



# **Virtual Coaching for Smoking Cessation: What are Users Preference in Ethical Principles for Human Feedback Allocation**

**Glebs Labunskis**

**Supervisors: Dr. ir. W.P. Brinkman, Ir. N. Albers**

**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  
June 22, 2024

Name of the student: Glebs Labunskis  
Final project course: CSE3000 Research Project  
Thesis committee: Dr. ir. W.P. Brinkman, Ir. N. Albers, Jonathan van Oudheusden

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

## Abstract

**Background.** Quitting smoking is a challenge nowadays. Virtual coaches offer autonomous, personalized guidance for smoking cessation. However, such systems cannot replace human coaches completely. In situations, when human coaches cannot provide help to everyone - a virtual coach could follow a set of ethical principles to decide on who should get the feedback from a human.

**Objective.** Our study aims to identify users' preferences on ethical principles that a virtual coach should follow to decide when to allocate human feedback to individuals preparing to quit smoking.

**Methods.** Our research was based on pre-gathered data, that included participants' responses to open and closed questions regarding feedback allocation principles. Thematic analysis was conducted on these responses. Triangulation was performed using a qualitative literature review and quantitative data analysis.

**Results.** Four main themes were identified: (1) **Struggling the Most** (63.75%), (2) **Increasing Chances of Success the Most** (13.75%), (3) **Equal Treatment** (11.25%), and (4) **Appreciating the Most** (11.25%). Participants prioritized support for those experiencing the greatest difficulty in smoking cessation. The triangulation supported the validity of these themes.

**Conclusions.** Our study highlights the importance of integrating user-preferred ethical principles in virtual coaching systems for smoking cessation. Prioritization of users who struggle the most can increase the effectiveness and fairness of such systems, potentially increasing success rates. Future research should explore additional ethical principles, combining several principles into systems, and real-world application of these findings to further refine virtual coaching in healthcare.

## 1 Introduction

Smoking is a widely spread harmful addiction today. According to the World Health Organization (WHO), over 22 percent of the world is smoking or vaping daily, exposing themselves to significant health risks [29]. Smoking cessation, however, is a challenging process accompanied by a wide range of withdrawal symptoms [8]. Therefore, willpower and dedication are essential during the whole process of quitting smoking, including the preparation stage [26]. Among various strategies like group therapy or nicotine replacements, a virtual coach is a tool that stands out with its independence, autonomy, and personal approach [25].

Having a virtual assistant on a mobile device provides motivation and constant guidance for users, especially for those who find quitting smoking tough and demanding [11]. Although powerful, virtual coaches cannot fully replace humans

yet, as it would require a much broader understanding of the world than such AI would have today [23].

*Albers and Brinkman et al.* held an experiment involving over 850 at the start and finishing with more than 470 participants, in which a virtual helped users prepare for smoking cessation [3]. Although the chatbot was the main consultant for the participants, Human coaches with a background in psychology were involved too, providing human feedback. *Balch et al.* concluded, that participants are more likely to receive and perceive help from trusted people - such as psychologists or successful quitters [4]. During the *Albers and Brinkman et al.* experiment the decision of whether to give a particular participant human feedback was made randomly, to be more precise - 20 percent of the time. As a result, some users might have needed human feedback, however, due to limited resources, have not received it. *Gromov et al.* [13] stressed the importance of optimal limited aid allocation. Furthermore, *Furnham et al.* [12] state that many physicians and hospital managers refer to the allocation of resources as one of the most demanding and unresolved problems within their profession. To make preparing for smoking cessation more helpful and consistent, the chatbot should have a setting on when to assign human feedback to the user. Human opinions should be gathered and analyzed to fit the solution as tightly to the problem as possible [16].

Other medical fields already researched what principles to follow when deviating from the default treatment methods. For example, *Persad et al.* [24] presented eight moral principles of allocation and three systems of combining those. They concluded that no principle could solely recognize all morally relevant values, therefore a list of such principles should be considered.

*Abma et al.* [1] compared ethical principles guidelines that four countries use in their ICU program. The author concludes that even when discussing aid allocation within one field (ICU), opinions on ethical principles differ per country.

*Krüti et al.* [18] state and prove with the research that a universal list of ethical principles does not exist. The author concludes that users prefer allocating the aid differently, depending on the case of study.

Therefore, the specifics of smoking cessation problem should be taken into account, when choosing the principles to follow.

This research aims to derive users' opinions on which principles of the virtual assistant help prepare for quitting smoking. It results in the main question for the research: Which (ethical) principles do users think should be used by the virtual coach to decide when a human coach should give feedback to people who are preparing to quit smoking/vaping? To answer this research question the following sub-questions will be considered:

- Which ethical principles do users who are preparing to quit smoking/vaping find the most valuable?
- What could be the underlying reasons for principles preferences of people who are preparing to quit smoking/vaping by virtual assistants?

To identify the list of preferences, the thematic analysis [5] was performed. The base for the theme derivation was the data from the *Albers and Brinkman et al.* study [3]. Then they were triangulated [6] on the literature and the user's preferences.

Our study will contribute to the design of more effective, fair, and responsible virtual coaching systems by integrating user-preferred ethical considerations, potentially leading to higher success rates in smoking cessation. Additionally, the insights derived can inform the broader application of AI in healthcare, particularly in areas requiring human-AI collaboration for personalized treatment strategies.

## 2 Methodology

### 2.1 Data Gathering

Out of all the participants who took part in a study by Albers and Brinkman, 459 people received human feedback after at least one of the five sessions.

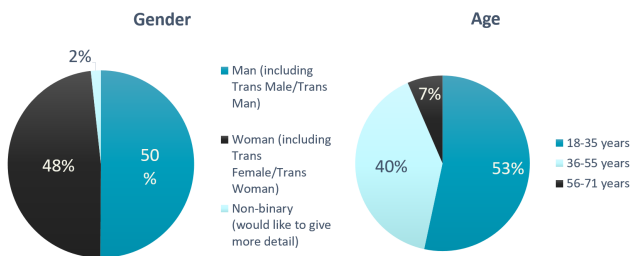


Figure 1: The pie charts represent the groups of participants.

During the process, a virtual coach named Kai provided help in preparing for smoking cessation [2]. After all the sessions were finished, the participants filled in the post-questionnaire. Among other questions, the participants were asked the following open question: "When a human coach cannot give feedback to everybody after each session due to time constraints, which principles/rules do you think the virtual coach should follow to decide when a human coach should give feedback to people who are preparing to quit smoking/vaping?".

Another crucial question in the post-questionnaire was a closed question. The question asked participants to assign 100 points across 11 principles a virtual coach could follow to allocate human feedback properly. The principles are derived from the *Persad et al.* paper [24].

The data gathered from the responses was then used in the thematic analysis. The thematic analysis used in this research is a variation of a *Braun and Clarke et al.* [5] supplemented by an idea of using ChatGPT-generated responses for gathering initial ideas for the thematic analysis. To ensure the reliability and quality of the results, two types of triangulation were applied: researcher [15] and method [6] triangulation. The former involved another researcher in applying the coding scheme derived from the first stages of the thematic analysis. The latter - method triangulation - supported the claims

and results of the later stages of the thematic analysis, which involved free-text responses, quantitative data, and literature. This approach allows to extract the essential information from the free-text responses, that would reflect the users' thoughts.

### 2.2 Thematic Analysis

The following four steps were undertaken.

**Stage 1: Familiarizing with the Data and Generating Initial Codes.** To gather the ideas for the initial codes an inductive approach was used in two different ways. The first is the one described by *Braun and Clarke et al.* [5] suggests scanning through the answers and noting down the ideas of codes. Another method we used was to generate artificial answers to the same question asking it to ChatGPT [22]. After deducting codes from 60 generated answers, a second coder was trained on the first 10 responses and then he coded the rest. The initial codes were formed by analyzing and discussing the results with the second coder. Such an approach was used since the discovery of ethical principles in the responses was not a trivial task. Gathering information from and training the second coder based on actual data would shrink the testable sample size. Generating synthetic data to train on allowed us to keep as many responses for the actual coding as possible.

**Stage 2: Coding of answers.** The coding of the main dataset is similar to the coding of synthetic data. After completing the coding, the first coder provided the dataset, codes, and descriptions of the latter to the second coder. First, the training set of 22 responses was coded together, serving as a guideline. The set consisted of specifically chosen answers by the following criterion: the answer is either a clear representative of a label or, the other way around, it is too vague to identify the label. Again, the identification of principles inside of a free-text response was a very subjective matter, therefore having a clear example of what is a representation of a label was crucial to reach an agreement on descriptions of the labels. Therefore, a specific set of responses was used to achieve that synchronization. After reaching an agreement on the training set, the second coder finished the rest of the dataset.

**Stage 3: Code reliability verification** The results of both coders were compared to identify if the coding scheme was reliable enough for the research. Cohen Kappa's and Brennan-Prediger coefficients [14] were calculated to achieve that.

First, to indicate the overall reliability of the scheme, the average Cohen Kappa coefficient was calculated for the whole scheme, resulting in 0.68. The codes had the same value for the research. Therefore, for the calculations, a non-weighted coefficient was used. The resulting number confirms reaching a substantial agreement [14]. This was also supported by the Brennan-Prediger coefficient equal to 0.69. To further justify the use of each code, Cohen Kappa was calculated per each code. The results can be found in Appendix A.

The coding scheme was changed after individual Cohen Kappa coefficients were calculated. To keep substantial

agreement along the whole coding scheme, codes that had individual coefficients lower than 0.60 were reconsidered. While reconsidering, the following features were taken into account: frequency of use and added value to answering the research question. Additionally, the code *Dissatisfaction with the Program* was removed due to the low frequency of use (less than 1 percent) and low impact on the research results. The results of the changes can be found in Appendix B. After reaching a substantial agreement, the resulting codes were again analyzed and discussed with the second coder. During the discussion, we agreed that the coding scheme is representing the dataset to full extent.

**Stage 4: Searching and Deriving the Themes** This last step involves a cyclic process of understanding and evaluating the themes. The process started with forming a mindmap that includes all the codes on the first iteration. Then some codes were grouped resulting in more isolated and stand-alone themes. The steps of the process included a discussion with the second coder. The discussion included such topics as representation of the names of the themes and the way the themes were grouped. It allowed for a more regulated result. After a few iterations, the final themes were formed.

### 2.3 Triangulation

Data triangulation is a process of comparing and looking for a correlation between the results of an experiment and the other sources of information [6]. The results of the thematic analysis of this research were compared to two sources: the literature (qualitative analysis) and data from sessions and post-questionnaire (quantitative analysis).

**Literature.** To successfully prepare for quantitative analysis, the thematic analysis results were first triangulated on the literature. After finding sufficient support for the results, further reading provided ideas and inspiration that formed a base for the quantitative analysis.

**Quantitative analysis.** The final part of the research was to support the resulting themes and literature triangulation. The findings of the thematic analysis were triangulated on the responses to the *closed question*.

The principles presented in the *closed question* were split into several groups. Each group then represented one of the themes derived from the thematic analysis. For instance, such principles as *Preferring those who have received the least amount of feedback so far* and *Preferring those who have spent the longest time without feedback* were grouped together representing the **Equal treatment** theme, as it has the same nature. The full list of principles and grouping of it can be found in the Appendix C.

To support the findings of the thematic analysis, the correlation coefficients between the themes and point allocation were calculated. For further calculation, the data was pre-processed by stating for each response for each theme whether the answer related to that theme (1 - "Yes", 0 - "No"). For this, the Point-Biserial [17] coefficient was used according to *Dischoso and Cabautan et al.* guideline [9].

The last stage of the research included the analysis of the data gained during the sessions and the post-questionnaire in turn to find informative correlations with the results of

the thematic analysis. For each of the potential correlations Point-Biserial coefficient was calculated, again consulting *Dischoso and Cabautan et al.* guideline [9].

## 3 Results

### 3.1 Themes

The performing of the thematic analysis resulted in deriving four themes. The resulting themes are the following: **Struggling the Most**, **Equal Treatment**, **Appreciating the Most**, and **Increasing Chances of Success the Most**.

The theme **Struggling the Most** covered all the responses where the participants indicated that human feedback should be received by those who struggle the most in different ways. Although *Those who struggle the most* was a self-contained code, during the theme analysis it was broadened by first splitting it into two subthemes: **Participant showing the most difficulty based on the indirect indicators** and **Participant showing the most difficulty based on direct indicators**. The former group consisted of ideas of tracking the participant's performance and noticing a decline or absence of progress, e.g. *"If it detects there's trouble keeping on with the sessions..."* (P652) and *People who are struggling to complete their tasks* (P480). The other group consists of answers indicating a preference for prioritizing those who struggle judging on the direct indicators, such as lack of motivation - *"Human coach should intervene when the participant is not feeling motivated."* (P890) and *"If the person trying to quit needs more motivation"* (P348); or noticing red flags - *"Maybe it should detect some words that can demonstrate a greater need for human interaction/feedback..."* (P663) and *"key words that may be controversial"* (P227).

The next theme detected by the thematic analysis was **Equal Treatment**. This theme is corresponding to the code with the same name. Apart from responses indicating exactly equality, such as *"Principles of equity and Equality"* (P901), this theme also included principles that behave similarly. For example, *"Randomly to make it fair I suppose"* (P22) and *"choose randomly"* (P751) for randomness, scheduled approach as in *"At least once a week per person"* (P222) and *"After every session"* (P930), and prioritizing those who have not received feedback the longest as can be derived from *"as soon as possible"* (P270).

The responses that expressed the willingness of the participants to receive human feedback were included in **Appreciating the Most** theme. Again, this theme is a direct derivation from the code *Those who would appreciate the most*. This preference was expressed directly - *"Use these questionnaires as guides as to who will appreciate the feedback, and then concentrate on those."* (P765); or having the same meaning, as in *"It should be someone who asks to"* (P950).

The final theme - **Increasing Chances of Success the Most** - consisted of the responses that propose that human

feedback should be provided to those participants, whose chance of successfully quitting smoking would rise the most if given the human feedback. The answers marked with this code provided a spectrum of ideas on how to indicate a successful participant. Those include highly-motivated people, as indicated in "It should be offered to those who really want to quit." (P385) and "Give feedback to those who really seek it from human coach" (P1082). Also, participants with great progress during the program - "Whoever seems to be engaging in the project the most." (P609). Finally, some answers also directly proposed this variant, as seen in "Who would be most likely to benefit from it should be prioritised" (P616).

The mind map of themes with corresponding usage percentages can be found in Figure 2:

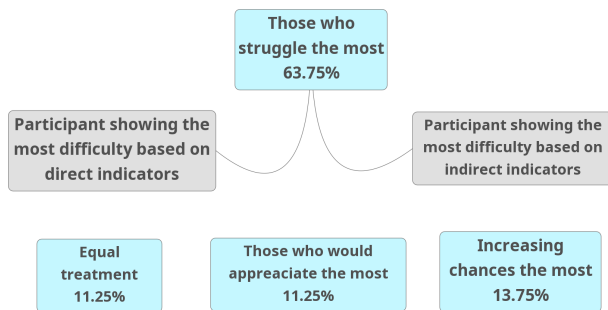


Figure 2: Overview of the themes with corresponding usage percentages (themes in blue and subthemes in gray).

### 3.2 Triangulation

The main goal of this research was to create a preference list by evaluating each theme to determine if the principle could be the main priority.

**Struggling the Most.** Prioritizing those who struggle the most is an ethical principle very commonly discussed in medicine [1][20][24]. *Mckie et al.* [20] examined the preferences of the public and professionals on the importance of the four principles, one of which was the severity of the condition. The results of their study have shown, that 51 percent of the participants preferred this principle over the others, making it the most important principle overall. The same applies to the results of this study: the participants in the vast majority - 63.75 % - preferred the principle of *those who struggle the most* over the others.

The literature also points out the cases when those who struggle the most are not the highest priority [1]. *Abma et al.* study of the triage guidelines indicated, that similarly to three other countries, the Dutch Intensive Care Unit (ICU) gives this principle the least priority. Naming it "Last resort" strategy, the author mentions that this principle is either not accepted at all, e.g. in Switzerland, or, when it comes down to those who suffer the most, the other principles are taken into account, e.g. in The Netherlands and Canada.

**Increasing Chances of Success the Most.** *Persad et al.* [24] give two formulations on how to allocate the help consulting **maximising total benefits** principle: **number of lives saved** and **prognosis or life-years saved**. *Abma et al.* [1] show in their study, that increasing the number of saved lives by allocating help to those whose chances increase the most when given priority, is the main preference across all the studied country guidelines on ICU triage.

In the case of our study, one of the increasing chances strategies was to prioritize the most motivated and engaged participants. *McCuller et al.* [19] conducted a study on how motivation affects success rates at smoking cessation. The results show, that motivated people are 26% more likely to quit smoking.

Although the literature has different opinions about *increasing chances the most* principle when it comes to smoking cessation - this strategy tends to be effective. Our study shows, that 13.75 percent of the participants prefer such a principle over the others.

**Equal Treatment.** Regarding the principle of equality, *Persad et al.* [24] suggested two applications: lottery and first-come-first-served. On top of that, the participants in their responses usually mentioned a scheduled approach. When it comes to the ICU guidelines, medics suggest that this approach should not be used, or used when no other principle is applicable [1].

When it comes to people's preferences, *Mckie et al.* [20] concluded that a significant amount (31%) prefer the equality principle. Furthermore, in *Nord et al.* [21] equal priority was preferred by the majority (45%) of the participants.

The results of our research show the same tendency as in *Abma et al.* [1] studies, contradicting the findings of *Nord et al.* and *Mckie et al.* Only 11.25% of the participants mentioned the *equal treatment* principle as the most preferred one.

**Appreciating the Most.** *Ende et al.* [10] concluded that, in general, people prefer not to be the decision-makers for their own treatments. On a scale from 0 to 100, where 0 indicates a very low and 100 indicates a very high preference for decision-making, the mean resulting score was 33.2. The same studies show, that among the people who prefer to decide on their treatments, 75% would do that when suffering from a minor illness.

*Soulakova et al.* [27] researched how reliance on other people affected the smoking cessation process. The authors concluded, that smokers who relied on social support from friends and family had a lot higher chances of quitting smoking. For our study, it means, that it is important to provide human feedback to those, who rely on it.

Therefore, the principle of giving human feedback to those, who would appreciate it and want to get it, is viable. The results of the research have shown, that 11.25% of participants would prefer if the help would go to those who *appreciating it the most* first.

Theme	Point-Biserial Coef.	P-value
Struggling the most	.282	p < .001
Equal treatment	.321	p < .001
Appreciating the most	.352	p < .001
Increasing chances of success the most	.146	0.009

Table 1: The table of correlations, expressed in Point-Biserial coefficients, between the themes and closed question allocation groups.

### 3.3 Quantitative Findings.

**Triangulation on closed question.** The results of calculating the Point-Biserial [17] coefficient for the themes and point allocation can be found in Table 1. The table only includes the correlation between the theme and the corresponding group of principles. The complete table with correlation identification between all the themes and groups of principles can be found in Appendix D. The low p-value (less than 0.05) indicates a significant result. According to guideline [9] the coefficient of **Equal Treatment** (0.32) and **Appreciating the Most** (0.35) themes indicate a *reasonably good* correlation; the coefficient of **Struggling the Most** (0.28) theme means *acceptable* correlation. Finally, **Increasing Chances of Success the Most** (0.15) coefficient indicates poor correlation. However, according to *Boyd et al.* [28], social sciences and psychology often consider even such small numbers to be valuable results, considering the low p-value ratios.

Good correlation scores indicate a connection between the open-question responses, where the participants would only mention the most valuable principle for them, and their point allocations, where the preferences were defined more precisely. Therefore, the results of this triangulation support the results of the thematic analysis.

### 3.4 Correlations.

The further investigation of data revealed two points of interest: the influence of the willingness to drop out of the experiment and the willingness to receive human feedback.

**Drop out and themes correlation.** For this correlation responses to the following pre-screening question were analyzed: *"Currently you are taking part in a paid experiment. Imagine this was an unpaid smoking cessation program. How likely would you then have quit the program or returned to this session?"* [3]. The answers were ratings on a scale from -5 ("definitely would quit a program") to 5 ("definitely would have returned to the sessions"); since the question was asked during each session, the mean of the responses was considered. The hypothesis for the correlation was that participants, who tend to want to quit the program, and therefore are not motivated to continue, would be less willing to opt for equal distribution of human feedback, as it would then mean willing to receive it and continue the sessions.

To prove the presence or absence of the correlation, the Point-Biserial coefficient was calculated, following *Dischoso and Cabauatan et al.* guideline [9]. The resulting numbers can be found in the Table 2.

Unfortunately, low Point-Biserial coefficients and high p-values indicate the absence of support for the hypothesis.

Theme	Point-Biserial Coef.	P-value
Struggling the most	.003	0.964
Equal treatment	.052	0.356
Appreciating the most	-.020	0.717
Increasing chances of success the most	-.032	0.571

Table 2: The Point-Biserial correlation coefficient between the drop-out question responses and themes with corresponding p-values.

Theme	Point-Biserial Coef.	P-value
Struggling the most	-.023	.687
Equal treatment	.136	.015
Appreciating the most	-.164	.003
Increasing chances of success the most	.059	.297

Table 3: The Point-Biserial correlation coefficient between the willingness to receive human feedback question responses and themes with corresponding p-values.

**Willingness to receive human feedback and themes correlation.** This correlation was based on the data gained from responses to the following question: *"How would you view receiving a feedback message from a human coach after this session?"* [3]. The responses were rated on a Likert scale from -10 ("very negatively") to 10 ("very positively"). Again, the mean of the ratings throughout the session was taken into consideration. The correlation was used against the following hypothesis: "Participants who are less willing to receive human feedback would opt to **Appreciating the Most** theme since it would express their feeling about receiving one."

The Point-Biserial coefficient was calculated. The results are presented in Table 3.

The resulting numbers indicated two conclusions. First, the low p-value (less than 0.05) and sufficiently high, according to *Boyd et al.* [28], negative correlation coefficient of the **Appreciating the most** theme concludes that the hypothesis holds, and there is a negative correlation between the willingness to receive human feedback and prefer allocating the human feedback to those who appreciate it the most.

Another conclusion that could be made was the positive correlation between the willingness and **Equal Treatment** theme. Therefore, when people indicated a willingness to receive human feedback, they would more likely prefer that it would be allocated equally.

## 4 Responsible Research

This research adheres to responsible research ideas by using different analysis methods and reflecting on researcher biases. The contributions of this research extend beyond smoking cessation, offering valuable insights into the ethical integration of AI in health care, ultimately aiming to improve user outcomes and experiences.

### 4.1 Informed Consent, Confidentiality and Anonymity

The data used in the research was collected during *Albers and Brinkman et al.* study [3]. It received ethical approval from the TU Delft Human Research Ethics Committee (HREC). Data was stored securely in a cloud drive, and access was restricted to authorized researchers only. After collection, the

data was anonymized and trimmed to exclude the risk of re-identification. Only protected data was accessible by the research team and was not shared further. The responses were not altered or modified to avoid data falsification. P-hacking was avoided by presenting the initial hypotheses.

## 4.2 Thematic Analysis and Triangulation

The thematic analysis was conducted to ensure the reliability and validity of the findings. To generate initial codes, an inductive approach was used. The codes were then tuned through a collaborative process involving multiple researchers. To ensure the validity of the results, method triangulation was used, comparing qualitative findings with quantitative data and literature. This process gave an understanding of the participants' preferences for ethical principles.

## 4.3 Synthetic Data Generation

The artificial data, generated by ChatGPT, was only used to enrich and broaden thematic analysis. ChatGPT provided a set of artificial answers to supplement the initial coding process. This approach assisted in identifying initial codes that might not have been immediately apparent from the participant answers data alone.

## 4.4 Contribution to Smoking Cessation Strategies

This research contributes to the development of more effective smoking cessation strategies. The identification of the ethical principles that users prefer in receiving human feedback, and virtual coaching systems can be suited to better meet the needs of the participants. Understanding these preferences allows for a more personalized and supportive smoking cessation method, increasing the success rates of quitting smoking. The conclusions gained from this research can help with designing virtual coaches, making them more responsive to the needs of users.

# 5 Discussion and Future work

## 5.1 Main results and Recommendations.

This study has identified and analyzed users' preferences on ethical principles for when a virtual coach should allocate human feedback during smoking cessation programs. The key findings are:

- **Struggling the Most:** 63.75% of participants prioritize support for those who experience the greatest difficulty in quitting smoking.
- **Increasing Chances of Success the Most:** 13.75% favor allocating feedback to those with the highest potential for succeeding in smoking cessation.
- **Appreciating the Most:** 11.25% prefer feedback to be given to those who would most value it.
- **Equal Treatment:** 11.25% of participants value fairness in the distribution of feedback.

Based on the results, we provide the recommendations for developers of such virtual assistant systems:

- **Prioritize struggling users:** Develop virtual coaching systems that identify and prioritize users, who show significant difficulty in quitting smoking. This could include monitoring indicators such as the declining performance of a user over the program or lack of progress.
- **Leverage motivation and progress indicators:** Use motivational and progress indicators to allocate feedback to those with the highest potential for success, such as highly motivated users or those showing significant engagement.
- **User-centric design:** Ensure that user feedback on the topic of willingness to receive human feedback is being collected regularly, to keep the system aligned with user needs and preferences.
- **Incorporate fairness:** When all the other principles fail to allocate the feedback, ensure equitable distribution of feedback. Consider randomization, scheduled feedback, and prioritizing those who have not received feedback recently.

## 5.2 Limitations

**Data.** The dataset used in this research was not gathered specifically for the topic of the study. This may have resulted in an incomplete representation of the smoking cessation population's opinion. The thematic analysis was conducted based on the answers to an open-form question. As a result, some participants did not understand the question or could not come up with an opinion, trimming the amount of useful data. Additionally, some participants might have presented themselves in a more favorable light or may not accurately recall their experiences during the smoking cessation process.

**Thematic analysis.** When coding, only the responses were considered to form an opinion. However, using data from the closed question points allocation might have provided a more reliable coding scheme. Additionally, every response was coded with only one, most suitable code.

However, in later stages, the data from the thematic analysis was found to be correlating with closed data, supporting the found themes. To ensure the reliability of the coding, the double-coder method and calculated agreement coefficients were used.

**Quantitative analysis.** To determine the relationship between themes and participants' preferences Point-Biserial correlation coefficients, that were used, have their limitations. These coefficients may not capture complex or nonlinear relationships, potentially oversimplifying the dynamics between variables.

The final sample size was modest, resulting in limited statistical power to detect small but meaningful effects. Potentially, it could lead to a failure to identify significant relationships, leading to potential type II errors.

## 5.3 Future work

**Correlations with Other Principles in Closed Question.** To explore complex interactions between multiple principles, a broader array of statistical methods, such as machine learning or logistic regression, could bring more value to the research.



This would provide a more detailed picture of how different ethical considerations are weighted by the participants.

Additionally, there were literature sources that mentioned that the age factor plays a crucial role in aid allocation preferences by the users [12] [7]. This research does not study *youngest first* principle, due to the lack of participants who mentioned that principle.

**Systems of principle usage.** *Persad et al.* [24] conclude, that instead of using a single principle, a system of principles exploitation is more effective. Revealing more principle preferences would allow us to then form different combinations of those, potentially increasing the benefits.

**Testing the application of the principles.** To further implicate the efficacy of the results, those should be applied in the real-world setting, helping validate findings. Controlled trials could be useful for understanding the impact of these principles on smoking cessation success rates.

## 6 Conclusions

This research aimed to find and analyze the users' ethical principles and preferences in deciding when a virtual coach should allocate human feedback in the context of preparing for smoking cessation. The study's data consisted of users' responses, supplemented by artificially generated data for initial ideas. By employing a thematic analysis and triangulating findings with quantitative data and literature, four main themes were identified: **Struggling the Most**, **Equal Treatment**, **Appreciating the Most**, and **Increasing Chances of Success the Most**.

The findings revealed that the majority of participants (63.75%) prioritize **Struggling the Most** principle, aligning with the ethical considerations commonly discussed in the medical literature. Although less preferred, **Equal Treatment** (11.25%) and **Appreciating the Most** (11.25%) also emerged as significant themes. The **Increasing Chances of Success the Most** (13.75%) principle is valid according to both thematic analysis and the literature, where several sources indicated such principle to be the most valuable.

The results underline the importance of tailoring virtual coaching systems to incorporate ethical principles that users value. **Struggling the Most** prioritization will allow virtual coaches to provide more targeted support, potentially increasing the overall success rates in smoking cessation. Additionally, the findings of the study offer valuable insights into the broader application of AI in healthcare and e-health, emphasizing the need for human-AI collaboration.

Future research should focus on widening the list of principles, gathering more information about such themes as age factor, exploring complex interactions between principles, and testing the real-world application of these, validating their efficacy. As the literature suggests, the integration of a combination of ethical principles could further enhance the effectiveness of virtual coaching systems in supporting individuals through their smoking cessation programs.

## 7 Acknowledgments

This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders orga-

nized 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.

Additional special thanks to the supervisors: Ir. Nele Albers and Dr. Ir. Willem-Paul Brinkman, and our team of researchers.

## References

- [1] I L Abma, G J Olthuis, and A J M Oerlemans. "Comparing the Dutch ICU triage guidelines to guidelines from abroad: differences and similarities". en. In: 29.6 (2021).
- [2] Nele Albers and Willem-Paul Brinkman. "Perfect Fit - Learning to propose useful preparatory activities for quitting smoking and becoming more physically active". en-us. In: (July 2023). Publisher: OSF. DOI: 10.17605/OSF.IO/NUY4W. URL: <https://osf.io/nuy4w> (visited on 05/05/2024).
- [3] Nele Albers and Willem-Paul Brinkman. "Perfect Fit - Learning when to involve a human coach in an eHealth application for preparing for quitting smoking or vaping". en-us. In: (Jan. 2024). Publisher: OSF. DOI: 10.17605/OSF.IO/78CNR. URL: <https://osf.io/78cnr> (visited on 06/03/2024).
- [4] George I. Balch. "Exploring Perceptions of Smoking Cessation among High School Smokers: Input and Feedback from Focus Groups". en. In: *Preventive Medicine* 27.5 (Sept. 1998), A55–A63. ISSN: 00917435. DOI: 10.1006/pmed.1998.0382. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0091743598903826> (visited on 06/11/2024).
- [5] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". en. In: *Qualitative Research in Psychology* 3.2 (Jan. 2006), pp. 77–101. ISSN: 1478-0887, 1478-0895. DOI: 10.1191/1478088706qp063oa. URL: <http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa> (visited on 04/27/2024).
- [6] Nancy Carter et al. "The Use of Triangulation in Qualitative Research". en. In: *Oncology Nursing Forum* 41.5 (Sept. 2014), pp. 545–547. ISSN: 0190-535X, 1538-0688. DOI: 10.1188/14.ONF.545-547. URL: <http://onf.ons.org/onf/41/5/use-triangulation-qualitative-research> (visited on 04/28/2024).
- [7] Paulo Antônio De Carvalho Fortes and Patricia Cristina Andrade Pereira. "Patient prioritization in medical emergencies: an ethical analysis". en. In: *Revista da Associação Médica Brasileira (English Edition)* 58.3 (May 2012), pp. 335–340. ISSN: 22554823. DOI: 10.1016/S2255-4823(12)70204-7. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2255482312702047> (visited on 06/01/2024).



- [8] Kerry Dechant. "Breaking the habit - nicotine replacement makes positive moves". en. In: *Inpharma Weekly* 856.1 (Sept. 1992). Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 1 Publisher: Springer International Publishing, pp. 7–9. ISSN: 2230-6056. DOI: 10.2165/00128413-199208560-00007. URL: <https://link.springer.com/article/10.2165/00128413-199208560-00007> (visited on 05/05/2024).
- [9] Anamarie A Dichoso and Rosmina Joy M Cabauatan. "Test Item Analyzer Using Point-Biserial Correlation And P-Values". en. In: 9.04 (2020).
- [10] Jack Ende et al. "Measuring patients' desire for autonomy: Decision making and information-seeking preferences among medical patients". en. In: *Journal of General Internal Medicine* 4.1 (Jan. 1989), pp. 23–30. ISSN: 0884-8734, 1525-1497. DOI: 10.1007/BF02596485. URL: <http://link.springer.com/10.1007/BF02596485> (visited on 06/05/2024).
- [11] Joseph Finkelstein and Jeffrey Wood. "Interactive mobile system for smoking cessation". en. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka: IEEE, July 2013, pp. 1169–1172. ISBN: 978-1-4577-0216-7. DOI: 10.1109/EMBC.2013.6609714. URL: <http://ieeexplore.ieee.org/document/6609714/> (visited on 06/10/2024).
- [12] Adrian Furnham, Kathryn Thomson, and Alastair McClelland. "The allocation of scarce medical resources across medical conditions". en. In: *Psychology and Psychotherapy: Theory, Research and Practice* 75.2 (June 2002), pp. 189–203. ISSN: 1476-0835, 2044-8341. DOI: 10.1348/147608302169643. URL: <https://bpspsychub.onlinelibrary.wiley.com/doi/10.1348/147608302169643> (visited on 06/11/2024).
- [13] Dmitry Gromov, Ingo Bulla, and Ethan O. Romero-Severson. "Optimal Resource Allocation for HIV Prevention and Control". en. In: *Trends in Biomathematics: Modeling, Optimization and Computational Problems*. Ed. by Rubem P. Mondaini. Cham: Springer International Publishing, 2018, pp. 121–137. ISBN: 978-3-319-91091-8 978-3-319-91092-5. DOI: 10.1007/978-3-319-91092-5\_9. URL: [http://link.springer.com/10.1007/978-3-319-91092-5\\_9](http://link.springer.com/10.1007/978-3-319-91092-5_9) (visited on 06/11/2024).
- [14] Kilem Gwet. "Handbook of inter-rater reliability". In: *Gaithersburg, MD: STATAxis Publishing Company* (2001), pp. 223–246.
- [15] Kilem Li Gwet and Kilem Li Gwet. *Analysis of categorical ratings*. en. Fifth edition. Handbook of inter-rater reliability / Kilem L. Gwet, Ph. D volume 1. Gaithersburg, MD: AgreeStat Analytics, 2021. ISBN: 978-1-79235-463-2.
- [16] Nicole Hassoun, Emir Malikov, and Nathan Lubchenco. "How people think about distributing aid". en. In: *Philosophical Psychology* 29.7 (Oct. 2016), pp. 1029–1044. ISSN: 0951-5089, 1465-394X. DOI: 10.1080/09515089.2016.1211259. URL: <https://www.tandfonline.com/doi/full/10.1080/09515089.2016.1211259> (visited on 06/11/2024).
- [17] Diana Kornbrot. "Point Biserial Correlation". en. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014. ISBN: 978-1-118-44511-2. (Visited on 06/09/2024).
- [18] Pius Krütli et al. "How to Fairly Allocate Scarce Medical Resources: Ethical Argumentation under Scrutiny by Health Professionals and Lay People". en. In: *PLOS ONE* 11.7 (July 2016). Ed. by Yoel Lubell, e0159086. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0159086. URL: <https://dx.plos.org/10.1371/journal.pone.0159086> (visited on 06/11/2024).
- [19] William Jason McCuller et al. "Motivation to quit as a mediator of tobacco cessation among at-risk youth". en. In: *Addictive Behaviors* 31.5 (May 2006), pp. 880–888. ISSN: 03064603. DOI: 10.1016/j.addbeh.2005.07.019. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0306460305001991> (visited on 06/05/2024).
- [20] John McKie and Jeff Richardson. "Social preferences for prioritizing the treatment of severely ill patients: The relevance of severity, expected benefit, past health and lifetime health". en. In: *Health Policy* 121.8 (Aug. 2017), pp. 913–922. ISSN: 01688510. DOI: 10.1016/j.healthpol.2017.05.010. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168851017301483> (visited on 06/01/2024).
- [21] Erik Nord. *Cost-Value Analysis in Health Care: Making Sense Out of QALYS*. en. Google-Books-ID: iEe0IdcYoWoC. Cambridge University Press, Sept. 1999. ISBN: 978-0-521-64434-1.
- [22] OpenAI. *ChatGPT (GPT-4)*. Accessed: 2024-06-03. 2024. URL: <https://www.openai.com>.
- [23] Jonathan Passmore and David Tee. "Can Chatbots replace human coaches? Issues and dilemmas for the coaching profession, coaching clients and for organisations". en. In: *The Coaching Psychologist* 19.1 (June 2023), pp. 47–54. ISSN: 1748-1104, 2396-9628. DOI: 10.53841/bpstcp.2023.19.1.47. URL: <https://explore.bps.org.uk/lookup/doi/10.53841/bpstcp.2023.19.1.47> (visited on 06/03/2024).
- [24] Govind Persad, Alan Wertheimer, and Ezekiel J Emanuel. "Principles for allocation of scarce medical interventions". en. In: *The Lancet* 373.9661 (Jan. 2009), pp. 423–431. ISSN: 01406736. DOI: 10.1016/S0140-6736(09)60137-9. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0140673609601379> (visited on 04/30/2024).
- [25] Simon Provoost et al. "Embodied Conversational Agents in Clinical Psychology: A Scoping Review". EN. In: *Journal of Medical Internet Research* 19.5 (May 2017). Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada,

e6553. DOI: 10.2196/jmir.6553. URL: <https://www.jmir.org/2017/5/e151> (visited on 05/21/2024).

- [26] Andrea L. Smith et al. “The Views and Experiences of Smokers Who Quit Smoking Unassisted. A Systematic Review of the Qualitative Evidence”. en. In: *PLOS ONE* 10.5 (May 2015). Ed. by Ruth Jepson, e0127144. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0127144. URL: <https://dx.plos.org/10.1371/journal.pone.0127144> (visited on 06/10/2024).
- [27] Julia N. Soulakova et al. “Motivational Benefits of Social Support and Behavioural Interventions for Smoking Cessation”. en. In: *Journal of Smoking Cessation* 13.4 (Dec. 2018), pp. 216–226. ISSN: 1834-2612. DOI: 10.1017/jsc.2017.26. URL: [https://www.cambridge.org/core/product/identifier/S1834261217000263/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1834261217000263/type/journal_article) (visited on 06/01/2024).
- [28] Boyd Spencer. “Correlations, Sample Size, and Practical Significance: A Comparison of Selected Psychological and Medical Investigations”. en. In: *The Journal of Psychology* 129.4 (July 1995), pp. 469–475. ISSN: 0022-3980, 1940-1019. DOI: 10.1080/00223980.1995.9914982. URL: <http://www.tandfonline.com/doi/abs/10.1080/00223980.1995.9914982> (visited on 06/09/2024).
- [29] *Tobacco*. URL: <https://www.who.int/news-room/factsheets/detail/tobacco> (visited on 04/27/2024).

## A Cohen Kappa coefficients per Code

Codes	Cohen Kappa
Those who struggle the most	.641
Increasing chances of success	.815
Those who are lacking motivation	.630
Those who would appreciate the most	.775
Equal treatment	.831
Performance decline across the preparation sessions	.652
Red flags detected	.800
Insufficient virtual coach support for specific topics	1
Data integrity concerns	.333
Dissatisfaction with the program	.667
Periods of increased temptation	.333

Table 4: Cohen Kappa coefficients for codes.

## B The finalized list of Codes

Codes	Cohen Kappa
Those who struggle the most	.641
Increasing chances of success	.815
Those who are lacking motivation	.630
Those who would appreciate the most	.775
Equal treatment	.831
Performance decline across the preparation sessions	.652
Red flags detected	.760
Insufficient virtual coach support for specific topics	1

Table 5: The resulting codes after reconsidering the coding scheme.

## C Grouping of the principles according to the Themes

Theme	Principle
Struggling the most	Least likely to success without the human feedback
	Most likely to experience negative consequences without human feedback
Equal treatment	Allocating randomly
	Least amount of feedback so far
	Longest time without human feedback
Appreciating the most	Would appreciate the human feedback the most
Increasing chances of success the most	Increasing chances of success if given human feedback
	Largest reduction in negative consequences if given human feedback

Table 6: Grouping of the closed question principles used to align with the themes.

## D Full version of Point-Biserial calculations

Theme	Group of principles	Point-Biserial Coef.	P-value
Struggling the most	Struggling the most	.282	p < .001
	Equal treatment	-.121	.031
	Appreciating the most	-.212	p < .001
	Increasing chances of success the most	-.008	.881
Equal treatment	Struggling the most	-.099	0.078
	Equal treatment	.321	p < .001
	Appreciating the most	-.095	.090
	Increasing chances of success the most	-.172	.002
Appreciating the most	Struggling the most	-.183	0.001
	Equal treatment	-.026	.642
	Appreciating the most	.352	p < .001
	Increasing chances of success the most	.023	.678
Increasing chances of success the most	Struggling the most	-.135	.016
	Equal treatment	-.099	.077
	Appreciating the most	.058	.300
	Increasing chances of success the most	.146	.009

Table 7: The complete table of correlations, expressed in Point-Biserial coefficients, between the themes and closed question allocation groups.