

Generative AI chatbot vs telephone counseling for youth tobacco cessation: a pragmatic pilot randomized trial

Tzu Tsun Luk

luktt@nus.edu.sg

National University of Singapore <https://orcid.org/0000-0002-7056-1976>

Xiaoyun Xie

The University of Hong Kong <https://orcid.org/0000-0003-3189-5583>

Keith Ching Ho Kwan

The University of Hong Kong

Annie Yip

The University of Hong Kong <https://orcid.org/0000-0003-4832-8286>

Joshua Ho

The University of Hong Kong <https://orcid.org/0000-0003-2331-7011>

Janet Yuen-Ha Wong

Hong Kong Metropolitan University

Sai Yin Ho

The University of Hong Kong

Tai Hing Lam

The University of Hong Kong <https://orcid.org/0000-0002-2033-9971>

Man Ping Wang

School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, SAR China <https://orcid.org/0000-0003-4000-2388>

Yee Tak Derek Cheung

The University of Hong Kong

Article

Keywords:

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8048239/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Generative artificial intelligence (GenAI) chatbots offer a promising and scalable intervention for youth tobacco cessation, but evidence from randomized trials is lacking. This pilot pragmatic trial evaluated the feasibility of implementing a GenAI chatbot in a youth quitline and compared tobacco abstinence outcomes with standard peer-led telephone counseling. The intervention group received a web-based GenAI chatbot prompt-engineered to deliver tobacco cessation counseling using the 5A's, 5R's, and motivational interviewing. The control group received standard telephone counseling, delivered by trained peer counselors using the same counseling framework. Both groups received follow-up calls at 3 months and 6 months. The primary feasibility outcomes were recruitment and retention rates, while the primary treatment outcome was biochemically validated tobacco abstinence at 6 months, defined by a negative salivary cotinine test. Of 164 eligible youths identified, 154 (94%; 95% CI, 90%-97%) were enrolled and randomized into the intervention (n=77) or control (n=77) group. At 6 months, 125 participants completed follow-up (81%; 95% CI, 75%-96%). In the intervention group, 55 participants (71%) used the chatbot at least once and 48 (62%) completed baseline counseling session. The median System Usability scale score was 68 out of 100. Biochemically validated abstinence was 23% in the intervention group vs 17% in the control group (relative risk, 1.38; 95% CI, 0.73-2.63; P=0.32). Overall treatment satisfaction (0-10) was similar between the intervention and control groups (8.0 vs 7.5; P=0.43). This trial first showed the feasibility of implementing a GenAI chatbot in a youth quitline to support cessation. ClinicalTrials.gov: NCT06470776.

Introduction

Most individuals begin tobacco use before age 25,¹ and early initiation is associated with lower cessation success² and increased premature mortality.³ However, youths differ from adults in smoking behaviors and quit determinants,⁴ and most effective treatments for adults have shown limited benefit for youths.⁵ Randomized controlled trials (RCTs) of varenicline, the most effective cessation intervention for adults, showed no clear benefit in adolescents.^{6,7} The US Preventive Service Task Force has called for more research to identify effective behavioral interventions for youths who use tobacco.⁸

Recent advances in chatbot technologies—computer programs that can simulate human-like interactions—have offered a potential platform to provide scalable and accessible tobacco cessation support. Recent reviews have shown high acceptability of cessation chatbots,^{9,10} and a recent meta-analysis of 3 RCTs reported limited but promising effect on tobacco cessation (relative risk [RR], 1.29; 95% CI, 1.13–1.46).¹¹ However, most existing chatbots were rule-based, relying on predefined flow to guide the conversation, or retrieval-based, using artificial intelligence (AI) to predict users' intent and reply with predetermined responses. These systems could not handle complex or unexpected user inputs, leading to unnatural or repetitive responses that may reduce engagement and effectiveness.¹²

The release of ChatGPT in late 2022 marked a major advancement in generative AI (GenAI), introducing chatbots powered by large language models (LLMs) that can generate original, contextually relevant responses. Unlike earlier rule-based or scripted chatbots, GenAI chatbots offer greater adaptability and conversational depth, potentially improving engagement. However, this flexibility also introduces risks of non-factual or clinically inappropriate responses (i.e., hallucination), raising important safety considerations. Emerging studies have found that ChatGPT's responses to inquiries related to tobacco and vaping were generally satisfactory and adhered to established clinical guidelines.^{13–15} Nevertheless, our PubMed search using the search terms ("Smoking"[Mesh] OR smok* OR tobacco OR cigar*) AND ("Generative Artificial Intelligence"[Mesh] OR chat*) did not find any completed RCT on GenAI chatbots for tobacco cessation.

Established in 2005, the University of Hong Kong (HKU) Youth Quitline was the first tobacco cessation service targeting youths in Hong Kong.¹⁶ The service offers multi-session telephone counseling delivered by peer counselors, achieving 23.6% self-reported and 9.6% biochemically validated abstinence at 6-month post-enrollment.^{17,18} However, the service's capacity and operating hours are constrained by its reliance on peer counselor. If shown to be non-inferior to telephone counseling in supporting cessation, GenAI chatbots could offer a scalable alternative to increase service access and uptake. To inform future definitive trials, this pilot trial evaluated the feasibility of implementing a GenAI chatbot within the youth quitline and compared biochemically validated tobacco abstinence rates with standard peer-led telephone counseling.

Methods

Study design and participants

This study was a 2-arm, parallel pragmatic pilot RCT conducted within the HKU Youth Quitline with 6-month follow-up. The trial was predominantly pragmatic across key PRECIS-2 domains.¹⁹ The only prominent deviations from routine practice were the addition of usability items in the follow-up survey and additional training for peer counselors to implement the chatbot support. The protocol was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (number, UW 05-185). To safeguard youth privacy and confidentiality, parental consent was exempted. The reporting followed the Consolidated Standards of Reporting Trials (CONSORT) guideline and the Development, Evaluation, and Assessment of Large Language Models (DEAL) Checklist.²⁰

From July 7 to November 23, 2024, participants were recruited through 3 existing channels of the HKU Youth Quitline: (1) incoming calls and referrals from teachers, parents, or friends; (2) online platforms, including the HKU Youth Quitline website and social networking sites (WhatsApp, Facebook page, Instagram); and (3) outreach in schools and community settings. Details on the recruitment method have been reported elsewhere.¹⁸ All youths enrolled in the HKU Youth Quitline service were screened for eligibility and invited to participate.

The eligibility criteria were identical to those of the Quitline's target users. Participants were Hong Kong residents aged 25 years or younger who had used any tobacco product, including cigarettes, electronic cigarettes, heated tobacco products, in the past 30 days; could communicate in Chinese; and owned a smartphone with internet access. Those who reported a diagnosis of a severe mental disorder requiring ongoing psychotropic medication or were participating in other tobacco cessation programs were excluded. This criterion was applied for participant safety, as the chatbot was not designed to manage co-occurring severe mental health conditions.

Randomization and blinding

A researcher not involved in participant recruitment used an online tool to generate the 1:1 randomization scheme, with random permuted blocks sizes of 2, 4, and 6 to achieve a similar number of participants in both groups. For allocation concealment, the researcher prepared sequentially numbered, opaque sealed, temper-evident envelopes, which were stored centrally at the call center. After obtaining consent from the participants, an envelope was opened by the peer counselor to determine the assignment. Due to the behavioral nature of the intervention, blinding of the participants and the researchers was not possible.

Procedures

All youths enrolled in the HKU Youth Quitline service were screened for eligibility via phone call. Those who met the criteria completed a baseline assessment, including sociodemographic characteristics, pattern of tobacco use, nicotine dependence, and quitting behaviors. Nicotine dependence was assessed using the Fagerström Test for Nicotine Dependence, without adaptation for non-cigarette products given the predominance of cigarette use in the study population. Participants were then invited to participate in the study. Participants in the control group received standard telephone counseling on the same phone call, while those in the intervention group received a WhatsApp message with instructions and a link to access a web-based chatbot. Both groups received standard telephone follow-up calls from a counselor at 3 months and 6 months post-baseline.

Control

Participants in the control group received the standard treatment provided by the Quitline, i.e., telephone counseling delivered by trained peer counselors at baseline, 1-week and 1-month post-baseline. The counseling model followed the 5A's (Ask, Advise, Assess, Assist, Arrange), 5R's (Relevance, Risks, Rewards, Roadblocks, Repetition) and motivational interviewing techniques.²¹ Participants received up to two WhatsApp reminders from counselors for scheduling a follow-up session, followed by several phone call attempts. The peer counselors were predominantly university students from healthcare-related or social science disciplines, supervised by a master's level project manager with over 5 years of

experience in smoking cessation counseling. All peer counselors attended a 1-day structured training workshop and passed a written and practical examination prior to serving at the Quitline call center.¹⁷ All telephone conversations were tape-recorded and randomly selected for quality checking.

Intervention

Participants in the intervention group received one baseline counseling session and two follow-up sessions at 1 and 2 months, delivered via a web-based chatbot accessible on smartphones and computers. The frequency of sessions matched those provided in telephone counseling, and participants could access the chatbot at any time for on-demand support. No unsolicited contacts were made outside the scheduled sessions. Up to 2 WhatsApp reminders were sent by counselors to initiate each counseling session, with an additional 2 reminders to encourage completion if needed.

Participants accessed the chatbot through a designated website and required authentication using an assigned study ID. Before baseline session, they completed screening questions including study ID (to confirm identity), preferred name, daily cigarette consumption and time to first cigarette (to calculate nicotine dependence), and readiness to quit. The chatbot then delivered a counseling session tailored to the participant's readiness level, using the 5A's or 5R's and motivational interviewing. During the 1-month and 2-month follow-up sessions, the chatbot inquired about quitting status and provided additional support (e.g., relapse prevention advice, re-assessment of motivation to quit). Upon completion of each session, the chatbot sent a closing remark with the next follow-up schedule and instructions to obtain support from the chatbot or the counselor if needed. Supplement Data 1 provides a de-identified conversation illustrating the chatbot's baseline counseling flow.

Chatbot design and development

The chatbot was powered by the GPT-4o model via Azure OpenAI services, without model fine-tuning and retrieval-augmented generation (RAG) technology. The temperature parameter was set at 0.7 to allow for moderate response variability while maintaining clinical appropriateness.

The chatbot was instructed in the system prompt to act as a youth tobacco cessation counselor. The chatbot autonomously generated responses dynamically based on participants' free-text inputs and conversation context, asked follow-up questions as needed, while being constrained by a protocol-based system prompt and session-specific task instructions (Table S1). Prompt engineering was based on the HKU Youth Quitline counseling protocol, incorporating the 5A's and 5R's framework and motivational interviewing techniques. Grounding information, such as local cessation resources and Cantonese linguistic norms, was also embedded directly into the chatbot's prompt design to mitigate the risk of hallucination and improve cultural appropriateness and relevance to the Hong Kong context.

The chatbot employed a multi-agent framework, with dedicated agents handling different counseling scenarios and sessions (Fig. 1). Each agent was a distinct instance of the GPT-4o model, configured with

role-specific prompts that specified its objectives and actions. A summary of prompt components and agent task sequences is provided in the Supplementary Information (Table S1 and S2). All agent transitions were handled internally in the background, presenting the participant with a unified, continuous counseling interaction.

The prompt underwent iterative refinement through internal testing by the research team and 6 experienced tobacco cessation counselors. Counselors assessed the chatbot's outputs in simulated user scenarios and provided feedback to guide revisions. Three HKU Youth Quitline users participated in usability testing to assess functionality and counseling clarity. Model stochasticity was managed by reviewing response variations across similar questions. While some variability was observed, the responses remained aligned with counseling objectives and were deemed acceptable by the research team.

Multi-layered approaches were used to ensure user privacy and data security. First, the chatbot was hosted on a cloud server, with user data stored in Firebase. The model service was provided by a trusted AI platform that contractually committed not to use user data for improving or training the model. Second, participants were advised not to disclose any personally identifiable information while interacting with the chatbot. Third, the system infrastructure was compliant with the Health Insurance Portability and Accountability Act (HIPAA), incorporating data encryption, strict access controls, and secure storage protocols. The handling of personal data also complied with the data protection requirements under the Personal Data (Privacy) Ordinance in Hong Kong. Additionally, live counselors had access to a backend interface to monitor chatbot conversations and steer the chatbot's responses when necessary, such as when the chatbot provided inappropriate responses.

Outcomes

Trial feasibility

The primary feasibility outcomes were the recruitment rate (proportion of eligible youths who joined the study) and retention rate (proportion of participants who completed the follow-ups). Based on a systematic review of 172 smoking cessation trials²² and our previous experiences,¹⁷ a recruitment rate of $\geq 70\%$ and a retention rate of $\geq 70\%$ would be considered satisfactory to justify a full trial.

Chatbot use and acceptability

We assessed the number of completed counseling sessions by checking the conversation logs. Session completion was defined as the participant receiving the chatbot's closing remark, which was generated when the chatbot determined that all the protocol-specified topics for that session had been addressed. The number of intervention participants who requested the alternate treatment, i.e., crossed over to the control group was documented. Chatbot usability was measured by the 10-item System Usability Scale

(SUS),²³ which assessed the participants' perceived effectiveness, efficiency, and satisfaction with the chatbot on a score of 0 to 100. A score of 68 or above indicated satisfactory usability.²³

Treatment outcomes

Participant-reported data were collected at baseline, 3 months and 6 months through telephone, prior to any counseling. The primary treatment outcome was biochemically validated tobacco abstinence at 6 months after treatment initiation (3 months after the end of treatment).²⁴ Participants who reported having stopped smoking for 7 days or longer at 6 months were invited to test for salivary cotinine using the Alere iScreen OFD mouth swab, with a cutoff level of 30 ng/mL.²⁵ Consistent with the routine practice of the Youth Quitline, research staff conducted the test at a place convenient to the participant, with a small cash incentive of HK\$300 (≈ US\$38). Secondary outcomes included self-reported 7-day and 30-day point prevalence abstinence (PPA), 24-hour quit attempts, and smoking reduction, defined by at least 50% reduction of baseline consumption in participants who continued to smoke, at 3 and 6 months.

Participant satisfaction

At the 6-month follow-up, all participants rated their overall satisfaction with the cessation program using the following question on a scale of 0 (not at all likely) to 10 (very likely): “How likely are you to recommend the HKU Youth Quitline service to others who need help quitting?”

Statistical analyses

The planned sample size for this pilot was based on feasibility targets, aiming to inform the design of a future definitive trial where a non-inferiority margin will be prespecified. A simulation study²⁶ and our prior cessation trial in adults²⁷ showed that a sample size of about 100 participants would be sufficient to provide precise estimates of the trial feasibility outcomes. Assuming a dropout rate of 35% to account for the greater unpredictability of youth participation, a sample size of 154 (77 per group) was planned for this pilot trial.

All analyses were prespecified in the study protocol unless otherwise indicated. Descriptive statistics were used to report the trial feasibility and chatbot acceptability outcomes. The primary analysis for treatment outcomes was by intention-to-treat, assuming participants with missing outcome to be non-abstinent.²⁸ To avoid convergence issues common with log-binomial regression, Poisson regression with robust variance was used to compare the treatment outcomes between-group, with and without adjusting for baseline prognostic factors, including sex, age, nicotine dependence, and intention to quit. We also conducted as-treated analyses, analyzing participants based on the intervention actually received; per-protocol analyses, excluding participants who received the alternate treatment; and complete-case analyses, excluding participants with missing outcome data.

All statistical analyses were conducted in Stata/MP software (version 15.1; StataCorp). A 2-sided $P < 0.05$ denoted statistical significance.

Results

Trial feasibility

Among 164 eligible youths identified, 154 (94%; 95% CI, 90%-97%) were enrolled and randomized to either the intervention (n = 77) or control (n = 77) group (Fig. 2). Baseline characteristics were similar between participants and those who declined participation (Table S3 in Supplementary Information). The overall follow-up rates were 86% (95%CI, 80%-91%) at 3 months and 81% (95% CI, 75%-96%) at 6 months, with no significant between-group differences ($P \geq 0.40$).

Baseline characteristics

The mean (SD) age of the participants was 18.3 (3.4) years; 114 (74%) were male. Most participants (92%, n = 141) currently smoked cigarettes, while half (50%, n = 77) currently used e-cigarettes; 44% (n = 68) used both products. Baseline characteristics were similar between study groups (Table 1).

Table 1
Baseline characteristics of the participants (N = 154)

Age (years), mean (SD)	Intervention (N = 77)	Control (N = 77)
	18.3 (3.3)	18.3 (3.5)
Sex, n (%)		
Male	55 (71)	59 (77)
Female	22 (29)	18 (23)
Past 30-day cigarette use, n (%)		
Nil	5 (6)	8 (10)
Non-daily	19 (25)	17 (22)
Daily	53 (69)	52 (68)
Past 30-day e-cigarette use, n (%)		
Nil	45 (58)	32 (42)
Non-daily	10 (13)	14 (18)
Daily	22 (29)	31 (40)
Number of cigarettes per day, n (%)		
≤ 10	54 (70)	57 (74)
11–20	21 (27)	17 (22)
≥ 21	2 (3)	3 (4)
Time to first cigarette of the day, n (%)		
> 60 min	35 (45)	34 (44)
31–60 min	8 (10)	6 (8)
6–30 min	8 (10)	20 (26)
^a Measured by the Fagerström test for nicotine dependence. Scores range from 0–10, with higher scores indicating greater nicotine dependence.		
^b Scores range from 0–10, higher score indicates higher levels of perception.		
^c Measured by Patient Health Questionnaire-2, scores range from 0–6, with higher scores indicating more symptoms of depression. A score of ≥ 3 indicates risk of depressive disorder.		
^{ld} Measured by Generalized Anxiety Disorder-2, scores range from 0–6, with higher scores indicating more symptoms of anxiety. A score of ≥ 3 indicates risk of anxiety disorder.		

Age (years), mean (SD)	Intervention (N = 77)	Control (N = 77)
	18.3 (3.3)	18.3 (3.5)
≤ 5 min	26 (34)	17 (22)
Nicotine Dependence ^a , n (%)		
Mild (0–3)	49 (64)	47 (61)
Moderate (4–5)	19 (25)	18 (23)
Heavy (6–10)	9 (12)	12 (16)
Readiness to quit, n (%)		
Not decided yet	34 (44)	37 (48)
Within next 6 months	16 (21)	12 (16)
Within next month	10 (13)	7 (9)
Within next 7 days	17 (22)	21 (27)
Past quit attempts, n (%)		
No	17 (22)	21 (27)
Yes	60 (78)	56 (73)
Perceptions of quitting ^b , mean (SD)		
Importance	7.3 (2.2)	6.9 (2.5)
Confidence	6.2 (2.6)	6.3 (2.7)
Difficulty	5.8 (2.9)	6.3 (3.0)
Depression symptoms ^c , n (%)		
No	61 (79)	64 (83)
Yes	16 (21)	13 (17)
^a Measured by the Fagerström test for nicotine dependence. Scores range from 0–10, with higher scores indicating greater nicotine dependence.		
^b Scores range from 0–10, higher score indicates higher levels of perception.		
^c Measured by Patient Health Questionnaire-2, scores range from 0–6, with higher scores indicating more symptoms of depression. A score of ≥ 3 indicates risk of depressive disorder.		
^{ld} Measured by Generalized Anxiety Disorder-2, scores range from 0–6, with higher scores indicating more symptoms of anxiety. A score of ≥ 3 indicates risk of anxiety disorder.		

Age (years), mean (SD)	Intervention (N = 77)	Control (N = 77)
	18.3 (3.3)	18.3 (3.5)
Anxiety symptoms ^{ld} , n (%)		
No	60 (78)	65 (84)
Yes	17 (22)	12 (16)
^a Measured by the Fagerström test for nicotine dependence. Scores range from 0–10, with higher scores indicating greater nicotine dependence.		
^b Scores range from 0–10, higher score indicates higher levels of perception.		
^c Measured by Patient Health Questionnaire-2, scores range from 0–6, with higher scores indicating more symptoms of depression. A score of ≥ 3 indicates risk of depressive disorder.		
^{ld} Measured by Generalized Anxiety Disorder-2, scores range from 0–6, with higher scores indicating more symptoms of anxiety. A score of ≥ 3 indicates risk of anxiety disorder.		

Treatment outcomes

Of 53 participants who self-reported abstinence for 7 days or longer at 6-month follow-up, 32 (60%) participated in the face-to-face biochemical validation, and 31 (97%) of them passed the validation. The one participant who failed reported having relapsed prior to the test. Participation rates were similar between 2 groups (58% vs 59%; $P = 0.94$). Table 2 shows the primary treatment outcome of biochemically validated tobacco abstinence at 6 months was 23% (18 of 77) in the intervention group vs 17% (13 of 77) in the control group (RR 1.38; 95% CI, 0.73–2.63; $P = 0.32$). Likewise, secondary outcomes of self-reported 7-day PPA, 30-day PPA, and 24-hour quit attempt were greater in the intervention group at both follow-ups except 30-day PPA at 3 months. Among youths who continued to smoke, the smoking reduction rates were lower in the intervention group. These results were consistent in multivariable analyses. As-treated, per-protocol, and complete case analyses also yielded similar point estimates of intervention effect (Table S4 in Supplementary Information).

Table 2
Primary and secondary outcomes in intervention and control group

	Intervention group (N = 77)	Control group (N = 77)	Crude RR (95% CI)	P value	Adjusted RR (95% CI)^a	P value
Biochemically validated abstinence						
6 months	18 (23)	13 (17)	1.38 (0.73–2.63)	0.32	1.27 (0.68–2.38)	0.45
Self-reported 7-day point prevalent abstinence						
3 months	24 (31)	20 (26)	1.20 (0.72–1.99)	0.48	1.16 (0.71–1.92)	0.55
6 months	31 (40)	22 (29)	1.41 (0.90–2.20)	0.13	1.34 (0.88–2.04)	0.17
Self-reported 30-day point prevalent abstinence						
3 months	15 (19)	16 (21)	0.94 (0.50–1.76)	0.84	0.89 (0.48–1.66)	0.71
6 months	23 (30)	14 (18)	1.64 (0.91–2.95)	0.10	1.58 (0.90–2.76)	0.11
24-hour quit attempt						
3 months	49 (64)	42 (55)	1.17 (0.89–1.52)	0.26	1.15 (0.89–1.48)	0.29
6 months (cumulative)	59 (77)	54 (70)	1.09 (0.90–1.32)	0.37	1.07 (0.90–1.28)	0.45
Smoking reduction by at least 50% of baseline^b						
3 months	19/49 (39)	28/55 (51)	0.76 (0.49–1.18)	0.22	0.74 (0.48–1.14)	0.17
RR = relative risk.						
^a Adjusted for sex, age, nicotine dependence, and intention to quit at baseline.						
^b Excluding participants who did not smoke cigarette but smoked other tobacco products at baseline and those reported abstinence from denominators.						

	Intervention group (N = 77)	Control group (N = 77)	Crude RR (95% CI)	P value	Adjusted RR (95% CI) ^a	P value
Biochemically validated abstinence						
6 months	4/43 (9)	5/50 (10)	0.93 (0.26–3.27)	0.91	0.90 (0.28–2.89)	0.86
RR = relative risk.						
^a Adjusted for sex, age, nicotine dependence, and intention to quit at baseline.						
^b Excluding participants who did not smoke cigarette but smoked other tobacco products at baseline and those reported abstinence from denominators.						

Chatbot use and acceptability

In the intervention group, 55 of 77 participants (71%) used the chatbot at least once, with a median (IQR) of 26 (15–48) conversation turns. Forty-eight participants (62%) completed the baseline counseling session and 13 (17%) completed all 3 counseling sessions (Table 3). In comparison, all participants in the control group completed the baseline session, and 64 (83%) completed all 3 sessions. Among the 15 intervention participants who provided reasons for not using the chatbot, 9 (60%) were “not interested,” 5 (33%) “think it is useless,” 4 (27%) reported “technical issue,” and 2 (13%) were “too busy.” None reported “privacy concern.” One participant explicitly requested human counseling and was crossed over to the control group. No major hallucinations or inappropriate responses requiring manual steering were identified throughout the trial.

Table 3
Chatbot usage in the intervention group (N = 77)

Session engagement	n (%)
Baseline counseling session	
Partially completed	7 (9)
Completed	48 (62)
Participant initiated on-demand contact ^a	4 (5)
Frequency of initiation, median (IQR)	2 (1-3.5)
Conversation turns ^b , median (IQR)	17 (13–30)
1-month counseling session	
Partially completed	7 (9)
Completed	19 (25)
Participant initiated on-demand contact ^a	3 (4)
Frequency of initiation, median (IQR)	1 (1–3)
Conversation turns ^b , median (IQR)	16 (6–28)
2-month counseling session	
Partially completed	2 (3)
Completed	15 (19)
Participant initiated on-demand contact ^a	0 (0)
Frequency of initiation, median (IQR)	NA
Conversation turns ^b , median (IQR)	7 (5–10)
All counseling sessions	
Completed at least one session	49 (64)
Completed all sessions	13 (17)
Not started	22 (29)
^a Defined as user sending message to the chatbot after completion of the counseling session (as indicated by the closing marker). If multiple messages were sent on the same day but pertain to the same topic, they were counted as a single instance.	
^b Excluding participants who did not use the chatbot.	

Session engagement	n (%)
Baseline counseling session	
Participant initiated on-demand contact ^a	5 (6)
Frequency of initiation, median (IQR)	2 (1–4)
Conversation turns ^b , median (IQR)	26 (15–48)
^a Defined as user sending message to the chatbot after completion of the counseling session (as indicated by the closing marker). If multiple messages were sent on the same day but pertain to the same topic, they were counted as a single instance.	
^b Excluding participants who did not use the chatbot.	

Of 55 participants who had used the chatbot, 41 (75%) responded to the SUS. The median (IQR) SUS score was 68 (48–78), reaching the satisfactory threshold. After reverse-coding the even-numbered items to account for their negative phrasing, the least rated item was item 1 “I think that I would like to use the chatbot frequently” with a median score of 2 over 5 (Table 4).

Table 4

System usability scale (SUS) scores for the chatbot in intervention participants (N = 41)^a

1. I think that I would like to use this chatbot frequently.	Median (IQR)
	2 (1–3)
2. I found the chatbot unnecessarily complex. ^b	4 (3–5)
3. I thought the chatbot was easy to use.	4 (3–5)
4. I think that I would need the support of a technical person to be able to use this chatbot. ^b	4 (4–5)
5. I found the various functions in this chatbot were well integrated.	3 (2–4)
6. I thought there was too much inconsistency in this chatbot. ^b	4 (3–5)
7. I would imagine that most people would learn to use this chatbot very quickly.	4 (3–4)
8. I found the chatbot very cumbersome to use. ^b	4 (2–5)
9. I felt very confident using the chatbot.	3 (3–4)
10. I needed to learn a lot of things before I could get going with this chatbot. ^b	4 (4–5)
Total score	68 (48–78)
^a All items were assessed on a scale of 1 (Strongly disagree) to 5 (Strongly agree).	
^b Reverse-coded for even items. Higher scores indicate more favorable perceptions of the chatbot usability.	

Participant satisfaction

At 6 months, the mean (SD) satisfaction score was 8.0 (2.7) in the chatbot group versus 7.5 (3.0) in the control group, with no significant difference ($P = 0.43$).

Discussion

To our knowledge, this was the first RCT evaluating a GenAI chatbot for youth tobacco cessation counseling. The high recruitment rate (93%) and retention rates at 3 months (end of intervention; 88%) and 6 months (81%) supported the feasibility of implementing the chatbot within an existing youth quitline service. The validated abstinence rate was higher in the chatbot group than in the telephone counseling group (23% vs 17%). The effect size (crude RR = 1.38) should be interpreted in light of the active, evidence-based cessation support in the control group, which has previously yielded a 9.6% validated abstinence rate at 6 months.¹⁸ However, the pilot was not powered for formal superiority or non-inferiority testing, and the confidence interval (95% CI 0.73–2.63) was imprecise. Other secondary

outcomes, including self-reported 7-day and 30-day PPA, smoking reduction and 24-hour quit attempt, also showed point estimates in the same direction. These preliminary findings warrant further investigation in a larger, adequately powered trial.

A few chatbots leveraging GenAI to support cessation have been developed, but none has been evaluated in an RCT. These included S.A.R.A.H by the World Health Organization and “BeFreeGPT”,¹⁴ both built on the GPT-4 model using prompt engineering approach and used RAG to mitigate hallucinations and misinformation. While both showed moderate adherence to clinical guidelines, they primarily operated in a question-and-answer format with brief responses and were not tailored to youth. A study evaluated a GenAI chatbot, “MIBot,” designed for motivational interviewing with 106 participants, who reported a mean increase of 1.7 points in confidence to quit on a 0–10 scale 1 week after use.²⁹ Our study contributes to the literature by evaluating a fully GenAI chatbot that delivers comprehensive tobacco cessation counseling (5A’s, 5R’s and motivational interviewing) using an RCT design with validated abstinence.

A pilot RCT of a rule-based chatbot, “QuitBot,” showed significantly higher self-reported abstinence among adults who completed the program.³⁰ While QuitBot’s scripted architecture engaged users through predictable, system-initiated conversations, our generative chatbot used probabilistic language generation to respond to open-ended user input. Whether this technical flexibility improves engagement through better conversational fit requires further investigation. Unlike QuitBot’s 4-year, 11-step user-centered design process, our accelerated timeline did not replicate this level of formative investment. Future iterations of GenAI chatbots could benefit from integrating elements of QuitBot’s development approach.

The moderate chatbot engagement rate in our trial aligns with the findings that most participants of digital health trials do not sustain use of the intervention, a phenomenon known as the “law of attrition.”³¹ Two large smoking cessation trials on chat-based support delivered by a live counselor showed that few participants (14.6%-30.6%) sustained engagement throughout a 3-month intervention period.^{32,33} A formative study of a ChatGPT-based cessation chatbot also reported that 16 out of 23 participants (70%) texted the bot at least once,³⁴ which is comparable to the 71% chatbot use observed in our study. Our chatbot was web-based to ensure compatibility by any device that has a web browser, but this may be less convenient and engaging compared with instant messaging apps (e.g., WhatsApp, Facebook Messenger, WeChat). Embedding the chatbot into widely used instant messaging apps may enhance accessibility and potentially improve engagement.

While the engagement pattern in the chatbot group was consistent with prior digital health studies, the all-session completion rate (17%) was considerably lower than that in the telephone counseling group (83%). Direct comparison is complicated by fundamental differences in how users engage with the modalities. Telephone calls were counselor-initiated and voice-based, while chatbot sessions required participants to actively log in to a web-based messaging system. Whether this differential completion translated into different therapeutic exposure cannot be ascertained. Higher completion does not

necessarily mean better outcomes; participants may disengage after receiving sufficient support.³⁵ Conversely, counselor-initiated calls may inflate completion rates without yielding superior therapeutic impact. These caveats warrant consideration when interpreting the abstinence outcomes.

The modest usability assessed by SUS also suggested room for improvement. However, this metric may not fully capture the value of a cessation tool. For instance, participants with low motivation to quit or those who had already quit may report low willingness to use the system frequently, regardless of its usability (SUS item 1). In a post-hoc exploratory analysis, the median SUS score increased to 72 after removing the item and rescaling the scale to 0–100. More targeted instruments, such as the Chatbot Usability Scale,³⁶ may better capture user experience with chatbot-based interventions and should be considered as primary measures in future work. Notably, despite moderate engagement and usability, participants in the chatbot group did not rate their willingness to recommend the Quitline service less favorably than those in the telephone counseling group. This suggests substituting some telephone counseling sessions with a chatbot was generally acceptable to participants.

Our trial suggests some youths may prefer live counselor support, as indicated by one participant's crossover from the chatbot group to the control group. Chatbots may appeal to youths who prefer texting over telephone communication or feel hesitant to disclose personal issues to another person. However, chatbots may miss verbal and non-verbal cues and lack the relational connection to establish therapeutic rapport. Our forthcoming post-intervention qualitative evaluation will provide additional insights into the participants' perceptions of our GenAI chatbot for cessation.

On safety monitoring, the absence of detected major hallucinations or inappropriate responses suggests that the chatbot operated with generally acceptable safety in this pilot context. Nevertheless, minor or subtle inaccuracies may have gone undetected. Our ongoing content analysis of chat logs will systematically evaluate message-level appropriateness to provide a more granular assessment.

Limitations

This trial had several limitations. First, as a pilot study, the sample size was intentionally small to assess feasibility and was underpowered to detect statistically significant differences in effectiveness. Second, participants were not blinded to group assignments, which might have influenced self-reported satisfaction and engagement. Third, participants in the intervention group were explicitly informed that the counseling was delivered by a GenAI chatbot, which might have shaped their expectations and perceptions, potentially leading to more critical evaluations or reduced emotional engagement. Fourth, counseling was delivered during the 3-month follow-up call per the service routine; the 6-month abstinence outcomes may not be attributed solely to the chatbot intervention. The first follow-up counseling was also scheduled at 1 and 2 months in the chatbot group versus 1 week and 1 month in the telephone group. This difference reflects the pragmatic adaptation to the asynchronous nature of chat-based interaction. These design considerations may limit direct comparison of the two modalities.

Conclusions

This trial showed the feasibility of implementing a fully GenAI-powered chatbot for youth tobacco cessation counseling in a youth quitline. Abstinence point estimates favored the chatbot when compared with standard peer telephone counseling but were imprecise and compatible with no difference. Fully powered randomized trials are warranted to confirm the chatbot's effectiveness.

Declarations

Contributions

T.T.L. conceptualized the study and was the principal investigator of the study. T.T.L., X.X., K.H.C.K., T.H.L., M.P.W., Y.T.D.C. contributed to the study design. X.X. and A.O.N.Y. contributed to participant recruitment and data acquisition. X.X. performed the statistical analysis. X.X. and T.T.L. wrote the first draft of the report. All authors interpreted the data and contributed to critical review of the manuscript for important intellectual content.

Competing interests

All authors declare no competing interests.

Acknowledgements

The research was funded by the Health and Medical Research Fund from the Health Bureau, The Hong Kong Special Administrative Region Government (grant no. 20212551). The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

We thank the participants for taking part and the youth counselors for implementing the study.

Data availability

The de-identified individual participant data generated during this study are available from the investigators on reasonable requests. Requests should be directed to the corresponding author by email.

References

1. Reitsma MB et al (2021) Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. *Lancet*

2. Pesce G et al (2019) Time and age trends in smoking cessation in Europe. PLoS ONE 14:e0211976
3. Choi SH, Stommel M (2017) Impact of Age at Smoking Initiation on Smoking-Related Morbidity and All-Cause Mortality. Am J Prev Med 53:33–41
4. United States Public Health Service Office of the Surgeon General & National Center for Chronic Disease Prevention and Health Promotion (US) (2020) Office on Smoking and Health. Patterns of Smoking Cessation Among U.S. Adults, Young Adults, and Youth. in *Smoking Cessation: A Report of the Surgeon General [Internet]*. US Department of Health and Human Services
5. Fanshawe TR et al (2017) Tobacco cessation interventions for young people. Cochrane Database Syst Rev 11:CD003289
6. Gray KM et al (2020) High-dose and low-dose varenicline for smoking cessation in adolescents: a randomised, placebo-controlled trial. Lancet Child Adolesc Health 4:837–845
7. Gray KM et al (2019) Efficacy and Safety of Varenicline for Adolescent Smoking Cessation: A Randomized Clinical Trial. JAMA Pediatr 173:1146–1153
8. US Preventive Services Task Force (2020) Primary Care Interventions for Prevention and Cessation of Tobacco Use in Children and Adolescents: US Preventive Services Task Force Recommendation Statement. JAMA 323:1590–1598
9. He L, Balaji D, Wiers RW, Antheunis ML, Krahmer E (2022) Effectiveness and acceptability of conversational agents for smoking cessation: a systematic review and meta-analysis. Nicotine Tob Res 25:1241–1250. 10.1093/ntr/ntac281
10. Whittaker R, Dobson R, Garner K (2022) Chatbots for Smoking Cessation: Scoping Review. J Med Internet Res 24:e35556
11. Bendotti H et al (2023) Conversational artificial intelligence interventions to support smoking cessation: A systematic review and meta-analysis. Digit Health 9:20552076231211634
12. Li H, Zhang R, Lee Y-C, Kraut RE, Mohr DC (2023) Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. Npj Digit Med 6:1–14
13. Amin S, Kawamoto CT, Pokhrel P (2023) Exploring the ChatGPT platform with scenario-specific prompts for vaping cessation. Tob Control 34:251–253. 10.1136/tc-2023-058009
14. Abrams LC, Yousefi A, Wysota CN, Wu T-C, Broniatowski DA (2025) Assessing the Adherence of ChatGPT Chatbots to Public Health Guidelines for Smoking Cessation: Content Analysis. J Med Internet Res 27:e66896
15. Xie XY et al (2025) ChatGPT-4's consistency, specificity, and inclusion of behavior change techniques in delivering smoking cessation advice in Traditional Chinese: A content analysis. Nicotine Tob Res ntaf267. 10.1093/ntr/ntaf267
16. Chan SSC et al (2008) The establishment and promotion of the first Youth Quitline in Hong Kong: challenges and opportunities. Eval Health Prof 31:258–271

17. Li WHC et al (2017) An Evaluation of the Youth Quitline Service Young Hong Kong Smokers. *J Adolesc* 60:584–591
18. Chen H et al (2024) Reaching Youth Smokers Through a Multipronged Approach: Comparison of Three Recruitment Methods of a Youth Quitline in Hong Kong. *J Adolesc Health* 75:883–889
19. Loudon K et al (2015) The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 350:h2147
20. Tripathi S et al (2025) Development, Evaluation, and Assessment of Large Language Models (DEAL) Checklist: A Technical Report. *NEJM AI* 2, Alp2401106
21. Tobacco Use and Dependence Guideline Panel (2008) Treating Tobacco Use and Dependence: 2008 Update. US Department of Health and Human Services
22. Bricca A et al (2022) Predictors of recruitment and retention in randomized controlled trials of behavioural smoking cessation interventions: a systematic review and meta-regression analysis. *Addict* 117:299–311
23. Brooke J (2013) SUS: a retrospective. *J Usability Stud* 8
24. Piper ME et al (2020) Defining and Measuring Abstinence in Clinical Trials of Smoking Cessation Interventions: An Updated Review. *Nicotine Tob Res* 22:1098–1106
25. Tourangeau R, Yan T, Sun H, Hyland A, Stanton CA (2019) Population Assessment of Tobacco and Health (PATH) reliability and validity study: selected reliability and validity estimates. *Tob Control* 28:663–668
26. Teare MD et al (2014) Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials* 15:264
27. Luk TT et al (2023) Mobile Chat Messaging for Preventing Smoking Relapse Amid the COVID-19 Pandemic: A Pilot Randomized Controlled Trial. *Nicotine Tob Res* 25:291–297
28. West R, Hajek P, Stead L, Stapleton J (2005) Outcome criteria in smoking cessation trials: proposal for a common standard. *Addict Abingdon Engl* 100:299–303
29. Mahmood Z et al (2025) Association for Computational Linguistics, Vienna, Austria,. A Fully Generative Motivational Interviewing Counsellor Chatbot for Moving Smokers Towards the Decision to Quit. in *Findings of the Association for Computational Linguistics: ACL 2025* 25008–25043 10.18653/v1/2025.findings-acl.1283
30. Bricker JB, Sullivan B, Mull K, Santiago-Torres M, Lavista Ferres JM (2024) Conversational Chatbot for Cigarette Smoking Cessation: Results From the 11-Step User-Centered Design Development Process and Randomized Controlled Trial. *JMIR MHealth UHealth* 12:e57318
31. Eysenbach G (2005) The law of attrition. *J Med Internet Res* 7:e11
32. Li Y et al (2024) Engagement With a Mobile Chat-Based Intervention for Smoking Cessation: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Netw Open* 7:e2417796–e2417796
33. Luk TT et al (2026) Mobile Chat Messaging for Smoking Relapse Prevention: A Randomized Clinical Trial. *JAMA Intern Med* 186:294–302

34. Abroms LC, Wysota CN, Yousefi A, Wu TC, Broniatowski DA (2025) ChatGPT-Based Chatbot for Help Quitting Smoking via Text Messaging: An Interventional Study. JMIR Form Res 9:e79402

35. Yardley L et al (2016) Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 51:833–842

36. Borsci S et al (2022) The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. Pers Ubiquitous Comput 26:95–119

Figures

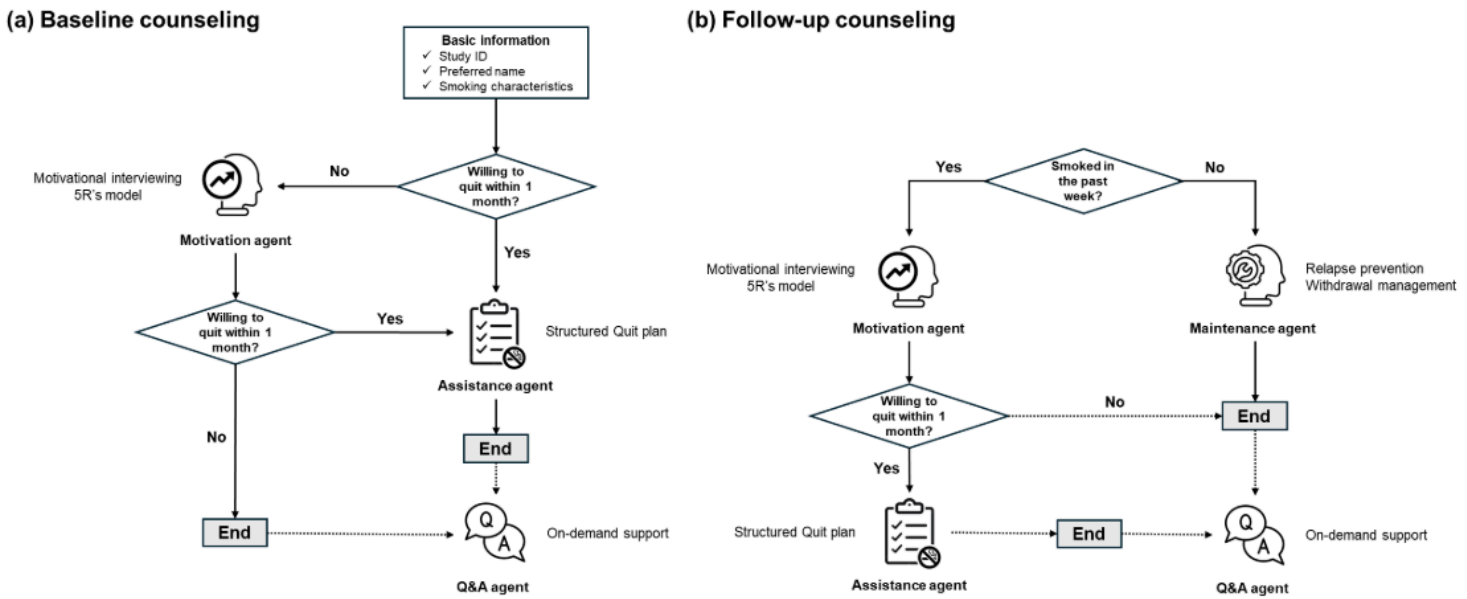


Figure 1

Multi-agent framework of the chatbot for (a) Baseline counseling and (b) Follow-up counseling. Diamond shapes represent decision nodes based on user state or input. During baseline counseling, users not ready to quit are routed to the motivation agent for motivation enhancement through motivational interviewing (MI) and the 5R’s model (relevance, risks, rewards, roadblocks and repetition). Ready users proceed to the Assistive agent to develop a structured quit plan. During follow-up counseling, users who reported smoking in the past week receive further cessation support from the motivation agent, while those who reported quitting receive withdrawal management and relapse prevention from the maintenance agent. After each counseling session, a dedicated question and answering (Q&A) agent takes over to provide on-demand support.

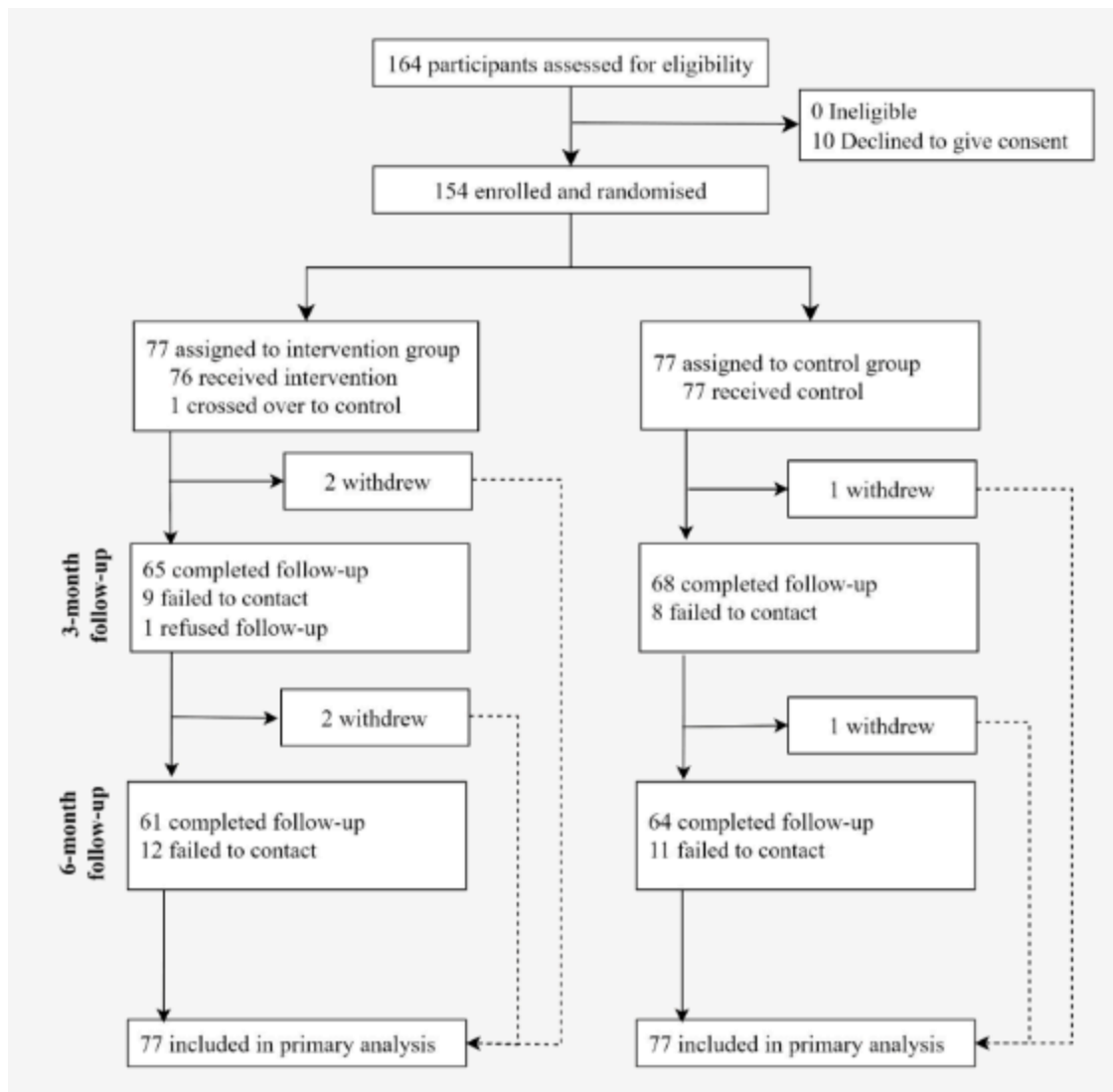


Figure 2

Participants flow diagram

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.docx](#)
- [machinelearningchecklist.pdf](#)
- [CONSORTchecklist.doc](#)
- [nrsoftwarepolicy.pdf](#)
- [nrreportingsummary.pdf](#)
- [SupplementData1.pdf](#)
- [NCOMMS26003312BReportingSummary.pdf](#)