

Conversational Chatbot for Cigarette Smoking Cessation: Report of the User-Centered Design Eleven Step Development Process

Jonathan Bricker, Brianna Sullivan, Kristin Mull, Margarita Santiago-Torres, Juan Lavista Ferres

Submitted to: JMIR mHealth and uHealth
on: February 12, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 56

 Multimedia Appendixes 57

 Multimedia Appendix 0..... 57

CONSORT (or other) checklists..... 58

 CONSORT (or other) checklist 0..... 58

Conversational Chatbot for Cigarette Smoking Cessation: Report of the User-Centered Design Eleven Step Development Process

Jonathan Bricker^{1,2} PhD; Brianna Sullivan¹ MS; Kristin Mull¹ MS; Margarita Santiago-Torres¹ PhD; Juan Lavista Ferres³ PhD

¹Fred Hutch Cancer Center Division of Public Health Sciences Seattle US

²University of Washington Department of Psychology Seattle US

³Microsoft Corporation AI for Good Lab Redmond US

Corresponding Author:

Jonathan Bricker PhD
Fred Hutch Cancer Center
Division of Public Health Sciences
1100 Fairview Avenue N
Seattle
US

Abstract

Background: Conversational chatbots are an emerging digital intervention for smoking cessation. No studies have reported on a user-centered design development process for creating cessation chatbots.

Objective: To describe the user-centered design development process for a novel quit smoking conversational chatbot called “QuitBot.”

Methods: The four years of formative research for developing QuitBot followed an eleven-step process: (1) specifying a conceptual model, (2) conducting content analysis of existing interventions, (3) assessing user needs, (4) developing the chat’s persona (“personality”), (5) prototyping content and persona, (6) developing full functionality, (7) programming the QuitBot, (8) conducting a diary study, (9) conducting a pilot randomized trial, (10) reviewing results of the trial, and (11) adding a free-form question and answer (QnA) function, based on user feedback from pilot trial

Results: A quit smoking program spanning 42 days of 2 to 3-minute conversations over two phases: (1) pre-quit phase (14 days) content covering topics ranging from motivations to quit, setting a quit date, choosing FDA-approved cessation medications; and (2) post-quit phase (28 days) content on coping with triggers and recovering from lapses/relapses. In a pilot randomized trial with 96% three-month outcome data retention, QuitBot demonstrated high user engagement and promising cessation rates compared to the National Cancer Institute’s SmokefreeTXT (SFT) text messaging program—particularly among those who viewed all 42 days of program content: 30-day complete-case, point prevalence abstinence (PPA) rates at three-month follow-up were 63% (39/62) for QuitBot vs. 38% (45/117) for SFT (OR = 2.58; 95% CI: 1.34, 4.99; P = .005). However, Facebook Messenger (FM) intermittently blocked participants’ access to QuitBot so we transitioned from FM to a standalone smartphone app as the communication channel. Participants’ frustration with QuitBot’s inability to answer their open-ended questions lead to us develop a core conversational feature enabling users to ask open-ended questions about quitting cigarette smoking and for the QuitBot to respond with accurate and professional answers. To support this functionality, we developed a library of 11,000 QnA pairs on topics associated with quitting cigarette smoking. Model testing results showed that Microsoft’s Azure-based QnA maker effectively handled questions that matched our library of 11,000 QnA pairs. A fine-tuned, contextualized GPT3.5 responds to questions that are not within our library of QnA pairs.

Conclusions: The development process yielded the first documented case of a comprehensive quit smoking program delivered as a conversational chatbot. Iterative testing led to significant enhancements, including improvements to the delivery channel. A pivotal addition was the inclusion of a core LLM-supported conversational feature allowing users to ask open-ended questions. Clinical Trial: ClinicalTrials.gov Identifier, NCT03585231

(JMIR Preprints 12/02/2024:57318)

DOI: <https://doi.org/10.2196/preprints.57318>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org/](#)

Original Manuscript

Conversational Chatbot for Cigarette Smoking Cessation: Report of the User-Centered Design Eleven Step Development Process

Jonathan B. Bricker,^{1,2} Brianna M. Sullivan,¹ Kristin E. Mull,¹ Margarita Santiago-Torres,¹ Juan Lavista-Ferres³

¹Fred Hutchinson Cancer Center, Division of Public Health Sciences, 1100 Fairview Avenue N., Seattle, Washington, 98109, USA.

²University of Washington, Department of Psychology, Box 351525, Seattle, Washington, 98195, USA.

³AI for Good Research Lab, Microsoft Corporation, Redmond, WA, 98052, USA

Word count: 4582

Running head: chatbot and smoking cessation

Declaration of competing interest: None of the authors have declarations.

Clinical trial registration: ClinicalTrials.gov Identifier, NCT03585231

Corresponding author

Jonathan B. Bricker, PhD

Fred Hutchinson Cancer Center

Division of Public Health Science, Cancer Prevention Program

1100 Fairview Avenue North, Seattle WA 98109, USA

Telephone: 206-667-5074

Fax: 206-667-5977

Email: jbricker@fredhutch.org

Preprint
JMIR Publications

Abstract (word count = 447)

Background: Conversational chatbots are an emerging digital intervention for smoking cessation. No studies have reported on the entire development process of a cessation chatbot.

Objective: To describe the user-centered design development process for a novel and comprehensive quit smoking conversational chatbot called “QuitBot.”

Methods: The four years of formative research for developing QuitBot followed an eleven-step process: (1) specifying a conceptual model, (2) conducting content analysis of existing interventions (63 hours of intervention transcripts), (3) assessing user needs, (4) developing the chat’s persona (“personality”), (5) prototyping content and persona, (6) developing full functionality, (7) programming the QuitBot, (8) conducting a diary study, (9) conducting a pilot randomized trial, (10) reviewing results of the trial, and (11) adding a free-form question and answer (QnA) function, based on user feedback from pilot trial results. The process of adding a QnA function itself involved a three-step process: (a) generating QnA pairs, (b) fine tuning Large Language Models (LLMs) on QnA pairs, and (c) evaluating the LLM model outputs.

Results: A quit smoking program spanning 42 days of 2 to 3-minute conversations covering topics ranging from motivations to quit, setting a quit date, choosing FDA-approved cessation medications, coping with triggers, and recovering from lapses/relapses. In a pilot randomized trial with 96% three-month outcome data retention, QuitBot demonstrated high user engagement and promising cessation rates compared to the National Cancer Institute’s SmokefreeTXT (SFT) text messaging program—particularly among those who viewed all 42 days of program content: 30-day complete-case, point prevalence abstinence (PPA) rates at three-month follow-up were 63% (39/62) for QuitBot vs. 38% (45/117) for SFT (OR = 2.58; 95% CI: 1.34, 4.99; $P = .005$). However, Facebook Messenger (FM) intermittently blocked participants’ access to QuitBot so we transitioned from FM to a standalone smartphone app as the communication channel. Participants’ frustration with QuitBot’s inability to answer their open-ended questions lead to us develop a core conversational feature enabling users to ask open-ended questions about quitting cigarette smoking and for the QuitBot to respond with accurate and professional answers. To support this functionality, we developed a library of 11,000 QnA pairs on topics associated with quitting cigarette smoking. Model testing results showed that Microsoft’s Azure-based QnA maker effectively handled questions that matched our library of 11,000 QnA pairs. A fine-tuned, contextualized GPT3.5 responds to questions that are not within our library of QnA pairs.

Conclusions: The development process yielded the first LLM-based quit smoking program delivered as a conversational chatbot. Iterative testing led to significant enhancements, including

improvements to the delivery channel. A pivotal addition was the inclusion of a core LLM-supported conversational feature allowing users to ask open-ended questions.



Introduction

Background

Cigarette smoking accounts for 8 million premature deaths and 25% of all cancer deaths annually [1, 2]. Despite advancements in government policies, anti-smoking campaigns, and shifting societal norms, existing smoking cessation interventions continue to have limited treatment engagement and cessation rates [3-9]. While this is a problem for the general population of people who smoke, the issue is particularly pronounced in marginalized communities, synonymous with vulnerable or disadvantaged groups, which are segments of society facing systemic disadvantages and barriers in accessing resources and opportunities. Marginalized populations, marked by factors such as racial or ethnic minority status, sexual or gender identity differences, low education and income levels, higher unemployment rates, and/or an increased prevalence of mental illness, encounter discrimination, social exclusion, and limited influence in decision-making processes.

Challenges in treatment engagement and cessation efficacy across all communities of people who smoke are compounded by a scarcity of trained clinicians and significant barriers, including cost and lack of insurance, hindering access to existing clinician-delivered interventions [10-14]. Given that 1.3 billion people in the world smoke cigarettes, with 70% of them wanting to quit, it would be impractical to have enough trained clinicians to help people quit smoking [15, 16]. Indeed, only 5% of cessation attempts are aided by a health professional [17]. Consequently, there is an enormous need for high-impact, cost-effective population-level interventions for smoking cessation.

A well-documented finding from research on clinician-delivered treatments has emphasized the significance of therapeutic conversations as powerful drivers of patient engagement [18-21]. Therapeutic conversations, which forms a social-emotional bond with the user, has predicted treatment engagement and, subsequently, health outcomes across various treatments and settings [12, 18, 22]. A new technology provides an opportunity to leverage engaging therapeutic conversations. Advances in machine learning, large language models (LLMs), and cloud computing are now making it possible to create and widely disseminate conversational chatbots for behavior change coaching.

Unlike the chatbots used in customer service contexts, conversational chatbots for behavior change coaching are designed to form long-term social-emotional connections with users—even as they are

made aware that chatbots are merely computer software that use language to communicate with users [23, 24]. Conversational chatbots for coaching are designed to be supportive and empathic, offering reflective listening, personalized responses, and timely advice aligned with the user's individual needs [25, 26]. In the context of cessation, conversational chatbots can enhance engagement through an informal therapeutic conversational style tailored to users' unique barriers to quitting smoking [27]. Conversational chatbots only require a text response to operate, making them ideal for all individuals who smoke, including those with low technology literacy.[26] Overall, conversational chatbots offer a cost-effective communication platform, accessible at any time, and have the potential for high population level reach, making them a valuable tool in smoking cessation interventions.

To date, research on conversational chatbots for smoking cessation is scarce. Existing literature revealed a limited number of empirical studies, often exhibiting low methodological quality [28]. There is a notable paucity of randomized controlled trials (RCTs) focusing on conversational chatbots for smoking cessation, and while promising results have emerged, they have been limited by low quit rates [29]. Several conversational chatbots for smoking cessation in the public domain include Florence [30], Bella [31], and Alex AI [32]. However, we are not aware of publications on their efficacy, with only the Florence app having reported user's receptivity results [33]. Critical to creating useful and engaging conversational chatbots is following a user-centered design development process [34]. Like most chatbots, the development of the chatbots listed above lacks context for how they were designed and any user-centered design that involved conducting a needs assessment or including user feedback during the development process [28, 35]. The few studies that have provided development details only describe early design phases, like coding 30 quit coaching calls for prototype development, without empirical efficacy data [28, 35, 36]. In sum, the literature on chatbots for smoking cessation offers only partial accounts on how they were developed or report on initial stages of development.

Objective

To address these gaps, this paper describes the comprehensive four-year, eleven-step user-centered design development process for a novel quit smoking conversational chatbot named "QuitBot". This single report aims to summarize the entirety of QuitBot development process.

Methods

Overview of Formative Research Process

The four years of formative research for developing QuitBot followed an eleven-step process, consistent with a user-centered design framework (**Figure 1**) [37, 38].

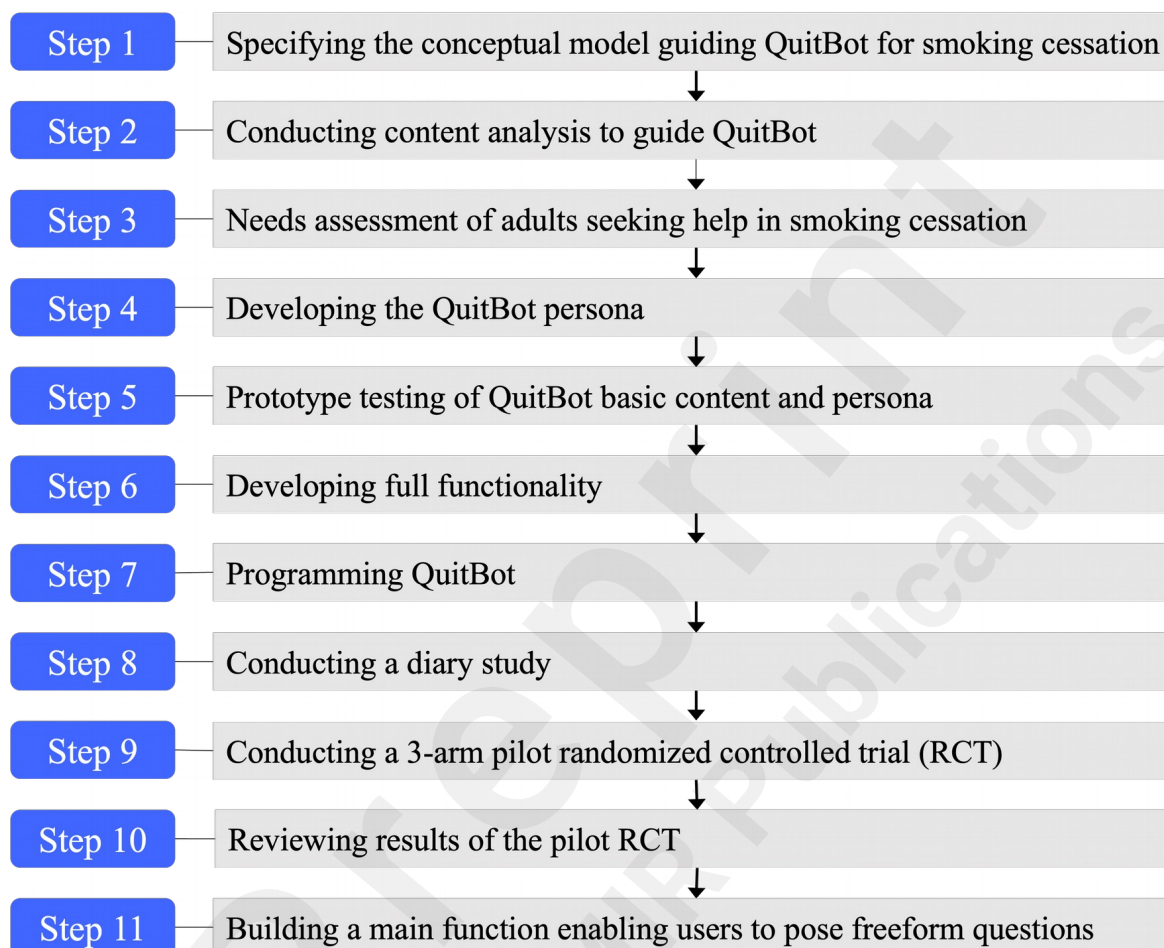


Figure 1. Overview of QuitBot's formative research process

The steps were: (1) specifying a conceptual model to guide the QuitBot intervention targets, (2) conducting content analysis of existing smoking cessation interventions to guide the QuitBot coaching conversations, (3) conducting a needs assessment to determine what an adult seeking help in quitting smoking would need from a cessation chatbot, (4) developing the QuitBot persona, or personality of the chatbot, to shape the user's experience of and bond with the QuitBot chatbot, (5) prototyping QuitBot's basic content and persona, (6) developing the full functionality of the QuitBot, (7) programming the QuitBot, (8) conducting a diary study for user feedback on their interactions with QuitBot and its design and content, (9) conducting a pilot randomized trial to test QuitBot for smoking cessation, (10) reviewing results of the pilot trial, and (11) adding a free-form question and answer (QnA) function, based on user feedback from pilot trial results. The process of adding the

QnA function itself involved a three-step process: (a) generating QnA pairs, (b) fine-tuning Large Language Models (LLMs) on the QnA pairs, and (c) evaluating the LLM model outputs.

Step 1: Specifying the conceptual model guiding QuitBot for smoking cessation

The conceptual model guiding the development of QuitBot for smoking cessation (Figure 2) focuses on impacting user engagement through four therapeutic alliance processes. The four processes are (1) *bond* with QuitBot, (2) agreement on smoking cessation *goal*, (3) agreement on *tasks* for achieving smoking cessation goal, and (4) *perception that QuitBot understands* user's current needs [39]. These working alliance processes have predicted smoking cessation [40] and quit attempts [41], and have mediated the impact of human therapist-delivered smoking cessation interventions [41]. QuitBot employs various strategies to establish a therapeutic alliance, including expressing empathy for the user's struggles [42, 43], engaging in social dialogue [44, 45], employing meta-relational communication (discuss the relationship) [46], and expressing happiness at interacting with the user [43]. Language constructs such as inclusive pronouns [42], politeness strategies [47], and the use of greetings and farewells rituals [48] contribute to the creation of this alliance as well. Compared to a technology that did not use these verbal behaviors, a conversational chatbot for physical activity increased these therapeutic alliance processes, which in turn, was predictive of higher engagement with the chatbot [49].

Agreement on smoking cessation goal starts by collaboratively setting a quit date (e.g., "Have you thought about a specific day you would like to quit? Generally, I recommend about 14 days away."). QuitBot enhances perceived understanding by promptly addressing the user's immediate needs (e.g., "You say you are tempted by friends who smoke. Here's a tip that might help..."). Additionally, self-disclosure [50] is employed to foster perceived understanding, generating various positive outcomes, especially when the listener responds with support and validation [51]. A chatbot which used self-disclosure increased the user's perception that the chatbot understood their needs, which in turn predicted more positive mood [52].

Step 2: Conducting content analysis to guide QuitBot

The content analysis aimed to establish a natural flow of coaching conversations for QuitBot, aligned with US Clinical Practice Guidelines for smoking cessation [53]. In the initial phase of the content analysis, we interviewed a panel of experts, including three smoking cessation counselors, a smoking cessation master trainer, and a tobacco cessation scientist from our team. This panel consisted of four women and one man, with 20% racial/ethnic minority. Among them, 40% held a PhD in clinical psychology, while 60% had master's degrees in counseling or social work. Collectively, they had a wealth of experience ranging from three to twenty years, with an average of eight years, in developing and delivering smoking cessation interventions. Deductive coding of these interviews and expert consensus iteratively lead to the formulation of common themes, domain-specific responses, and anticipated user interactions that QuitBot should address. We identified common conversation topics, about smoking cessation including triggers to smoke (physical, emotional, and situational triggers), motivations to quit, and barriers to quitting. Interviews also highlighted the importance of QuitBot's persona to be sensitive and empathetic to the user and to express that their concerns are being heard.

Guided by this expert consensus on conversation topics, the second phase was to extract the content and flow of smoking interventions as they naturally occur in actual interactions between cessation counselors and patients. To achieve this, we conducted semantic analysis of verbatim manually transcribed intervention conversation transcripts obtained from our telephone counseling intervention trial, randomly selected among those who did and did not quit smoking (R01 DA038411) [54]. A total of 159 call transcripts (equating to 63 hours and 23 minutes) were randomly selected, constituting a 20% sample from each of the five sessions (with an average session duration of 22.9 minutes) of an efficacious behavioral intervention for smoking cessation with a 25% 30-day point prevalence abstinence rate at the 12-month follow-up [54]. These sessions covered various topics, including motivations to quit, triggers to smoke, barriers to quitting, setting a quit date, developing a quit plan, education and proper use of FDA-approved medications for quitting smoking, coping skills for dealing with urges, enlisting social support, and strategies for avoiding external cues to smoke. Participants had a mean age of 47.4 (SD: 12.7), with 44% male and 22% identifying as racial/ethnic minority.

Transcripts underwent deductive coding using a predefined codebook to identify common conversation topics related to smoking cessation, such as triggers to smoke (physical, emotional, and situational triggers), motivations to quit, and barriers to quitting. Utilizing a supervised machine learning approach, these topics formed the basis of Quitbot's entity extraction, wherein elements of the unstructured transcript text were coded into predefined categories. Subsequently, we determined intent classifications, which involved discerning the meaning of the user's text. Finally, we mapped out the natural conversational flow of both the chatbot and the range of verbal responses and comments that users might provide in response to the chatbot. The entity extraction, intent classifications, and conversational mapping were conducted using the LUIS conversational AI program [55].

Step 3: Needs Assessment of adults seeking help in smoking cessation

Assessing the needs of adults seeking help in smoking cessation interventions shapes what the user should be able to do with a chatbot. To assess user needs, we first analyzed the results of the content analysis phase. Subsequently, we conducted interviews with five adults who had participated in our human-delivered smoking cessation interventions within the past year (including two who quit and three who did not quit) [56]. Participants were mean age 46.1 (SD: 10.4), 40% female, and 31% racial/ethnic minority. The interviews queried participants about their personal background and smoking history, expectations for a smoking cessation program, experiences with a human cessation coach, perceptions regarding setting, keeping and changing quit dates, coping skills for urges to smoke, and attitudes toward and expectations of what a chatbot could do for helping them quit smoking. Semi-structured interviews were conducted in-person at the lead's author's UX Research HABIT lab. The deductive thematic analysis organized the user's responses by grouping them into themes, reviewing the themes, and then interpreting them [57-59]. The themes of the key user needs identified were: (1) a coach who can make a personal connection, (2) on-demand help with urges, and (3) skills for preparation to quit and preventing relapse.

Step 4: Developing the QuitBot "persona"

The user's bond with the chatbot is impacted by its persona [49]. Based on interviews with smoking cessation coaches and our master trainer, we created the persona to foster a strong

bond with users. Presented to the user as a computer program (e.g., “I’m a bot designed to help you live smoke free”), elements of the QuitBot persona included expressions of empathy [42, 43], social dialogue [44, 45], meta-relational communication (discuss the relationship) [46], and expressing happiness to see the user [43]. Additionally, specific language constructs, including inclusive pronouns [42], politeness strategies [47], and greetings and farewells rituals were integrated to enhance the user experience and promote a respectful dialogue [48]. Finally, we established eleven core values for the persona, serve as guiding principles for QuitBot’s behavior throughout conversations.

Step 5: Prototype testing of QuitBot basic content and persona

The prototyping testing of QuitBot’s basic content and persona aimed to assess users’ initial responses to basic smoking cessation conversations between the user and the persona. Stimuli were built using botmock.com[60] to develop the chat dialogue, which was then integrated into Facebook Messenger using Chatfuel.com [52]. Participants had a guided initial chat conversation introducing the chatbot and program goals, querying about triggers for smoking, and setting a quit date. Subsequently, they interacted with QuitBot for a second conversation focusing on tracking triggers to smoke. For both conversations, a UX researcher frequently paused to prompt participants to think-aloud their experiences with QuitBot. Real-time interactions between the user and QuitBot were facilitated by a UX researcher using the Chatfuel program [61]. To evaluate this process, 75-minute individual interviews were conducted with eight adults interested in quitting smoking. Four were chosen because they thought a chatbot could be helpful for quitting smoking, while the remaining four were selected because they were unsure or skeptical that a chatbot would be helpful for quitting smoking. The mean age of participants was 42 (SD: 11.1), with 38% male, 50% female, and 12% transgender. Additionally, 38% had high school education or less and 25% reported minority race/ethnicity backgrounds.

Semi-structured interviews

Semi-structured interviews were conducted in-person at the lead author’s UX Research HABIT lab. A deductive thematic analysis method was used to organize user responses into

themes, review those themes, and then interpret them [57-59]. Despite initial skepticism from half of participants regarding the usefulness of interacting with a digital coach, the results showed a notable shift in the interest in QuitBot by the end of the interview: 100% reported that a chatbot like QuitBot would be valuable for helping someone quit smoking, with 86% expressing willingness to try this chatbot for quitting. Additionally, all participants (100%) found Quitbot easy to use, noting its conversational tone as “encouraging,” “polite”, and “reassuring.” They deemed the length and speed of onboarding conversations appropriate and felt comfortable providing conversational responses. Participants expressed surprise at the “humanness” of QuitBot’s avatar, noting its informal, reassuring, accessible, and easy-to-talk-to demeanor.

When discussing whether the avatar should have a gender or a name, there was consensus among participants in favor of a female persona, with the name “Ellen” deemed appropriate. (Interestingly, one of the initial participants spontaneously suggested “How about a woman named ‘Ellen?’”) Later participants concurred with this choice when asked by the UX researcher.

Participants also expressed a desire for more actionable suggestions and to open and close each conversation with a specific plan of action. In response, we added a plan outlining what to anticipate, letting them know that the avatar would initiate a chat the following day and introduce a new quitting smoking skill in the subsequent conversation. Some participants wanted additional time to decide on a quit date, prompting us to include a dialogue indicating that they postponed setting a quit date until they felt ready. Additionally, participants suggested visualizing their progress in quitting smoking, such as through a graph. In response, we added a progress chart displaying the number of cigarettes smoked over time. Overall, participants described feeling “captivated” by the content and expressed eagerness to learn more.

Step 6: Developing full functionality

Building upon the prototype as the foundation, we applied insights from the earlier steps to

develop a full program consisting of 42 days of 2 to 3-minute focused conversations. These conversations are distributed over several phases of treatment: a pre-quit phase (14 days of content), quit day (1 day of content), and post-quit phase (27 days of content). There are also conversations for those not ready to quit smoking by day 14 (6 days of content) and conversations for those who have relapsed (3 days of content). The content, described in Table 1, follows US Clinical Practice Guidelines for cessation interventions [62]. The program content is presented as a continuous conversation, built on user input from prior conversations. This ensures a personalized and adaptive approach based on the user's stated motivations to quit, triggers to smoke, and number of cigarettes smoked. QuitBot is proactive and provides daily prompts to start a structured text conversation with Ellen at the user's preferred time—e.g., “Hi Alex, are you free to chat?” Users also have the flexibility to reach out to Ellen at any time for on-demand help with urges, inspiration, mood, and slips.

Step 7: Programming QuitBot

We initially sought a development architecture with the flexibility to interact with QuitBot on any major consumer communication channel (e.g., as a standalone app, Facebook Messenger, Slack, etc.). Such flexibility adapts to current consumer trends in communication technology use, making QuitBot available for use on the channels with current high population level reach. For determining which communication channel for interacting with QuitBot, we conducted an online survey of 100 US adults who smoke, asking them which of these channels they would prefer for a chatbot: standalone app, WhatsApp, Facebook Messenger (FM), Skype, or Slack. The majority (74%) of respondents preferred FM, citing its familiarity, ubiquity, and ease of use. FM is an instant messaging service for online chats. At that time (2019), there were over 133 million FM users in the US (1.3 billion globally) and FM hosts over 300,000 chatbots, with 27% of them for healthcare (e.g., exercise) [63-65]. Following these findings, we hosted QuitBot on FM.

Therefore, we custom built an architecture using the Microsoft Bot Framework that employs Microsoft Azure for the cloud computing and Microsoft Language Understanding (LUIS) platform for the natural language understanding of the QuitBot guide “Ellen”. The preference for natural language understanding over an if-then decision-based conversation flow was made

to ensure a more natural and open-ended interaction, allowing for a broad range of responses and better conveying that the user is being heard and understood. QuitBot's LUIS allows it to understand common text shorthand. Users can respond freely or simply select from a menu of responses. If QuitBot does not understand a free response, it will say so and ask the participant to rephrase the response. QuitBot was written in the programming language of Node.js[66].

Step 8: Conducting a diary study

We (DK, BS, JB) conducted a diary study to obtain ongoing feedback on users' interactions with QuitBot, its design, and content. In user-centered design research, a diary study of two weeks with six to twelve participants is recommended to obtain this initial feedback [67, 68]. Accordingly, we conducted a single arm 14-day diary study of the program with nine adults who were smoking at least daily (all smoked 30 or more cigarettes per day), were interested in quitting smoking, and recruited from around the US via Facebook ads. Four were chosen because they were skeptical about chatbots being able to help someone quit smoking while the remaining five were neutral about them. Participant demographics: mean age 40.4 (SD: 13.4), 12% minority race, 44% female, and 70% had less than a bachelor's degree.

All nine participants completed: (1) the 60-minute video-based orientation focusing on how to use QuitBot and complete the daily diary entries, (2) 14 evening diary entries (15 minutes each) about their daily interactions with QuitBot, its design, and content, (3) on day seven, a mid-point 15-minute video call with a member of our user research team to review their impressions to date, and (4) a 60-minute video call exit interview with a member of our user research team. A PhD-level UX researcher with over 20 years of experience conducted the interviews. (Example questions from the exit interview: "Which parts of the app did you find the most helpful? Why?") Semi-structured interviews were conducted in-person at the lead's author's UX Research HABIT lab. The deductive thematic analysis organized the user's responses by grouping them into themes, reviewing the themes, and then interpreting them [57-59].

The results showed that, even though the focus was on usability, by day 14, three participants quit smoking and all remaining six reduced to 3 or 4 cigarettes per day. Ratings for usefulness

“Overall, how useful was the QuitBot app for helping you quit smoking?”), satisfaction (“Overall, how satisfied were you with the QuitBot?”), and likelihood of recommending QuitBot (“To what extent would you recommend QuitBot to someone who would like to quit smoking?”) were all high: 4.33, 4.67, and 4.88 on a 0 (not all) to 5 (extremely) scale. All nine users felt highly supported by Ellen and liked her persona. They liked the skills training for coping with smoking urges and lapses. Their feedback yielded minor content edits and fixes of technical bugs. Representative functionalities of QuitBot are shown in Figure 3.

Diversity and Inclusion in User Experience (UX) design

The diversity of race, gender, age, and educational characteristics of users who participated in our UX design studies described above, influenced the design in many ways, including Ellen’s persona design (e.g., men and women both preferred a female persona), Ellen’s stories of people who have quit smoking (e.g., they were diverse in age, gender, race, education), use of language (e.g., 5th grade reading level, informal and respectful), and user interface (e.g., larger response buttons, larger font size [69]).

Step 9: Conducting a 3-arm pilot randomized controlled trial (RCT)

The favorable feedback from the diary study led us to conduct a three-arm parallel pilot RCT comparing QuitBot (n = 200) to the Smokefree TXT (n = 149) intervention and to a QuitBot delayed access control group (n = 55). Following expert recommendations for pilot trial design [70, 71], the feasibility outcomes were the study’s primary focus in order to inform the further development of QuitBot and design of a future full-scale trial of QuitBot. Since this pilot RCT was the first time QuitBot was tested and no prior RCTs had been reported on any quit smoking chatbot, estimated effect sizes were unknown. Instead, the sample sizes were based on comparable sample sizes from prior pilot studies we have conducted in our lab [72, 73]. Participants were recruited nationwide and were randomized to intervention arm using randomly permuted blocks of size 2, 4 and 6, stratified by biological sex (male vs. female), heaviness of smoking index score (≤ 4 vs >4), and percent confidence in being smoke-free in 12 months ($\leq 70\%$ vs. $>70\%$). The study was double-blinded, with both interventions called “QuitBot.” All study procedures were approved by the Fred Hutch Cancer Center Institutional

Review Board (Protocol Number: 8659/RG1001766). The clinical trial protocol was approved by the Fred Hutch Scientific Review Committee (Approval number: FHIRB008659) and the trial was registered on ClinicalTrials.gov (Identifier: NCT03585231). There were no deviations to the registered protocol.

Eligibility criteria for the pilot RCT

Inclusion criteria included: (1) age 18 and older, (2) having smoked at least one cigarette a day for at least the past 12 months, (3) wanting to quit cigarette smoking within the next 14 days, (4) if concurrently using any other nicotine or tobacco products, wanting to quit using them within the next 14 days, (5) being interested in learning skills to quit smoking, (6) being willing to be randomly assigned to either condition, (7) residing in the US, (8) having daily access to their own smartphone, (9) having both text messaging and FM on their smartphone (criterion 8 and 9 were required to receive each interventions' content), (10) being willing and able to read in English, and (11) not using other smoking cessation interventions. Individuals deemed ineligible to participate were directed to the smokefree.gov website and the 800-QUIT-NOW number for access to their state's quitline resources.

SmokefreeTXT (SFT) comparison condition

For the past 20 years, mobile phone-delivered text messaging interventions have been a prominent technology for delivering smoking cessation interventions [74-78]. Each year text messaging smoking cessation interventions are reaching over 300,000 US adults who smoke and 6 million adults who smoke worldwide [79-81]. SFT's 42-day program was developed by the National Cancer Institute (NCI). SFT is the most widely accessible text messaging program in the US. SFT is non-proprietary and free to the public, thereby providing maximal accessibility and replicability. Daily messages are sent about the importance of quitting smoking, setting a quit date, preparing to quit, quitting, and maintaining abstinence. Daily messages check in about quit status. Three keywords can be proactively sent by users to receive anytime help:

“CRAVE” (on how to cope with urges), “MOOD” (on how to cope with moods triggering smoking), and “SLIP” (on how to cope with lapses). Participants do not

need to respond to or otherwise engage with SFT messages in order to complete the SFT program. See **Figure 4** for sample messages.

NCI's SFT contractor (ICF International; icf.com) provided us with the full content of SFT so that we could internally host a secured private version for research. In both SFT and QuitBot, participants receive two prompts per day (three on the quit day). Comparisons between QuitBot and SFT are shown in **Table 1**.

Table 1. QuitBot and SFT: Phases, Corresponding Content, and How Content is Communicated.

Phase (no. days of content)	Content	How Content is Communicated
Pre-Quit (14 days)	Triggers to Smoke; motivations for quitting; setting quit date; FDA-approved medication information; skills to be aware of and cope with urges; cessation progress tracking	<p><i>SFT:</i> Sends texts of the content; Answers to daily check-ins (e.g., #cigs/smoked today); get 1-2 text responses; answers to entering anytime help keywords (e.g., CRAVE); get 1 text response.</p> <p>vs. <i>QuitBot:</i> Digital coach sends user a greeting to start a 2–3-minute conversation; presents content in a dialogue with the user via engagement features described in Figure 3 (e.g., tailored responses, empathy); answers to entering anytime help keywords (e.g., CRAVE) initiate a dialogue.</p>
Quit Day (1 day)	Encouragement; Smoking status check-in	
Post-Quit (28 days)	Withdrawal symptoms education; slips and relapse prevention; managing mood; managing cravings; cessation progress tracking	
Not Ready/Quit Date $x > 14$ Days	Reviews motivations for quitting; cessation progress tracking	
Anytime Help	Skills to cope with urges, mood, and slips	

QuitBot delayed access comparison condition

To explore the unique impact of QuitBot on smoking cessation, considering that some participants might quit smoking without intervention, we introduced a delayed access comparison condition. In this condition, 55 participants received delayed access to QuitBot after completing the 3-month follow-up survey. The delayed access comparison condition served ethical purpose of providing participants access to a treatment (as opposed to no treatment at all).

Outcome measures

Outcome data were collected through an encrypted web-based survey. Participants not

completing the web-based survey were sequentially offered the survey via phone, mailed survey, and postcard. The primary feasibility outcomes were (1) sufficient accrual of the planned number of study participants, (2) balanced demographic and smoking characteristics at baseline between study arms, and (3) retention of the primary 30-day point prevalence abstinence (PPA) smoking outcome at the 3-month follow-up. Intervention engagement was assessed based on comparing the active treatment study arms on the number of times and number of days participants interacted with their assigned intervention. All engagements with the participants' assigned interventions were objectively logged using an internally hosted secure server. The primary smoking cessation outcome was 30-day PPA, based on complete-case analysis, and 7-day complete-case PPA was secondary.

Statistical Analysis for the pilot RCT

The feasibility of the pilot RCT was assessed based on sufficient accrual, balanced randomization, and adequate follow-up data retention rates that did not differ between arm. Baseline characteristics were compared between the three study arms using analysis of variance for continuous variables and Fisher's exact tests for categorical variables and were summarized with the 'arsenal' package in R version 4.2.3 [82, 83]. We used generalized linear models to assess differences between study arms in the number of days participants used their intervention.

We used negative binomial models, implemented with the R package 'MASS' [84], to compare treatment arms on total number interactions because the data were heavily right-skewed. Logistic regression models were used to test the effect of treatment arm on binary smoking cessation outcomes. Based on evidence from text messaging trials meta-analyses [85], all outcome models were adjusted for the three factors used in stratified randomization: biological sex (male vs. female), heaviness of smoking index score (≤ 4 vs >4), and percent confidence in being smoke-free in 12 months ($\leq 70\%$ vs. $>70\%$). Wald tests for pairwise comparisons of each outcome between study arms were adjusted for multiple comparisons with the Holm procedure [86]. Statistical tests were considered significant at $\alpha < .05$. Deductive thematic analysis organized participants' comments about QuitBot by grouping them into themes, reviewing the themes, and then interpreting them [57-59].

Results

Step 10: Main results of the pilot RCT

Recruitment was successful

Based on our successful methods for national recruitment [87], we developed and tailored Facebook ads with ongoing monitoring and adjustment of recruitment yield. These efforts yielded 2954 participants screened, 1380 eligible, 583 consented, and 418 randomized between September 2018 and June 2019. After completion of study participation, 14 participants were found to be cases of fraud, duplicate participants, or in the same household as another participant, leading to a total of 404 participants included in analyses.

Randomization

The three stratification conditions were balanced at baseline on all measured characteristics (all p-values > .05). As shown in Table 2, participants were on average, 36 years old, 70% female, 29% reported racial or ethnic minority background, 53% were unemployed, 84% had no college degree, 72% smoked more than one half pack daily, and 60% had high cigarette dependence (FTCD scores of 6 or more). The characteristics of this FM sample are very similar to those of other digital health intervention trials [85, 88, 89].

Table 2. Baseline participant characteristics by study arm.

Baseline characteristic	Total (N=404)	SFT (N=149)	Delayed (N=55)	QuitBot (N = 200)	P value
Age, mean (SD), years	36.0 (10.4)	36.2 (11.2)	35.6 (9.6)	35.9 (9.9)	.92
Female gender	283 (70.0%)	103 (69.1%)	39 (70.9%)	141 (70.5%)	.95
Race					.52
White	296 (73.3%)	110 (73.8%)	40 (72.7%)	146 (73.0%)	
Black or African American	51 (12.6%)	21 (14.1%)	9 (16.4%)	21 (10.5%)	
Asian	2 (0.5%)	0 (0.0%)	0 (0.0%)	2 (1.0%)	
Native American or Alaska Native	12 (3.0%)	4 (2.7%)	0 (0.0%)	8 (4.0%)	
Native Hawaiian or Pacific Islander	1 (0.2%)	0 (0.0%)	0 (0.0%)	1 (0.5%)	
Multiple races	31 (7.7%)	13 (8.7%)	4 (7.3%)	14 (7.0%)	
Unknown race	11 (2.7%)	1 (0.7%)	2 (3.6%)	8 (4.0%)	
Hispanic ethnicity	28 (6.9%)	7 (4.7%)	6 (10.9%)	15 (7.5%)	.27
Minority race or ethnicity ^a	116 (28.9%)	42 (28.2%)	17 (30.9%)	57 (28.9%)	.93
Married	104 (25.7%)	32 (21.5%)	16 (29.1%)	56 (28.0%)	.32
Employed	191 (47.3%)	80 (53.7%)	24 (43.6%)	87 (43.5%)	.14
At least some college	339 (83.9%)	126 (84.6%)	47 (85.5%)	166 (83.0%)	.87
Heavy alcohol use ^b	47 (11.9%)	18 (12.5%)	6 (11.3%)	23 (11.6%)	.96
Positive depression screening results ^c	223 (55.5%)	91 (61.5%)	28 (50.9%)	104 (52.3%)	.17
Smoking behavior					
FTCD score, mean (SD)	5.7 (2.0)	5.5 (2.0)	6.1 (2.2)	5.7 (2.0)	.17
High nicotine dependence	242 (59.9%)	88 (59.1%)	36 (65.5%)	118 (59.0%)	.66
Smokes more than one-half pack per day	289 (71.5%)	98 (65.8%)	42 (76.4%)	149 (74.5%)	.14
Smokes more than one pack per day	66 (16.3%)	20 (13.4%)	14 (25.5%)	32 (16.0%)	.11
First cigarette within 5 minutes of waking	205 (50.7%)	75 (50.3%)	34 (61.8%)	96 (48.0%)	.19
Smoked for ≥10 years	317 (78.5%)	112 (75.2%)	44 (80.0%)	161 (80.5%)	.46
Used e-cigarettes at least once in past month	122 (30.2%)	42 (28.2%)	16 (29.1%)	64 (32.0%)	.73
Quit attempts in past 12 months, mean (SD) ^d	1.6 (4.7)	1.6 (3.3)	1.1 (3.2)	1.7 (5.8)	.68
At least one quit attempt in past 12 months ^e	145 (38.5%)	51 (37.8%)	16 (30.2%)	78 (41.3%)	.33
Confidence to quit smoking, mean (SD)	64.1 (27.0)	62.6 (27.0)	72.2 (27.3)	62.9 (26.8)	.05
Friend and partner smoking					

Close friends who smoke, mean (SD)	2.8 (1.7)	2.8 (1.7)	2.7 (1.6)	2.8 (1.8)	.97
No. of adults in home who smoke, mean (SD)	1.5 (0.9)	1.4 (0.9)	1.7 (1.1)	1.5 (0.8)	.19
Living with partner who smokes	145 (35.9%)	51 (34.2%)	24 (43.6%)	70 (35.0%)	.43

^an=401; ^b n=395; ^cn=402; ^dn=377; ^en=377.

The 3-month follow-up rates were high

To maximize outcome data completion, we followed our team's successful protocol [87]: four sequential survey modalities (first Web, followed by phone, mail, and postcard). As agreed in the informed consent, participants received \$25 for submitting their responses and received an additional \$10 bonus for completing the Web survey within 24 hours. The achieved *outcome survey completion rate of 96%* provided confidence in the follow-up survey methods. The data retention did not differ between study arms ($P = .54$). Given the limitations of the pilot budget, cessation data were self-reported.

Engagement and cessation results were promising for QuitBot (See Table 3)

The number of times participants interacted with their assigned intervention was 1.3 times greater in QuitBot as compared to SFT (IRR = 1.33; 95% CI: 1.04, 1.70; $P = .02$). Participants used their assigned intervention 11 days longer in the QuitBot arm than in the SFT arm (Point estimate = 11.5; 95% CI: 4.9, 18.1; $P = .001$). QuitBot's intervention completion results are substantial when considering that each day's content involves a 2 to 3-minute conversation. (By contrast, SFT participants do not need to respond to or otherwise engage at all with their messages in order to complete their program: daily SFT text messages are sent automatically). Participant engagement was limited by QuitBot's inability to answer participants' open-ended questions (see below). Therefore, cessation results are reported for all participants and for participants who completed their assigned intervention.

For all participants, the 30-day point prevalence abstinence (PPA) rates at three-month follow-up were 31% for QuitBot vs. 35% for SFT (QuitBot vs. SFT: OR = 0.81; 95% CI: 0.50, 1.29; $P = .36$) vs. 7% for delayed treatment (QuitBot vs. Delayed: OR = 5.97; 95% CI: 2.04, 17.45; $P = .002$). For those who completed their assigned intervention (i.e., viewed all 42 days of planned content), the 30-day complete-case, PPA rates at three-month follow-up were 63% for QuitBot vs. 38% for SFT (QuitBot vs. SFT: OR = 2.58; 95% CI: 1.34, 4.99; $P = .005$). The pattern of results was highly similar for the outcome of 7-day complete-case PPA rates at three-month follow-up—albeit with higher abstinence rates in each study arm.

Table 3. Comparison of QuitBot, SFT, and Delayed Intervention on Three-Month Engagement and Cessation Outcomes.

Study outcomes	SFT ^a (N=149)	Delayed ^b (N=55)	QuitBot ^c (N = 200)	QuitBot ^c vs. SFT ^a IRR ^d , PE ^e , or OR ^f	QuitBot ^c vs. Delayed ^b OR ^f
Three-Month Engagement or Cessation Outcome	Summary	Summary	Summary	(95% CI); <i>P</i> value	(95% CI); <i>P</i> value
Number of times interacted mean (SD), median	24.2 (25.8), median=15	---	32.9 (29.0), median=25	IRR: 1.33 (1.04, 1.70); <i>P</i> = .02	---
Days from randomization to last input, mean (SD), median	44.1 (22.7), median=54	---	55.7 (36.0), median=70	PE: 11.5 (4.9, 18.1); <i>P</i> = .001	---
30-day cigarette abstinence among all participants, (%)	50 (35%), n=144	4 (7%), n=54	59 (31%), n=190	OR: 0.81 (0.50, 1.29); <i>P</i> = .36	5.97 (2.04, 17.45); <i>P</i> = .002
30-day cigarette abstinence among program completers, (%)	45 (38%), n=117	---	39 (63%), n=62	OR: 2.58 (1.34, 4.99); <i>P</i> = .005	---
7-day cigarette abstinence among all participants, (%)	76 (53%), n=144	5 (9%), n=54	91 (48%), n=190	OR: 0.79 (0.51, 1.22); <i>P</i> = .28	10.08 (3.79, 26.80); <i>P</i> < .001
7-day cigarette abstinence among program completers, (%)	70 (60%), n=117	---	50 (81%), n=62	OR: 2.63 (1.24, 5.55); <i>P</i> = .01	---

^aSFT=SmokefreeTXT intervention; ^bDelayed= 3-month delay in receiving QuitBot; ^cQuitBot=QuitBot intervention; ^dIRR=Incidence Rate Ratio; ^ePE=Point Estimate; ^fOR=Odds Ratio

Representative QuitBot Comments

Comments from QuitBot arm trial participants reflected a strong overall bond with the chatbot's persona: "I loved Ellen. She was always there when I needed her"; "Ellen was always there for me when I had a craving"; "I love how engaged she was, I could really quit with her there to talk to"; "She made me feel like I was not alone"; "She was there without making me feel ashamed"; "She was kind, non-judgmental"; "She held me accountable"; "Felt like a friend encouraging me."

Conversely, participants were frustrated by QuitBot's inability to respond to their specific questions about quitting smoking: "I could not ask questions and get real answers back"; "I could not ask it real live questions"; "I wanted to write my own questions"; "Can't ask any question"; "Not being able to respond to my questions"; "I wish you could talk to her...without it being a constant couple of options"; "I didn't like how it selected responses"; "The fact that you cannot ask a question and [it] has no idea what you are saying unless you select one of the options."

Main conclusions from the pilot RCT

Our main conclusions were: (1) the intervention demonstrated potential for rigorous testing based on sufficient accrual, balanced randomization, and high retention rates, (2) overall, there was a strong engagement with QuitBot, and (3) promising quit rates were observed—particularly among participants who completed the content of their assigned intervention. The effectiveness of QuitBot was evident, as quit rates in the delayed condition group were significantly lower (31% vs. 7%), indicating a net increase in smoking cessation of 24%. Therefore, it is highly unlikely that effects of QuitBot were merely due to the passage of time or baseline motivation to quit smoking (i.e., few participants quit without offering intervention).

Challenges were also identified, potentially impacting participant engagement, and quit rates. Specifically, QuitBot's inability to respond to participants' own questions about quitting smoking led to a significant level of frustration. While the participant can answer questions asked by the QuitBot (e.g., "Tell me what is triggering your urge."), the reverse was not possible: participants could not ask QuitBot their own questions. A QuitBot feature which allows participants to ask freeform questions would be needed to address this limitation.

Technical limitations of the Facebook Messenger platform

While FM was the preferred communication platform from our survey results, Facebook introduced changes that would limit participants' engagement with QuitBot as well as our own access to user data: (1) Facebook made policy changes that revoked access permissions to proactively outreach (e.g., to invite participant to check in or start a conversation), effectively removing our ability to proactively contact users (restricting that ability to news-related apps only) and (2) Facebook made platform changes that restricted our ability to access demographic information of users, inhibiting data collection. Facebook's changes raised concerns about the feasibility of QuitBot's conversational functionality and data collection. This critical limitation could be addressed by transitioning to a standalone smartphone app communication platform, enhancing accessibility and control for both participants and the development and research teams.

Step 11: Building a main function enabling users to pose freeform questions about smoking

The goal of this specific QuitBot refinement was to build a main function of QuitBot that would enable users to pose freeform questions about quitting cigarette smoking and for the QuitBot to respond with accurate, concise, professional, and non-repetitive answers. This was an iterative three-step process detailed in **Table 4**.

Table 4. Steps, sources, and results of QuitBot's question and answer (QnA) iterative development process.

Step	Source (Year)	Results
1. Generate QnA Pairs	<ul style="list-style-type: none"> ▪ Alexander Street Therapy Transcripts (2020) ▪ National Cancer Institute Call Center Transcripts (2020) ▪ HABIT Lab Cessation Counseling Intervention Transcripts (2020) ▪ HABIT Lab Digital Intervention Content (2020) ▪ HABIT Lab Clinical Team Generates QnA (2020-2021) ▪ Prolific Survey of Adults Who 	<ul style="list-style-type: none"> ▪ 11000 smoking QnA pairs ▪ 8223 chitchat QnA pairs

	Smoke (2021)	
2. Training LLM models on QnA Pairs	<ul style="list-style-type: none"> ▪ Azure API (2020-2023) ▪ DialoGPT (2021) ▪ ParlAI (2021) ▪ Davinci GPT-3 (2021) ▪ Curie GPT-3 (2021) ▪ Ada GPT-3 (2021) ▪ Contextualized GPT-3.5 (2022) ▪ GPT 4.0 (2023) 	<ul style="list-style-type: none"> ▪ Models with higher self-scored confidence about answers provided: processed by Azure API ▪ Models with lower self-scored confidence about answers provided: handled by Curie GPT-3 (2021), replaced by Contextualized GPT-3.5 (2022)
3. Evaluating LLM Outputs	<ul style="list-style-type: none"> ▪ Automated Evaluation: pertinence & grammaticalness (2020-2023) ▪ Manual Evaluation: accuracy & tone (2021-2023) ▪ 14-day User Testing (2021) 	<ul style="list-style-type: none"> ▪ Identified the answers that were repetitive, incorrect, or had impersonal/non-professional tone.

Step 11.1. Generating QnA Pairs

The first step was to develop a knowledge base of QnA pairs focused on the topics of quitting cigarette smoking. Smoking cessation clinical data sources included the Alexander Street datasets of counseling transcripts [90], NCI Call Center transcripts of smoking cessation coaching conversations [91], and transcripts of counseling sessions from our Fred Hutch Cancer Center research lab's prior smoking cessation intervention trials [54]. The categories of smoking cessation questions spanned a wide range, including medications to aid smoking cessation, the role of vaping e-cigarettes in quitting smoking, health consequences of smoking on self and others, staying motivated to quit smoking, triggers to smoke, barriers to quitting smoking, tips for managing cravings and withdrawal, relapse prevention and recovery.

For generating a diversity of QnA sources, the knowledge base was broad, drawn from Alexander Street transcripts of therapy sessions [90], NCI Call Center transcripts of online Live Chats (livehelp.cancer.gov), Fred Hutch Cancer Center's HABIT lab cessation counseling intervention transcripts [54], written clinical content from HABIT lab digital smoking cessation interventions [92, 93], and manual generation of questions and answers by HABIT lab clinical team members. We created a sequence-to-sequence (seq2seq) training model and processed cleaned transcripts, generating more than 8,000 question and answer (QnA) pairs specifically

focused on the topic of quitting cigarette smoking. As detailed in Step 3, our evaluations led us to revisit Step 1. In this iteration, we generated 2,000 new questions posed by 32 adults interested in quitting cigarette smoking in a Prolific online survey. Then our HABIT lab clinical team members generated answers to those questions. We also generated 1,000 additional new questions from NCI call transcripts and transcripts related to our lab's intervention research. This iterative process yielded over 11,000 QnA pairs.

Step 11.2. Training Large Language Models (LLMs) on QnA Pairs

The second step was to use these QnA pairs to train a series of LLMs. The first LLM deployed was an Azure API customized for the chatbot's six-level chat architecture:

1. Motivations to quit smoking; triggers to smoke; and frequency of smoking.
2. Dispatcher dialogue flow, to interpret and route utterances to the correct target, either Azure's 8,000 QnA library or a custom AI model.
3. QnA-maker-based "Caring" personality chitchat to express empathy and provide supportive interactions with the user.
4. QnA-based custom "help" to respond to trigger words like suicide.
5. Custom transformer-based model to handle chitchat QnA scenarios in dialogues that do not align with any specific intent.
6. Custom transformer-based model designed to address smoking-related QnA scenarios in dialogues that do not align with any specific intent.

In 2021, we tested four neural network models: a pre-existing DialoGPT Chitchat model, two smoking content-trained DialoGPT models (one trained on our 8,000 QnA pairs, one on NCI chatlogs), and one smoking content-trained ParlAI model. We then accessed OpenAI's GPT-3 and conducted a comparative analysis of its model performance against our then best-performing model DialoGPT. GPT-3 outperformed the specially trained DialoGPT. Subsequently, to further improve GPT-3 performance, we compared model options: (a) Davinci, (b) Curie, and (c) fine-tuned Curie, concluding that the fine-tuned Curie model provided the best performance. Upon the release of GPT-3.5 (March 2022), we compared its performance with the fine-tuned GPT3 Curie model [94]. This evaluation involved batch testing of questions that were not present in the

QnA library, followed by a ranked review of answers. GPT3.5 provided better or comparable answer quality but required additional contextualizing parameters to match fine-tuned performance tailored for smoking-specific content.

As a result, we employed GPT3.5 to expand the Azure QnA library by generating variations of existing questions and answers. This procedure expanded the Azure QnA library's capacity to accommodate a wider range of user question formulations, providing QuitBot with greater control over the quality of answers delivered to users.

Table 5. Comparisons of Large language Model answers to three sample questions on quitting smoking.

<i>Sample Question 1: What is the quickest way to stop smoking?</i>	
Davinci	The quickest way to stop smoking is to not start!
Curie	Tapering down is the best way to stop smoking. Quitting smoking is a process. Make a quit plan and notice what triggers you to smoke.
Fine-tuned Curie	Our program was created to give you a plan that will work for you. We start by making a quit plan. Write down what inspires you to quit smoking. Choose a quit date and start cutting back on cigarettes.
GPT 3.5	The quickest way to stop smoking is to quit cold turkey, which means stopping smoking abruptly and without any aids such as nicotine replacement therapy or medication. However, quitting cold turkey can be difficult and the withdrawal symptoms can be challenging. We found that it works best to follow the program steps: make a quit plan and notice what triggers you to smoke. Then use quit smoking medication to help with withdrawal and use our tools to help urges pass as you quit gradually. Either way, it's important to find a method that works for you and stick with it.
<i>Sample Question 2: I smoke when I am lonely. What should I do?</i>	
Davinci	You are addicted to the nicotine in cigarettes. The only way to truly stop smoking is through nicotine replacements or prescription medications.
Curie	It's hard to feel lonely. Notice what thoughts and feelings are there and try not smoking them away.
Fine-tuned Curie	The first step is to make a quit plan. Write down what inspires you to quit smoking. Choose a quit date and start cutting back on cigarettes. Be aware of what triggers you to smoke.
GPT 3.5	It's hard to feel lonely. Notice what thoughts and feelings are there and try not smoking them away. Trying to be a friend to yourself in that moment and do something that would be nourishing to you.

<i>Sample Question 3: Aren't e-cigarettes like the nicotine inhaler?</i>	
Davinci	No, because unlike nicotine inhalers, e-cigarettes are used as simultaneous consumption of nicotine and air.
Curie	No. While e- cigarettes/vaping don't have the harmful contaminants of tobacco smoke they are not safe.
Fine-tuned Curie	They appear similar but are actually very different. The content of the nicotine inhaler is regulated by the FDA and when used stays mostly in throat and mouth. E-cigarettes are less harmful than regular cigarettes however, they are still addictive and contain noxious chemicals which can damage the heart and blood vessels.
GPT 3.5	Although they may look similar, the content of nicotine inhalers and e-cigarettes are actually different. The nicotine inhalers are regulated by the FDA and the nicotine content stays mostly in the throat and mouth. The e-cigarettes are not regulated, and the nicotine content can vary greatly, sometimes having more nicotine than a cigarette. They can also have other harmful substances besides nicotine. Because it is heated into a mist, the vapor goes into the lungs.

Step 11.3. Evaluating Large Language Model Outputs

The third step in the development of the QuitBot's QnA feature involved evaluating the results of the LLMs using both automated and manual evaluation methods, as well as user testing from adults who wanted to quit smoking. The automated evaluation employed a scoring system of entropy (pertinence) and PPL (grammaticalness) scores to rate the quality of answers generated by the LLMs. Additionally, we implemented a filtering and scoring system to enhance the QuitBot's ability to communicate when it fails to comprehend a question. In such cases, the QuitBot asks users to be more specific if the confidence score for an answer falls below a certain threshold. The manual evaluation was conducted by trained raters in our HABIT lab, who hand scored the results of the models on measures of answer *accuracy* (Yes, No), *repetitiveness* (Yes, No), and *tone* (Acceptable or Needs Improvement). Answers requiring improvement underwent manual revisions and were included into future iterations of model testing.

After retraining the model, in 2022, adults who smoke ($N = 14$) were recruited from nationally placed Facebook ads to provide feedback on QuitBot prototype over a period of 14 days. Participant demographics: mean age 44.6 (SD = 9.9) years, 43% minority

race/ethnicity, 64% female, and 50% had less than a bachelor's degree. A PhD-level UX researcher with 20 years of experience conducted the interviews. Deductive thematic analysis organized the user's responses by grouping them into themes, reviewing the themes, and then interpreting them [57-59].

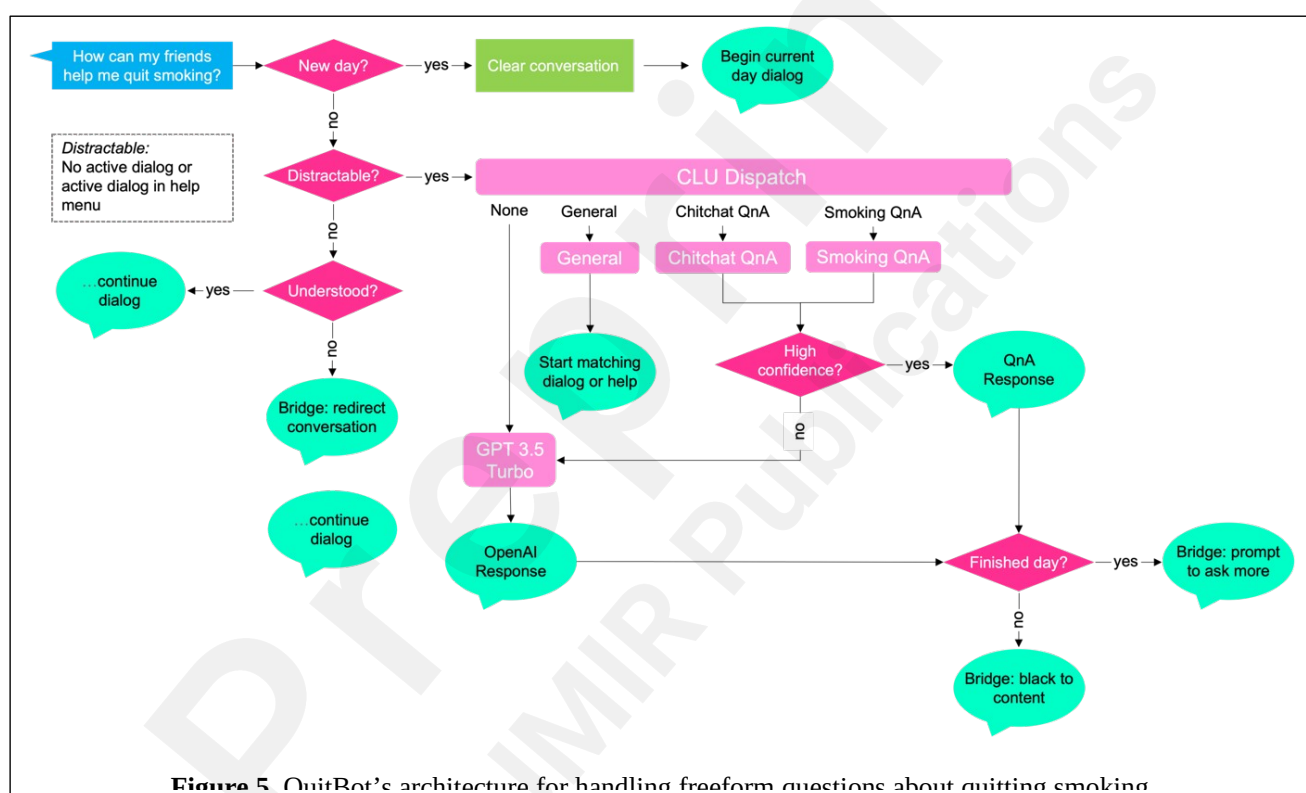
Participants were highly receptive to the structured clinical conversations and noted that the freeform chats required additional fine-tuning to enhance comprehension of the users' questions. Based on this feedback, we determined the final organization of the chatbot architecture, combining two different models (**See Figure 5**):

1. Microsoft's Azure QnA maker to handle questions matching our library of 11,000 QnA pairs that are based on scientific and clinical expertise. QnA maker uses Microsoft Language Understanding framework (LUIS) to leverage transformer models for responding to structured questions, with vetted answers stated in a professional manner.
2. Fine-tuned GPT-3.5 Turbo model to respond to questions that are not within our library of QnA pairs. Answers accurately with human-like variability, with different wording variations each time. GPT-3.5 is a backup to our QnA library.

From the perspective of computer science, the QuitBot's chatbot's program runs on a "finite state machine" [95]—which is a model of a system that runs on a limited set of modes. Depending on which mode the program is in, the QuitBot will behave in one manner or another. Ellen's state machine tracks what it is currently doing and combines interactions with the user to determine the next state. Since the user can say anything, a hierarchy of possible intentions ranked by importance to the current state is used to decide the response to the user and the next state of the conversation.

A daily welcome exchange is an example. In this example, it is the start of the day and Ellen is in the "welcome" state. The user opens the app and says, "Hi!". To handle this prompt, the user's text goes through several steps in the finite state machine illustrated in **Figure 5**: (1) categorize the user's intent; (2) determine if the intent is relevant to the current state of the conversation; accordingly, (3) move the current state of the conversation; and (4) formulate a response. In this case, "Hi!" is interpreted as a greeting intent, which is relevant to the current conversation. Ellen

moves the state of conversation to “daily check-in” and responds with a greeting of her own, “Good morning. Thanks for checking in.” When formulating a response, user’s intent determines which AI model will be used. General banter goes to a “chit chat” model powered by our Azure QnA library, smoking questions go to the Azure QnA model specifically trained on smoking questions, and unknown or low confidence intents go to GPT 3.5 Turbo. Lastly, pre-written responses from our Azure QnA library that fit into the current conversation are used for specific scenarios.



Final Version of QuitBot

The final version of QuitBot is a standalone app that features (a) a personal coach (named “Ellen”) who supports the user, (b) a series of 42 days of 2 to 3-minute structured clinical conversations with Ellen, guiding the user through distinct stages of quitting smoking and (c) the ability for users to pose any freeform question related to quitting smoking. The structured conversations provide the valuable function of a clear step-by-step program for staying motivated, learning about one’s triggers to smoke, setting a quit date, and staying smoke-free. Complementing the structured conversations, the freeform question feature provides users the

freedom to ask their own questions, the option to address unique clinical needs, and the opportunity to follow-up on the content provided in the structured conversations. The combination of both structured and freeform conversation features is intended to balance their main strengths and limitations: the structured clinical format offers a guided program on quitting smoking, albeit with limited user question flexibility, while the open-ended format provides freedom but may encounter instances of not fully understanding the user's questions to give them clear guidance—despite the positive performance of the QnA feature thus far. Representative screenshots of QuitBot are provided in the Appendix.

Discussion

This paper described the research group's four-year process of developing a conversational chatbot for cigarette smoking cessation ("QuitBot"). The user-centered development process yielded a comprehensive quit smoking program that follows a series of 42 days of 2 to 3-minute structured clinical conversations. The program content covers topics ranging from motivations to quit, setting a quit date, choosing FDA-approved medications, identifying, and coping with a wide range of triggers to smoke, and recovering from lapses or relapses. The program content, which follows the US Clinical Practice Guidelines for smoking cessation, is presented as a continuous conversation, built on user input from prior conversations. QuitBot is both proactive and on-demand for anytime help. Users can continue to interact with QuitBot after completing the 42 days of conversations.

Pilot randomized trial testing of QuitBot showed the intervention had high user engagement and promising cessation rates—especially among participants who completed their assigned intervention. However, Facebook made policy changes that revoked access permissions to proactively outreach (e.g., to invite participant to check in or start a conversation), effectively removing our ability to proactively contact users (restricting that ability to news-related apps only). This limitation was addressed by changing the FM communication platform to a standalone smartphone app that is fully accessible and controllable by the development and research teams.

The primary feedback from users was their frustration that the QuitBot could not respond to their own questions about quitting smoking. Therefore, we created the core conversational feature that would allow users to ask freeform and open-ended questions about quitting cigarette smoking and for the QuitBot to respond with accurate, concise, professional, and non-repetitive answers. We developed a library of 11,000 QnA pairs on the topic of quitting cigarette smoking. The results of our model testing showed that Microsoft's Azure-based QnA maker could handle any question that matched our library of 11,000 QnA pairs. In contrast, a fine-tuned, contextualized GPT3.5 could answer new questions that are not within our library of QnA pairs.

QuitBot has several key limitations that might present a challenge for users who expect fast responses to their questions. The QuitBot was designed for users to wait until the end of the 2 to 3-minute structured clinical conversations before they can ask freeform questions. This design element was necessary to prevent breaking the logic of each of the structured conversations, and thereby going off on tangents without an ability to return to the structured conversation. We address this design element by asking the user to hold onto their questions until the end of the structured conversation at various times throughout the program. To date, this message appears to have been effective at training the user to wait until the end of the structured conversation to ask freeform questions. The second major limitation is the freeform question response time latency when the GPT servers are running at capacity. While usually the response latency is only a few seconds, we have observed some instances where it can take up to 30 seconds. We are addressing this potential delay by telling the users that it may take a moment to answer their question, so their patience is appreciated.

On the other hand, this study has numerous strengths which have the potential to advance clinical intervention development research and practice to aid smoking cessation. Most importantly, this study illustrates the value of following a methodical, user-centered design framework in the development of technology interventions. The framework has yielded a chatbot with a comprehensive step-by-step clinical program for quitting smoking and possesses a broad knowledge base on the topic of quitting smoking. QuitBot allows users to ask freeform and open-ended questions about quitting smoking with answers informed by a broad set of clinical experience and scientific research. This technical capability has been afforded by the LLMs that

underlie the state-of-the-art versions of Azure QnA Maker and GPT. The result is that users can obtain accurate and informative answers to their questions which would otherwise be difficult to glean and evaluate from other accessible digital resources like internet searches.

By contrast, prior reports of chatbots only address certain aspects of the quit smoking process, like providing reflections on the pros and cons of smoking or helping ambivalent adults who smoke in contemplating a quit attempt. Only 8% of participants rated such chatbots as helpful [36]. Similarly, early iterations of QuitBot, which relied on a forced-choice answer format, left participants wishing responses tailored to their individual needs [96]. Likewise, a study of six users of a tablet-based chatbot, aimed at encouraging them to contemplate quitting and set a quit date, was limited by a forced-choice answer format [97]. In the only prior randomized trial of a smoking cessation chatbot, responses were confined to preset scripts, and had an outcome data retention rate of only 45% [98].

Even though users are informed that QuitBot is only a computer program, the supportive and conversational tone of the messages has the potential to lead to a long-term social-emotional connection. Indeed, the interim trial result of a mean of 72 days from first to last use is longer than we have observed in rule-based text messaging interventions for smoking cessation (which typically last about 7 days) [99-101], and longer than typical human clinician-delivered interventions such as telephone quit coaching (which typically last about 7 days) [102, 103]. Length of intervention engagement is a strong predictor of treatment success [104, 105], so these initial results on QuitBot's engagement certainly appear promising.

While we developed and tested QuitBot in the US for an English-speaking audience, the program could be tailored to other nationalities and languages across the world. As a health behavioral change platform, QuitBot has the potential to be adapted to other behavior changes including alcohol and drug use, dietary change, and physical activity.

Lessons Learned

QuitBot was developed in the midst of rapid changes in LLM technology, during what is arguably one of the most rapidly transformative periods of AI history (2020 to 2023) [106, 107]. Thus, the most important lesson we learned was the value of investing the time in continuing to iterate and improve on our freeform QnA feature as new LLMs were continuously being released. Indeed, when we started to develop the freeform QnA feature in 2020, LLM capabilities were primitive by today's (January 2024) standards. Despite being based on 345 *million* parameters, we learned that DialoGPT was limited in its ability to determine the intent of our questions. By the time GPT3.5 was released (based on 175 *billion* parameters), the performance of the freeform QnA feature was far superior, which in turn allowed us to improve on the quality of answers provided by Azure QnA.

The second most important lesson we learned is the challenge of training an LLM model for a specific clinical domain. The popular press has provided ample examples of LLMs like GPT providing very detailed answers to questions in a wide variety of topics [106, 108, 109]. While much has been written about the tendency for LLMs to “hallucinate” (i.e., providing confident-sounding answers that are factually wrong or fabricated) [110], the more common problem we encountered in our development process was in providing an extensive knowledge base to address highly specific questions within a clinical domain. In our experience, off-the-shelf LLMs are like dilettantes: they possess broad knowledge but lack depth in a particular subject. From this project, we glean that this characteristic holds particularly true when the subject matter requires clinical expertise and familiarity with scientific literature within a specific clinical domain. Overcoming this challenge required multiple iterations to build a knowledge base grounded in empirically supported best practices for smoking cessation. The responses needed to be accurate and clinically sensitive, suggesting a similar knowledge-building process will be essential for developing chatbots in any other clinical domain.

Conclusions

The development process yielded a comprehensive, fully developed, quit smoking program delivered through a conversational chatbot. Iterative testing led to improvements in the delivery platform, and a core LLM-supported conversational feature was integrated, enabling users to

pose open-ended questions about quitting cigarette smoking. Our next step is testing QuitBot's efficacy for smoking cessation in a full-scale randomized controlled trial.

Acknowledgements

We gratefully acknowledge the National Cancer Institute for providing their Call Center transcripts; James Jacoby and Tara Ward at Moby Inc. for providing QuitBot development expertise and technical project management; Jessica Harris, Sarah Hwang, Julie Packard, Victoria Sanborn, and Christeine Terry in the Fred Hutch HABIT Lab for supporting the QnA development; Daniella Kim at Halibut Flats for leading the user testing; Andrew Pham and Adam Rhine for their design and programming expertise in the early phase of developing QuitBot; the research staff at OpenAI for providing access to and technical support for fine tuning GPT; the research staff, including Anusua Trivedi, Marci Strong, and Thomas Roca at Microsoft's AI for Good for providing technical expertise in LLMs. An interview about QuitBot with the lead author and Trevor Noah is available here: <https://www.youtube.com/watch?v=6bTQfWsWC7g>

References

1. Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nat Rev Cancer*. 2009;9(9):655-6410.1038/nrc2703.
2. Safiri S, Nejadghaderi SA, Abdollahi M, Carson-Chahhoud K, Kaufman JS, Bragazzi NL, et al. Global, regional, and national burden of cancers attributable to tobacco smoking in 204 countries and territories, 1990-2019. *Cancer Med*. 2022;11(13):2662-78.PMC9249976. 10.1002/cam4.4647.
3. U.S. Department of Health and Human Services. Smoking Cessation. A Report of the Surgeon General. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Atlanta, GA; 2020.
4. Patnode CD, Henderson JT, Coppola EL, Melnikow J, Durbin S, Thomas RG. Interventions for Tobacco Cessation in Adults, Including Pregnant Persons: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2021;325(3):280-9810.1001/jama.2020.23541.
5. Littman D, Sherman SE, Troxel AB, Stevens ER. Behavioral Economics and Tobacco Control: Current Practices and Future Opportunities. *Int J Environ Res Public Health*. 2022;19(13).PMC9266334. 10.3390/ijerph19138174.
6. Choi K, Jones JT, Ruybal AL, McNeel TS, Duarte DA, Webb Hooper M. Trends in Education-Related Smoking Disparities Among U.S. Black or African American and White Adults: Intersections of Race, Sex, and Region. *Nicotine Tob Res*. 2023;25(4):718-28.PMC10032197. 10.1093/ntr/ntac238.
7. Ho JY, Elo IT. The contribution of smoking to black-white differences in U.S. mortality. *Demography*. 2013;50(2):545-68.PMC3606656. 10.1007/s13524-012-0159-Z.
8. Kcomt L, Evans-Polce RJ, Engstrom CW, West BT, McCabe SE. Discrimination, Sexual Orientation Discrimination, and Severity of Tobacco Use Disorder in the

- United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions-III. *Nicotine Tob Res.* 2021;23(6):920-30.PMC8150127. 10.1093/ntr/ntaa197.
9. Arrazola RA, Griffin T, Lunsford NB, Kittner D, Bammeke P, Courtney-Long EA, et al. US Cigarette Smoking Disparities by Race and Ethnicity - Keep Going and Going! *Prev Chronic Dis.* 2023;20:E45.PMC10240929. 10.5888/pcd20.220375.
 10. Girvalaki C, Mechili EA, Papadakis S, Nikitara K, Demin A, Trofor A, et al. Current practices and perceived barriers to tobacco-treatment delivery among healthcare professionals from 15 European countries. The EPACTT Plus project. *Tob Prev Cessat.* 2020;6:6.PMC7291907. 10.18332/tpc/115033.
 11. Sakka S, Al-Shatanawi TN, Bataineh DZ, Haddad W, Al Tamimi S, Al Salamat H, et al. Knowledge, attitude, practice and perceived barriers towards smoking cessation services among community pharmacists. *Pharm Pract (Granada).* 2022;20(1):2637.PMC9014907. 10.18549/PharmPract.2022.1.2637.
 12. Julius RJ, Novitsky MA, Jr., Dubin WR. Medication adherence: a review of the literature and implications for clinical practice. *J Psychiatr Pract.* 2009;15(1):34-4410.1097/01.pra.0000344917.43780.77.
 13. Twyman L, Bonevski B, Paul C, Bryant J. Perceived barriers to smoking cessation in selected vulnerable groups: a systematic review of the qualitative and quantitative literature. *BMJ Open.* 2014;4(12):e006414.PMC4275698. 10.1136/bmjopen-2014-006414.
 14. Pipe AL, Evans W, Papadakis S. Smoking cessation: health system challenges and opportunities. *Tob Control.* 2022;31(2):340-710.1136/tobaccocontrol-2021-056575.
 15. Center fo Disease Control and Prevention. Smoking Cessation—The Role of Healthcare Professionals and Health Systems. Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion: Center fo Disease Control and Prevention; 2020.

16. **World Health Organization. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.; 2023.**
17. **Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults — United States, 2000–2015. MMWR Morb Mortal Wkly Rep. 2017.**
18. **Horvath AO, Greenberg LS. The Working Alliance: Theory, Research, and Practice. New York, NY: John Wiley & Sons, Inc.; 1994.**
19. **Meier PS, Barrowclough C, Donmall MC. The role of the therapeutic alliance in the treatment of substance misuse: a critical review of the literature. Addiction. 2005;100(3):304-1610.1111/j.1360-0443.2004.00935.x.**
20. **Becker MH, Rosenstock IM. Compliance with medical advice. In: A. Steptoe and A. Matthews (Ed.), Health care and human behavior, pp. 135-152. London, UK: Academic Press.1984.**
21. **Boardman T, Catley D, Grobe JE, Little TD, Ahluwalia JS. Using motivational interviewing with smokers: do therapist behaviors relate to engagement and therapeutic alliance? J Subst Abuse Treat. 2006;31(4):329-3910.1016/j.jsat.2006.05.006.**
22. **Joe GW, Simpson DD, Dansereau DF, Rowan-Szal GA. Relationships between counseling rapport and drug abuse treatment outcomes. Psychiatr Serv. 2001;52(9):1223-910.1176/appi.ps.52.9.1223.**
23. **Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns. 2017;100(9):1720-9.PMC5559098. 10.1016/j.pec.2017.04.015.**
24. **Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res. 2017;19(5):e151.PMC5442350. 10.2196/jmir.6553.**

25. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. *J Med Internet Res*. 2016;18(1):e1.PMC4717285. 10.2196/jmir.5239.
26. King AC, Campero I, Sheats JL, Castro Sweet CM, Garcia D, Chazaro A, et al. Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemp Clin Trials*. 2017;61:115-25.PMC5987528. 10.1016/j.cct.2017.07.020.
27. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med*. 2023;183(6):589-96.PMC10148230. 10.1001/jamainternmed.2023.1838.
28. Whittaker R, Dobson R, Garner K. Chatbots for Smoking Cessation: Scoping Review. *J Med Internet Res*. 2022;24(9):e35556.PMC9514452. 10.2196/35556.
29. Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: A systematic review and meta-analysis. *Digit Health*. 2023;9:20552076231211634.PMC10623979. 10.1177/20552076231211634.
30. World Health Organization. Using AI to quit tobacco.2021.[Available from: <https://www.who.int/news-room/spotlight/using-ai-to-quit-tobacco>.].
31. Solutions4Health. Quit with Bella.2018.[Available from: <https://www.quitwithbella.com/>.].
32. Alex Therapeutics AB. Alex.2021.[Available from: <https://www.alexquitsmoking.com/>.].
33. Loveys K, Antoni M, Donkin L, Sagar M, Broadbent E. Comparing the Feasibility and Acceptability of a Virtual Human, Teletherapy, and an e-Manual in Delivering a Stress Management Intervention to Distressed Adult Women: Pilot Study. *JMIR*

Form Res. 2023;7:e42390.PMC9951078. 10.2196/42390.

34. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. *Comput Inform Nurs.* 2009;27(3):175-83.PMC2818536. 10.1097/NCN.0b013e31819f7c7c.
35. Bendotti H, Ireland D, Lawler S, Oates D, Gartner C, Marshall H. Introducing Quin: The design and development of a prototype chatbot to support smoking cessation. *Nicotine Tob Res.* 2023;10.1093/ntr/ntad217.
36. Almusharraf F, Rose J, Selby P. Engaging Unmotivated Smokers to Move Toward Quitting: Design of Motivational Interviewing-Based Chatbot Through Iterative Interactions. *J Med Internet Res.* 2020;22(11):e20251.PMC7671850. 10.2196/20251.
37. International Organization for Standardization. Ergonomics of human-system interaction -- Part 210: Human-centered design for interactive systems.2017. [Available from: <https://www-iso-org.offcampus.lib.washington.edu/standard/52075.html>].
38. International Organization for Standardization. Human-centred design processes for interactive systems Internet.2017.[Available from: <https://www-iso-org.offcampus.lib.washington.edu/standard/21197.html>].
39. Horvath AO, & Greenberg, L.S. *The Working Alliance: Theory, Research, and Practice.* New York: John Wiley & Sons; 1994.
40. Warlick C, Richter KP, Catley D, Gajewski BJ, Martin LE, Mussulman LM. Two brief valid measures of therapeutic alliance in counseling for tobacco dependence. *J Subst Abuse Treat.* 2018;86:60-4.PMC5810957. 10.1016/j.jsat.2017.12.010.
41. Klemperer EM, Hughes JR, Callas PW, Solomon LJ. Working alliance and empathy as mediators of brief telephone counseling for cigarette smokers who are not ready to quit. *Psychol Addict Behav.* 2017;31(1):130-5.PMC5302829. 10.1037/adb0000243.
42. Havens L. *Making Contact: Uses of language in psychotherapy.* Cambridge: Harvard University Press; 1986.

43. Okun B. **Effective helping: interviewing and counseling techniques.** . Pacific Grove, CA: Brooks/Cole; 1997.
 44. Laver J. **Communicative functions of phatic communion** In: Kendon A, Harris, R., Key, M. , editor. **The organization of behavior in face-to-face interaction** The Hague: Mouton; 1975. p. 215-38.
 45. Dainton M S, L. . **Routine maintenance behaviors: A comparison of relationships type, partner similarity and sex differences.** *Journal of Personal and Social Relations.* 1993;10:255-71
 46. Stafford LC, D. . **Maintenance strategies and romantic relationship type, gender, and relational characteristics** *Journal of Social and Personal Relations.* 1991;8:217-42.
 47. Brown P, & Levinson, SC. **Politeness: some universals in language usage.** Cambridge: Cambridge University Press; 1987.
 48. Laver J. **Linguistic routines and politeness in greeting and parting.** . In: Coulmas F, editor. **Conversational routine.** The Hague: Mouton; 1981. p. 289-304.
 49. Bickmore T, Gruber A, Picard R. **Establishing the computer-patient working alliance in automated health behavior change interventions.** *Patient education and counseling.* 2005;59(1):21-3010.1016/j.pec.2004.09.008.
 50. Archer RL. **Self-disclosure** In: Wegner DM, & Vallacher, R.R., editor. **The Self in Social Psychology** New York: Oxford University Press; 1980. p. 183-205.
 51. Shenk CE, & Fruzzetti, A.E. **The impact of validating and invalidating responses on emotional reactivity** *Journal of Social and Clinical Psychology* 2011;30:163-83
 52. Ho A, Hancock J, Miner AS. **Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot.** *J Commun.* 2018;68(4):712-33.PMC6074615. 10.1093/joc/jqy026.
 53. Fiore M, Jaén CR, Baker TB, Bailey WC, Bennett G, Benowitz NL, et al. **A clinical**
- 43

- practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. *Am J Prev Med.* 2008;35(2):158-76.PMC4465757. 10.1016/j.amepre.2008.04.009.
54. Bricker JB, Sullivan BM, Mull KE, Torres AJ, Carpenter KM. Full-scale Randomized Trial Comparing Acceptance and Commitment Therapy Telephone-Delivered Coaching With Standard Telephone-Delivered Coaching Among Medicare/Uninsured Quitline Callers. *Nicotine Tob Res.* 2022;24(10):1556-66.PMC9575982. 10.1093/ntr/ntac052.
55. Microsoft. What is Language Understanding (LUIS)?2024.[Available from: <https://learn.microsoft.com/en-us/azure/ai-services/luis/what-is-luis>].
56. Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C. Smoking Cessation Smartphone App Use Over Time: Predicting 12-Month Cessation Outcomes in a 2-Arm Randomized Trial. *J Med Internet Res.* 2022;24(8):e3920810.2196/39208.
57. Rogers Y, Sharp H, Preece J. *Interaction Design: Beyond Human-Computer Interaction.* West Sussex: John Wiley & Sons; 2011.
58. IDEO. *Human-Centered Design Toolkit: An Open-Source Toolkit To Inspire New Solutions in the Developing World.* 2nd ed. Canada2011.
59. Braun V., V. C. Using thematic analysis in psychology. *Qual Res Psychol.* 2006;3(2):77–10110.1191/1478088706qp063oa.
60. crunchbase. Botmock.[Available from: <https://www.crunchbase.com/organization/botmock>].
61. Chatfuel. Chatfuel.2023.[Available from: <https://chatfuel.com/>].
62. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, et al. Treating tobacco use and dependence: 2008 update. *Clinical Practice Guideline.* Rockville, MD: U.S. Department of Health and Human Services, Public Health Service; 2008.

63. Kraus R. Messenger is for bots. Mashablecom. 2018.
64. Statista. Facebook Messenger usage reach in the United States 2019.[Available from: <https://www.statista.com/download/MTU2NDE4MzcwNSMjNzcwODQzIyM1MjAwNCMjMSMjcGRmIyNTdHVkeQ==>].
65. Statista. Acceptance of artificial intelligence chatbots by customers worldwide, as of 2017. by service.2017.[Available from: <https://www.statista.com/statistics/717098/worldwide-customer-chatbot-acceptance-by-industry/>].
66. Node.js® is an open-source, cross-platform JavaScript runtime environment. [Internet]. OpenJS Foundation. 2024 [cited February 12, 2024]. Available from: <https://nodejs.org/en>.
67. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?:An Experiment with Data Saturation and Variability. *Field Methods*. 2006;18(1):59-8210.1177/1525822x05279903.
68. Daae J, Boks C. A classification of user research methods for design for sustainable behaviour. *Journal of Cleaner Production*. 2015;106:680-9<https://doi.org/10.1016/j.jclepro.2014.04.056>.
69. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the Technology Acceptance Model to Explore User Experience, Intent to Use, and Use Behavior of a Patient Portal Among Older Adults With Multiple Chronic Conditions: Descriptive Qualitative Study. *J Med Internet Res*. 2019;21(4):e11604.PMC6475817. 10.2196/11604.
70. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res*. 2011;45(5):626-9.PMC3081994. 10.1016/j.jpsychires.2010.10.008.
71. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: Getting started and moving on from stage I. *Clinical Psychology: Science*

and Practice. 2001;8(2)10.1093/clipsy/8.2.133.

72. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. *Nicotine Tob Res.* 2013;15(10):1756-64.PMC3768336. 10.1093/ntr/ntt056.
73. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. *Drug Alcohol Depend.* 2014;143:87-94.PMC4201179. 10.1016/j.drugalcdep.2014.07.006.
74. Boland VC, Stockings EA, Mattick RP, McRobbie H, Brown J, Courtney RJ. The Methodological Quality and Effectiveness of Technology-Based Smoking Cessation Interventions for Disadvantaged Groups: A Systematic Review and Meta-analysis. *Nicotine Tob Res.* 2018;20(3):276-8510.1093/ntr/ntw391.
75. Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. *Health Technol Assess.* 2012;16(38):1-205, iii-v10.3310/hta16380.
76. Graham AL, Amato MS. Twelve Million Smokers Look Online for Smoking Cessation Help Annually: Health Information National Trends Survey Data, 2005-2017. *Nicotine Tob Res.* 2019;21(2):249-52.PMC6319446. 10.1093/ntr/nty043.
77. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. *Annu Rev Public Health.* 2015;36:393-415.PMC4406229. 10.1146/annurev-publhealth-031914-122855.
78. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tob Control.* 2005;14(4):255-61.PMC1748056. 10.1136/tc.2005.011577.
79. Vercammen L. Estimated 2017 US reach of SMS text messaging programs for smoking cessation ICF International 2018.

80. Cheung YT, Chan CH, Lai CK, Chan WF, Wang MP, Li HC, et al. Using WhatsApp and Facebook Online Social Groups for Smoking Relapse Prevention for Recent Quitters: A Pilot Pragmatic Cluster Randomized Controlled Trial. *J Med Internet Res.* 2015;17(10):e238.PMC4642789. 10.2196/jmir.4829.
81. Durmaz S, Ergin I, Durusoy R, Hassoy H, Caliskan A, Okyay P. WhatsApp embedded in routine service delivery for smoking cessation: effects on abstinence rates in a randomized controlled study. *BMC Public Health.* 2019;19(1):387.PMC6454636. 10.1186/s12889-019-6727-z.
82. Heinzen E SJ, Atkinson E, Gunderson T, Dougherty G. *_arsenal: An Arsenal of 'R' Functions for Large-Scale Statistical Summaries_*. 3.6.3 ed2021.
83. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Accessed April 16, 2024. 2023
84. Venables WN, Ripley BD. *Modern Applied Statistics with S*. 4th ed. Springer. <https://doi.org/10.1007/978-0-387-21706-22002>.
85. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev.* 2016;4:CD00661110.1002/14651858.CD006611.pub4.
86. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics.* 1979;6:65–70
87. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. *J Med Internet Res.* 2018;20(8):e10351.PMC6128955. 10.2196/10351.
88. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text Messaging-Based Interventions for Smoking Cessation: A

- Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth. 2016;4(2):e49.PMC4893152. 10.2196/mhealth.5436.
89. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev. 2017;9:CD007078.PMC6703145. 10.1002/14651858.CD007078.pub5.
90. Counseling and Psychotherapy Transcripts: Volume I | Alexander Street, part of Clarivate [Internet]. [cited November 21, 2023]. Available from: <https://search.alexanderstreet.com/psyc>.
91. Get live, online assistance from NCI's Cancer Information Service [Internet]. National Cancer Institute. [cited November 21, 2023]. Available from: livehelp.cancer.gov.
92. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction. 2018;113(5):914-23.PMC5930021. 10.1111/add.14127.
93. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of Smartphone Applications for Smoking Cessation: A Randomized Clinical Trial. JAMA Intern Med. 2020;180(11):1472-80.PMC7506605. 10.1001/jamainternmed.2020.4055.
94. Models Overview [Internet]. 2024 [cited February 12, 2024]. Available from: <https://platform.openai.com/docs/models>.
95. Wang J. Formal Methods in Computer Science. 1st ed. New York, NY: Chapman and Hall/CRC; 2019.
96. Alphonse A, Stewart K, Brown J, Perski O. Exploring Users' Experiences With a Quick-Response Chatbot Within a Popular Smoking Cessation Smartphone App: Semistructured Interview Study. JMIR Form Res. 2022;6(7):e36869.PMC9305398. 10.2196/36869.
97. Abdullah AS, Gaehde S, Bickmore T. A Tablet Based Embodied Conversational

- Agent to Promote Smoking Cessation among Veterans: A Feasibility Study. *J Epidemiol Glob Health*. 2018;8(3-4):225-30.PMC7377562. 10.2991/j.jegh.2018.08.104.
98. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano ME, Martinez-Suberviola FJ, et al. Effectiveness of a Conversational Chatbot (Dejal@bot) for the Adult Population to Quit Smoking: Pragmatic, Multicenter, Controlled, Randomized Clinical Trial in Primary Care. *JMIR Mhealth Uhealth*. 2022;10(6):e34273.PMC9274388. 10.2196/34273.
99. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. *Addict Behav*. 2016;62:47-53.PMC5144826. 10.1016/j.addbeh.2016.06.016.
100. Coa KI, Wiseman KP, Higgins B, Augustson E. Associations Between Engagement and Outcomes in the SmokefreeTXT Program: A Growth Mixture Modeling Analysis. *Nicotine Tob Res*. 2019;21(5):663-9.PMC6608548. 10.1093/ntr/nty073.
101. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev*. 2012;11:CD006611.10.1002/14651858.CD006611.pub3.
102. Pineiro B, Wetter DW, Vidrine DJ, Hoover DS, Frank-Pearce SG, Nguyen N, et al. Quitline treatment dose predicts cessation outcomes among safety net patients linked with treatment via Ask-Advise-Connect. *Prev Med Rep*. 2019;13:262-7.PMC6351387. 10.1016/j.pmedr.2019.01.009.
103. Matkin W, Ordonez-Mena JM, Hartmann-Boyce J. Telephone counselling for smoking cessation. *Cochrane Database Syst Rev*. 2019;5(5):CD002850.PMC6496404. 10.1002/14651858.CD002850.pub4.
104. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. *Am J Drug Alcohol Abuse*. 2016;42(4):450-810.3109/00952990.2016.1149591.

105. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: Adherence to a smartphone app for smoking cessation. *Addict Behav.* 2016;63:120-4.PMC5501958. 10.1016/j.addbeh.2016.07.007.
106. Toloka Team. The history, timeline, and future of LLMs. Lucerne, Switzerland.2023. [Available from: <https://toloka.ai/blog/history-of-llms/>].
107. Karjian R. Tech Accelerator. A guide to artificail intelligence in the enterprise. The history of artificial intelligence: Complete AI timeline. TechTarget. 2023.
108. Bushwick S. What the New GPT-4 AI Can Do. *Scientific American.* 2023.
109. Roose K. GPT-4 Is Exciting and Scary. *The New York Times.* 2023 March 15, 2023.
110. Metz C, Collins K. 10 Ways GPT-4 Is Impressive but Still Flawed. *The New York Times.* 2023 March 14, 2023.

Supplementary Files

Multimedia Appendixes

Representative Screenshots of QuitBot.

URL: <http://asset.jmir.pub/assets/3d7f4a6b30eff9952b69a5ef15fdff35.docx>

CONSORT (or other) checklists

CONSORT eHealth checklist.

URL: <http://asset.jmir.pub/assets/0e37c1ba374d55d439e402d96e88204f.pdf>