme from the remainder of the study, and we were left with a list of 8 themes.

Some of the responses in the dataset expressed what the user needed indirectly. An example would be the one provided by participant P446 during their third session with the chatbot (*"I thought about it but just forgot"*). In this case, the underlying theme would be that users would benefit from adding reminder functionalities to the application. Certain quotations, however, contained a direct request from a user describing what they need, such as the case of participant P112, who wrote after their fourth session: *"I really don't like the idea of recording myself ... I felt like that was too invasive and would ruin my routines"*. A response such as this one can provide insight into the fact that users require less confidence-damaging activities.

Each of the 8 themes corresponds to a subgroup of quotations which expressed similar complaints or suggestions from the users. Most of the quotations in the dataset corresponded to a single theme, while a few which were more expressive corresponded to 2 or more themes simultaneously. A small number of the quotations could not be tied to any specific theme, as they only expressed a general dislike of the application, without providing enough detail for the reason behind the dislike. The names of the 8 themes, together with their descriptions, are provided below:

- **Less distressing activities** - The user needs activities that are less focused on the negative consequences of smoking. Certain users explained how they found activities that required them to search for pictures of ill smokers distressing. Thinking about or visualizing the consequences of smoking is connected to increasing motivation to not smoke, as mentioned in the paper by Krosnick et al. (2017). However, putting the participant in contact with shocking concepts that they are not accustomed to thinking about will inevitably increase their stress levels, which has been correlated with increased chances to smoke again, as discovered by the study of McKee et al. (2010). As such, a possible solution would be to accommodate the user with the negative consequences of their habits gradually. Instead of some users being asked directly to think about these negative aspects, the chatbot can always ask them first about their smoking and physical activity habits. Doing this will help users reach conclusions about the negative consequences of smoking by themselves before being directly asked to do so.
- **More interesting activities** - A general theme referring to users' need for more engaging and meaningful activities. This category contains responses that described participants' disinterest in their activities either because they already performed similar activities, or because they found the activities pointless or boring. To analyse this theme, an important distinction needed to be made: the user's interest in the activities is different from the user's interest in quitting smoking. As such, the responses that suggested the user was disinterested in quitting smoking were not taken into consideration for this theme. By providing a wider range of activity types, as well as a wider range of activity complexities, the application can ensure that a bigger range of participants who are willing to quit smoking can more easily follow the activities posed by Mel.
- **More knowledge** - This theme contains responses which indicated that users wanted to receive more detailed knowledge on the effects of smoking and how to reliably quit. A number of users mentioned how the educational videos that Mel suggested did not contain any information they did not already know, while other users showed that they did not yet understand the purposes of certain aspects of smoking cessation, such as nicotine replacement therapy. Moreover, there was a division amongst the quotations that were part of this category: a number of people watched or read the entire material and afterwards realised that the material did not contain any new information, while the rest stopped before finishing the material. Only the responses which indicated that the user stopped before finishing the activity were selected for this theme, as one of the main goals of the chatbot is to ensure the involvement and completion of activities. To reduce the frequency of this theme in the future, the chatbot can ask each user about their previous knowledge related to smoking cessation techniques. This can help the chatbot give mostly new knowledge to each user, so as to combat misconceptions about smoking cessation.
- **More motivation** - A theme containing the responses of people that were motivated to quit smoking, but were not sufficiently motivated to do their assigned activities. While motivation to quit smoking is a must when undergoing healthy behavior changes, the scope of this study does not include building motivation to quit smoking. It does, however, include building motivation to participate in the activities for people who already wish to quit smoking. Increasing motivation can help promote healthy behavior change, as outlined by the study of Sheeran et al. (2021), and since the ultimate purpose of the study is to help people reliably perform such changes, meeting this need is important.
- **Confidence-maintaining activities** - The chatbot needs to suggest activities that maintain or increase users' self-confidence. Participants that are self-conscious or uncertain about themselves tend to not follow activities that require them to expose themselves. Even though this theme only contains 2 quotations, the quotations were unique enough that we decided they deserved their own

theme. Both quotations in this theme mentioned privacy as a complaint (*"I really don't like the idea of recording myself, or how much I smoked in a situation. I felt like that was too invasive and would ruin my routines. So unfortunately I haven't even tried to do the activity"*, *"I have done this in the past and actually have a list on my phone to look at when I'm in quit mode. I do not like having notes around due to privacy issues in my living space."*. The activities suggested by Mel did not require users to expose themselves in any way. Analysing the responses with this consideration in mind, however, suggests that these users were unwilling to perform activities because they could damage their self-confidence. One solution for these users would be providing multiple options for completing a single activity. The activities in question can then suggest making a video or writing about the user's habits, and if they do not feel confident enough to do so they can just think about the topic.

- **Reminders** - A large number of users mentioned not doing their activities because they forgot amidst their busy schedule. Helping users remember about their assigned activities can help increase their involvement. This idea is supported by the study of Fry and Neff (2009), which suggests that regular reminders have positive effects on health behavior interventions.
- **Analytics of tasks** - Certain users benefit from being able to track their smoking cessation progress. These participants typically used third-party applications for taking notes or maintaining schedules. The chatbot program could then include its own habit-tracking features, so that people do not have to rely on multiple applications for their smoking cessation needs. A number of the most successful applications for smoking cessation as of 2024, such as Flamy and QuitNow! use habit tracking features to incentivize users (Offlinefirst, 2017; Fewlaps, 2010). This theme ties in with **More motivation**, as the whole purpose of the analytics features is to motivate users by helping them visualize their progress.
- **Weight-managing tips** - Some users mentioned their concerns about weight increase when quitting smoking. This theme is tied to the user's need for meaningful knowledge, since improper understanding of the consequences of smoking cessation on weight will increase their stress levels, which will lead to more smoking, as mentioned by Hughes (2003). For these users, the application can include more information about how weight is affected by quitting smoking and how they can reliably manage their weight in these situations. Gathering of user knowledge statistics about weight management before the use of the application can thus help provide more persuasive information during the use of the application, as in the case of the **More knowledge** theme.

These 8 themes formed the answer for our first research sub-question. Fortunately, these themes also helped us directly outline a clear set of needs, as all of these responses directly or indirectly suggest what the user did not enjoy. As such, each of these 8 themes describes a specific need, and thus our answer for the second research sub-question. The needs corresponding to the themes are as follows:

1. **Users should not be distressed by the activities.**
2. **Users need more engaging and meaningful activities.**
3. **Users need to be provided with knowledge that is varied and deep enough to maintain interest.**
4. **Users that are motivated to quit smoking need to be motivated to also maintain high participation with the activities.**
5. **Users need to be assigned activities that maintain or increase their self-confidence.**
6. **Users need to be reminded of their assigned activities and future sessions with the chatbot.**
7. **Users should be able to track their smoking-related habits from within the application itself.**
8. **Users need to be well informed by the application about the effects of smoking cessation on weight, as well as any advice about managing weight as a quitter.**

From this point onwards, we will refer to the 8 user needs by their respective themes, since they are more concise. Hence, whenever we will refer, for example, to the **More interesting activities** theme, we are essentially referring to the user's need for more engaging and meaningful activities.

2.6 Understanding the participants

To correlate the results of the study with the characteristics of the participants, certain anonymous information about them was gathered. Each participant's age, gender and education level was recorded into a file. To connect each participant's characteristics to their free-text responses, their random ID (called PID) was used. To better understand the types of people that the study contained, we created a Python script to calculate various statistical data about them. Of the initial 671 participants to the study, 646 passed the checks required to be taken into consideration for the purpose of this study (i.e. they passed all the attention checks given by Mel during the sessions).

The results of the script show that the ages of these 646 participants have a mean of 38.4 years and a standard deviation of 11.4 years. 344 of the 646 (53.2%) participants identified as female, 292 (45.2%) identified as male and 10 (1.5%) identified as non-binary. No user withheld the information about their gender identity, despite the fact that the option "Rather not say" was present in the study's pre-questionnaire. Regarding their education level, 5 people had no formal qualifications, 78 finished secondary education, 164 received their high school diploma or A-levels, 105 finished a technical or community college, 200 received their undergraduate degree, 86 received their graduate degree and 5 finished their doctorate. 3 participants selected the answer "Don't know / not applicable" for this question in the pre-questionnaire. The ages of the participants ranged from 18 to 85 years, with a mean of 38.4 and a standard deviation of approximately 11.4.

After filtering the dataset and keeping only the 131 quotations that were relevant for understanding the user needs, we were left with the following statistics: 65 (49.6%) participants identified as female, 6 (4.5%) as non-binary and 60 (45.8%) as male. Their education levels included 1 person who had no formal qualifications, 12 who finished secondary education, 38 who finished high-school or A-levels, 23 who finished technical or community college, 39 who received their undergraduate degree, 16 who received their graduate degree and 2 who received their PhD. The ages of these users ranged from 21 to 73 years old, with a mean of 40.2 and a standard deviation of 12.2.

2.7 Correlation with user characteristics

To understand the presence of the themes in the filtered dataset, we have performed statistical analyses in which we computed the total number of quotations that exhibited each of the 8 needs, the number of participants who identified as each of the the 3 available gender options (i.e. female, non-binary and male respectively), as well as the mean, standard deviation, minimum and maximum age of the participants that exhibited each need. A table which contains all of this statistical data has been added to this end (see Table 1), as well as a bar chart in which this data can be more intuitively visualized (see Figure 1).

	Number quotations	Number female	Number non-binary	Number male	Mean age	SD age	Min. age	Max. age
Easier/simpler activities	29	12	1	16	45.5	13.7	22	72
Less distressing activities	4	3	0	1	48.2	14	24	58
More interesting activities	44	24	3	17	41	11.6	21	73
More knowledge	15	7	3	17	42	9.9	27	56
More motivation	9	3	1	5	42	12	23	56
Confidence-maintaining activities	2	1	0	1	32	-	31	33
Reminders	29	14	1	14	37.2	10.7	22	64
Analytics of tasks	5	2	0	3	36.4	9	30	54
Weight managing tips	3	2	0	1	45.3	-	35	51

Table 1: General statistics regarding user characteristics

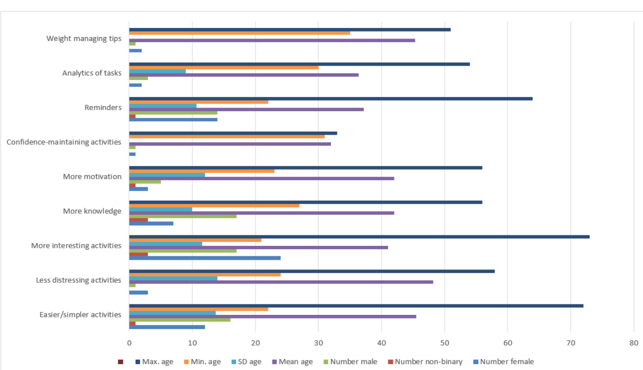

Figure 1: User characteristics bar chart

To compute the correlation of the themes with the user characteristics, we created separate Python scripts, each with a different type of test. The details of each of these tests are mentioned in the following sub-sections. Their results constitute our answer for the third research sub-question.

Gender identity

To compute the correlations between users' genders and each of the themes, we calculated the chi-square test value between the number of quotations given by women and the number of quotations given by men. The test was implemented using the *chi2_contingency* method in the *scipy.stats* library in Python. However, since one of the pre-requisites of the chi-square test is that the numbers in its contingency table should be at least 5, this meant that we could only compute the P-values for 3 themes, namely **More interesting activities**, **More knowledge** and **Reminders** (see Table 2). Unfortunately, most of the themes did not contain enough quotations to perform the test, and we were not able to perform any tests on the frequencies of non-binary people, as no theme contained more than 3 quotations from them. For the **More knowledge** and **Reminders** themes, the P-values are the maximum possible of 1.0, which is understandable considering the exact same frequencies of men and women (see Table 1). For the **More interesting activities** theme, the P-value is 0.58, which is still too high from the usual threshold of 0.05 to be considered significant.

	Gender chi-square P-value	Age Pearson P-value	Education Spearman P-value
Less distressing activities	-	0.26	0.84
More interesting activities	0.58	0.16	0.28
More knowledge	1	0.2	0.1
More motivation	-	0.31	0.95
Confidence-maintaining activities	-	0.74	0.34
Reminders	1	0.54	0.81
Analytics of tasks	-	0.96	0.68
Weight management tips	-	0.29	0.05

Table 2: Values related to the correlation tests

Age distributions

To calculate the correlations of the themes with the ages of the participants, we applied a Pearson correlation using the *pearsonr* method from the *scipy.stats* library in Python. For this analysis we applied age bins of 10 years, so the age intervals of the bins were 20-29, 30-39, until 70-79. Since the dataset of 131 quotations did not contain participants with ages younger than 20 or other than 79 years old, we did not have to add a separate bin for the intervals 18-19 and 80-85 (which was the maximum recorded age in the entire study). The results of the script can be seen in Figure 2, and the corresponding P-values can be seen in Table 2. 5 of the themes have correlation values around 0.10, the highest being that of **More interesting activities** with 0.125. This might suggest a slightly noticeable trend of people exhibiting those 5 themes as they increase in age. The other 3 themes have values that are too low to be taken into consideration. All P-values corresponding to the age correlations are considerably higher than the standard threshold of 0.05, which indicates that there is no significant correlation between participants' ages and the themes.

Highest completed education levels

To correlate the themes with users' highest completed education levels, we employed the use of Spearman's rank correlation using the *spearmanr* method from the *scipy.stats* library in Python. To make the education-related data compatible with this test, we had to order the options from least to most

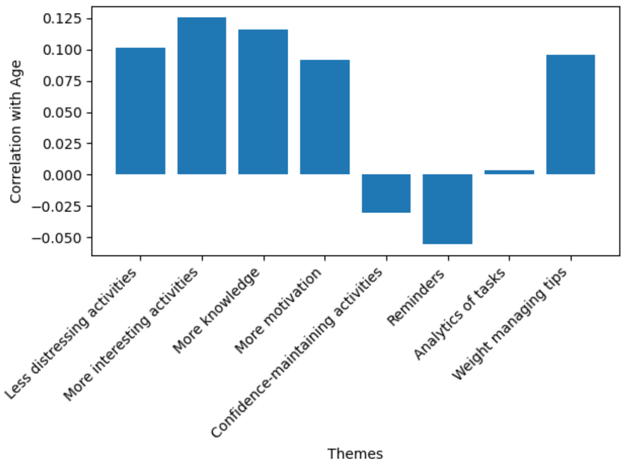

Figure 2: Pearson correlations between the themes and age groups

amount of education. More specifically, we mapped all options for education levels to scalars, so the "no formal education" option was mapped to 0, "Secondary education (e.g. GED/GCSE)" was mapped to 1, etc. The results of the script can be seen in Figure 3, and the P-values corresponding to the correlations can be observed in Table 2.

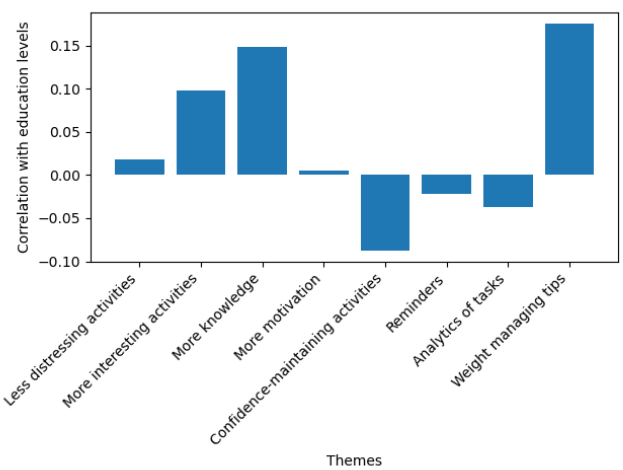

Figure 3: Correlations between the themes and highest completed education levels

The theme **Weight-managing tips** has the highest value of approximately 0.17. What is interesting is that its corresponding P-value is 0.05, which is exactly the value of the usual threshold for this type of test. In general, if a P-value is below 0.05, the correlation is deemed significant. In our case, however, since it is directly at the limit of significance, we can say that there might be a slight connection between users with more qualifications and concerns about weight management, but more data is required to know whether the correlation is significant. Other noteworthy themes in the figure are **More knowledge** and **Confidence-maintaining activities**, but their P-values are too large to be considered significant. However, due to the low frequencies of these themes within the dataset, more research must be performed to draw any conclusions.

3 Discussion

3.1 Consequences of the needs

What is surprising about this study is the overall stability of the observed needs in relation to the user characteristics. None of the results clearly shows a strong correlation between a specific subgroup of users and a particular need, which should be the desirable outcome for the chatbot application. The chatbot should meet people's smoking cessation needs regardless of characteristics such as their gender, age or education. While some of the higher-valued correlations deserve more in-depth analysis in the future to determine whether or not they are meaningful, we can generally say that the users have similar needs. Some of the results, such as the low correlation between the need for **Reminders** and people's age, may seem improbable due to the general idea that older people are more forgetful (Gold and Korol, 2014). Such results can be attributed to the following factors:

- The number of quotations which were relevant to the study of user needs was quite low. Out of 646 quotations related to preparatory activities, only 131 quotations were selected, which reduced the dataset from 671 to 121 different people. As such, the correlations that were calculated might not have been stabilized due to low sample size. Moreover, for the correlation with users' gender, most themes did not have high enough frequencies to compute reliable statistics.
- A lot of participants did not share enough meaningful feedback with the chatbot and did not describe how they perceived the preparatory activities, and as such we were left with a relatively small dataset of relevant data. This can be attributed to user preconceptions, as well as the financial incentive that users received when entering the study.

3.2 Limitations

The main limiting factor of this research is the lack of data that is relevant for the purpose of understanding user needs. From a dataset of interactions with 671 daily smokers, only 131 responses were related to users' needs, with a large percentage of these responses requiring us to read between the lines for any meaningful information. People might not yet be accustomed to conversing with chatbots in that much detail, as they might not trust it. The financial incentive that the study participants received might have altered the results too, considering the high number of users which provided no meaningful information. Moreover, the results might not be applicable to the general population due to the high age mean and a lack of people of certain educational levels.

Furthermore, the methods used in this study are inevitably subject to researcher bias. Even though the dataset was anonymized before we first came into contact with it, the coding of the individual quotations is reliable on several subjective factors. The researcher might be influenced by the manner in which a user phrases their feedback. Sometimes, the user's choice of words leads to the researcher being confused. The researcher might be influenced by other quotations they recently coded, which can heavily influence the manner in which they code new quotations. Bias was undeniably introduced since the thematic analysis was performed by one person, with the addition of a double coder, both of which had the same educational background. The researcher and the second coder might have had very similar biases, leading to the high Kappa values by chance. To help reduce potential bias, future studies can reproduce the thematic analysis outlined in this paper with multiple coders of different backgrounds.

One other limitation arises from the reduced scope of the performed analyses. Due to limited time, the thematic analysis might have led to less quotations being thoroughly investigated. There were a lot of aspects from this research that deserved more time and analysis, such as the correlations of the needs and users' effort values, as well as their non-smoker and quitter self-identities. Hopefully, once this data becomes available to the public, more research can be done into these aspects.

4 Responsible Research

4.1 Ethical considerations

To maintain user anonymity, all sensitive information that could help identify them has been removed before we received the dataset. The dataset from the study of Albers and Brinkman (2023) is not currently available for public use, as it is currently being used for another study, but will be made available on the TU Delft Repository once it is finished. The link to the entry on the TU Delft Repository will be made available on the OSF-form for that study (https://osf.io/nuy4w). The dataset used for this study will be made available on 4TU.ResearchData, DOI: 10.4121/257decd4-5980-4731-a368-67abfca7c308.

4.2 Reproducibility

We consider that this research paper provides sufficient information regarding the employed methods and analysis choices, as well as the over-arching study that it is part of, so that readers will be able to reproduce, test, and build upon it. The file containing the 131 quotations used for the correlations contains codes from the *"Needs"* category in the *"Codes"* column. These codes represent the 8 themes and their respective needs, and were added for ease of computing the final results.

The steps taken to perform the thematic analysis, as well as the double coding process and the following quantitative analysis have been thoroughly explained in the paper. The paper also mentions the statistical tests used to correlate the themes to the user characteristics, as well as the software tools that have been used to implement them. Lastly, any observed results, limitations and potential improvements have been stated, as they can prove useful to future research.

5 Conclusions and Future Work

To summarize, we performed an analysis on the needs of daily smokers who interacted with a chatbot designed to help them quit. A thematic analysis was performed on the dataset of interactions, developing a set of 8 themes, each describing a recurring aspect of the interactions which led to sub-optimal user experiences. The reliability of the thematic analysis results was verified by the means of double coding. Each of the discovered themes led us to a corresponding need of the users, which we correlated with their gender, age and highest completed education level. Most of these correlations showed that the themes were not significantly connected to the user characteristics, and our results suggest that in general users have similar needs regardless of gender, age and education level. However, a couple of these correlations require further analysis to determine whether the results are significant.

This study provides a building block for further advancements in the domain of healthy behavior changes, especially in the context of smoking cessation. In order to ensure the effectiveness of eHealth applications, we must properly understand what users need. To build upon this study, we suggest investigating the highlighted correlations further. We also recommend implementing and studying the effectiveness of the solution ideas outlined in the Methods section. Moreover, this study will benefit from performing the thematic analysis on a bigger dataset, with more coders from different backgrounds, so as to reduce bias.

6 Acknowledgement

This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders organized by the Netherlands Organization for Scientific Research (NWO), program Commit2Data - Big Data & Health (project number 628.011.211). Besides NWO, the funders include the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, the Ministry of Health, Welfare, and Sport (VWS), Health Holland, and the Netherlands eScience Center.

We are acknowledging the contributions of Aaron Maguire, who acted as our second coder during the double coding phase of the study, and who provided various pieces of code that aided us in accomplishing the goal of this paper. To this end we are also acknowledging the contributions of Jonathan van Oudheusden.

Appendices

A Thematic analysis codes

Code - Comment
Behavior change
Compulsive behaviors - Shows that smoking is a compulsion for them. They have trouble withholding themselves from smoking.
Coping strategies - Activities that the user does to refrain from smoking. Not necessarily a distraction, but something that helps with cravings.
Habits - Any talk about what users habitually do, about their routine activities.
Lifestyle changes - Major changes to one's lifestyle. Could be starting physical activity, changing environments, quitting cold turkey, etc.
Obstacles - Any mention of obstacles to attaining one's goal.

Desire
Cravings - Any mention of cravings or powerful wishing to smoke.
Desire for change - The user shows that they clearly want to change.

Effort
All of activity - User performed all the required parts for the activity.
Nothing - User didn't do any of the activity.
Part of activity - The user performed only a part of the activity.
Related to activity - The user didn't perform what the activity required, but prefered doing something similar.

Emotions
Anxiety - User exhibits feelings of anxiety, worriness or ruminates over an activity.
Depression - User has a general state of negativity, sadness, hopelessness, or other similar feelings.
Disinterest - User exhibits a lack of interest in the activity or in general in quitting smoking.
Dislike - User doesn't like the activity.
Dissapointment - User is dissapointed with themselves or their performance with the activities/journey.
Frustration - User is frustrated with parts of their journey (e.g. inability to quit, previous failures, etc.)
Happiness - User exhibits a generally positive/uplifting attitude. This is an overencompassing code for clear positive emotions.
Ineffectiveness - The user doesn't feel thta the activity worked, or it is clear for the coder that it didn't have the expected results.
Interest - User exhibits interest in the activity and performing it.
Kindness - User experienced kindness from someone or did something kind to other people.
Motivation - User mentions motivation or is motivated to quit smoking/perform activities.
Uncertainty - User is uncertain on themselves or whether the activities will work.

Financial
Costs of living - User mentions the financial aspects of day-to-day life and routine.
Costs of nicotine replacements - User mentions the costs of nicotine replacement products (i.e. nicotine patches).
Costs of smoking - User mentions the costs of their smoking habit.

General factors
Irrelevant to study - The quotation is not relevant for the study. This is what I used for the 'none' quotations.
Lack of confidence - User is not confident in themselves or the process.
Lack of involvement - User is not involved enough with the activities, procrastinates or in general doesn't put the required effort into their interactions.
Lack of privacy - User mentions a lack of privacy either within their interactions with Mel or within their personal lives.
Superficial analysis - User didn't provide enough details in their free-text responses so as to perform a meaningful analysis.
Visualization - User visualized something. Imagining things, fantasies, visual mementos and thinking about something that feels vivid are all deemed to be visualizations.

Health
Health awareness - User shows that they are aware of health aspects. I didn't consider mentioning health benefits or concerns as health awareness, since I have separate codings for those things. If the user mentions health as a factor in general then it is health awareness.
Health benefits - User mentions the health benefits of quitting smoking, doing more physical activity, etc.
Health concerns - User is concerned about health or mentions negative health aspects (e.g. diseases, shortness of breath, etc.)
Mental health - User mentions or is concerned with mental health aspects.
Weight - User mentions struggling with weight management, the impact smoking cessation has on weight, etc.

Human experience
Alcohol - User mentions consuming alcohol.
Caffeine - User mentions consuming coffee or other caffeinated products.
Excuses - User gives an excuse, usually to justify why they didn't do/complete an activity.
Food - User mentions food or eating.
Forgetfulness - User forgot to do an activity.
Impact on others - User mentions the impact smoking/quitting smoking has had/will have on people around them (usually when they mention family/children, also when they mention close ones that suffered from smoking).
Inspiration - User mentions inspiration or shows that they drew inspiration from different sources/people/activities/etc.
Remembrance - User recalls things from the past..

Personal empowerment
Goal setting - User clearly set a goal for themselves.
Knowledge enhancement - User learned something new.
Positive attitude - User has a generally positive attitude or mindset towards the activities or their journey. This is used more frequently than the Happiness coding, and is used when happiness isn't that clearly present.
Self-confidence - User mentions or is confident in their own abilities/progress/etc.
Self-discipline - User maintained a routine, had self-control in dealing with their cravings, and in general was very stoic or disciplined.
Self-improvement - User exhibits an improvement with regards to their smoking journey. Some cases might be the increase in frequency of healthy behaviours, reducing smoking frequency, and in general feeling like they improved.
Self-reflection - User reflects on themselves or the activites.

Smoking cessation
Addiction - User directly mentions addictions.
Consequences of smoking - User mentions the consequences of smoking. It is more general than the "Physical effects of smoking" coding, but I was a bit more loosey-goosey with these 2 things. It depends more on the nuance.
Impact of quitting smoking - User talks/is concerned about the negative aspects of QUITTING smoking (e.g. weight gain, sleeplessness, etc.)
Nicotine replacement - User mentions any options for nicotine replacement, except other forms of smoking (so nicotine patches essentially, not smoking).
Physical effects of smoking - User talks about the more physical effects of smoking (lack of fitness, shortness of breath, coughing, etc.)
Smell of smoke - User mentions the smell of smoke.
Smoking alternatives - User mentions any smoking alternatives (i.e. vape, e-cigarette).
Triggers - User mentions triggers or situations that gets them to smoke.

Well-being
Alternative therapy - User mentions alternative forms of therapy when dealing with smoking. Here I included all things that feel more New Agey, like breathing techniques, yoga, etc.
Busyness - User mentions being busy or it can be inferred that they had a lot of things to do.
Distraction - User mentions distracting themselves from smoking or can be inferred that they distracted themselves.
External support - User had support from any tool or person outside the study (e.g. used Google to search for things, looked at other videos on YouTube, or talked to other people about their journey).
Physical activity - User mentions performing any sort of physical activity except yoga (including walking).
Physical limitations - User mentions any sort of physical limitations or disabilities.
Relaxation - User mentions feeling relaxed or doing activities with the purpose of relaxation.
Sleep quality - User mentions anything related to sleep.
Stress management - User mentions anything related to stress and managing it.
Time constraints - User mentions any sort of constraints on their schedule, usually with the purpose of excusing themselves for a lack of participation.
Time management - User mentions anything related to managing their time and schedule.

References

Aggarwal, A., Tam, C. C., Wu, D., Li, X., and Qiao, S. (2023). Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research*, 25.

Albers, N. and Brinkman, W.-P. (2023). Perfect fit - learning to propose useful preparatory activities for quitting smoking and becoming more physically active.

Albers, N., Neerincx, M. A., Aretz, N. L., Ali, M., Ekinci, A., and Brinkman, W. P. (2023). Attitudes toward a virtual smoking cessation coach: Relationship and willingness to continue. volume 13832 LNCS, pages 265–274. Springer Science and Business Media Deutschland GmbH.

Albers, N., Neerincx, M. A., Penfornis, K. M., and Brinkman, W. P. (2022). Users' needs for a digital smoking cessation application and how to address them: A mixed-methods study. *PeerJ*, 10.

Asbjørnsen, R. A., Wentzel, J., Smedsrød, M. L., Hjelmesæth, J., Clark, M. M., Nes, L. S., and van Gemert-Pijnen, J. E. (2020). Identifying persuasive design principles and behavior change techniques supporting end user values and needs in ehealth interventions for long-term weight loss maintenance: Qualitative study. *Journal of Medical Internet Research*, 22.

ATLAS.ti Scientific Software Development GmbH. (2023). ATLAS.ti Windows (version 23.4.0.29360). https://atlasti.com. [Qualitative data analysis software].

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3:77–101.

Burla, L., Knierim, B., Barth, J., Liewald, K., Duetz, M., and Abel, T. (2008). From text to codings: Intercoder reliability assessment in qualitative content analysis. *Nursing Research*, 57:113–117.

Cohen, J. (1992). Quantitative methods in psychology a power primer.

Cooper, J., Borland, R., Yong, H.-H., McNeill, A., Murray, R. L., O'Connor, R. J., and Cummings, K. M. (2010). To what extent do smokers make spontaneous quit attempts and what are the implications for smoking cessation maintenance? findings from the international tobacco control four country survey. *Nicotine & Tobacco Research*, 12(suppl_1):S51–S57.

Dai, X., Gil, G. F., Reitsma, M. B., Ahmad, N. S., Anderson, J. A., Bisignano, C., Carr, S., Feldman, R., Hay, S. I., He, J., et al. (2022). Health effects associated with smoking: a burden of proof study. *Nature Medicine*, 28(10):2045–2055.

Fang, Y. E., Zhang, Z., Wang, R., Yang, B., Chen, C., Nisa, C., Tong, X., and Yan, L. L. (2023). Effectiveness of ehealth smoking cessation interventions: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 25:e45111.

Fewlaps (2010). Quitnow: Quit smoking for good (version 6.2.8) [mobile app].

Fry, J. P. and Neff, R. A. (2009). Periodic prompts and reminders in health promotion and health behavior interventions: Systematic review. *Journal of Medical Internet Research*, 11:e1138.

Gold, P. E. and Korol, D. L. (2014). Forgetfulness during aging: An integrated biology. *Neurobiology of Learning and Memory*, 112:130–138.

He, H., Pan, Z., Wu, J., Hu, C., Bai, L., and Lyu, J. (2022). Health effects of tobacco at the global, regional, and national levels: Results from the 2019 global burden of disease study. *Nicotine Tobacco Research*, 24:864–870.

Hughes, J. R. (2003). Motivating and helping smokers to stop smoking epidemiology of smoking cessation.

Krosnick, J. A., Malhotra, N., Mo, C. H., Bruera, E. F., Chang, L. C., Pasek, J., and Thomas, R. K. (2017). Perceptions of health risks of cigarette smoking: A new measure reveals widespread misunderstanding. *PLOS ONE*, 12:e0182063.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159.

Lentferink, A. J., Oldenhuis, H. K., Groot, M. D., Polstra, L., Velthuijsen, H., and Gemert-Pijnen, J. E. V. (2017). Key components in ehealth interventions combining self-tracking and persuasive ecoaching to promote a healthier lifestyle: A scoping review. *Journal of Medical Internet Research*, 19:e7288.

McKee, S. A., Sinha, R., Weinberger, A. H., Sofuoglu, M., Harrison, E. L., Lavery, M., and Wanzer, J. (2010). Stress decreases the ability to resist smoking and potentiates smoking intensity and reward. *http://dx.doi.org/10.1177/0269881110376694*, 25:490–502.

Meijer, E., Korst, J. S., Oosting, K. G., Heemskerk, E., Hermsen, S., Willemsen, M. C., van den Putte, B., Chavannes, N. H., and Brown, J. (2021). "at least someone thinks i'm doing well": a real-world evaluation of the quit-smoking app stopcoach for lower socio-economic status smokers. *Addiction science & clinical practice*, 16(1):1–14.

Mokdad, A. (2004). Actual causes of death in the united states, 2000. *Jama*, 291:1238–1245.

Offlinefirst, L. (2017). Stop smoking with - flamy (version 6.2.8) [mobile app].

Sheeran, P., Wright, C. E., Avishai, A., Villegas, M. E., Rothman, A. J., and Klein, W. M. (2021). Does increasing autonomous motivation or perceived competence lead to health behavior change? a meta-analysis. *Health Psychology*, 40:706–716.

Sun, S. (2011). Meta-analysis of cohen's kappa. *Health Services and Outcomes Research Methodology*, 11:145–163.

Warrens, M. J. (2015). Five ways to look at cohen's kappa. *Journal of Psychology Psychotherapy*, 5:4.

Zhang, J., Oh, Y. J., Lange, P., Yu, Z., and Fukuoka, Y. (2020). Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint. *Journal of Medical Internet Research*, 22.