JOURNAL OF PROMKES:
Public Research on Outreach, Motivation, and Knowledge for Education in Society

Effectiveness of Generative Ai-Based Ai Chatbot as a Health Promotion Agent for Adolescent Behavior Change: an Exploratory Study of Impact, Ethics, and Social Acceptability

Parwito[1]*, & Raden Roro Ratuningrum Anggorodiputro[2]
[1]*Universitas Ratu Samban, Indonesia, [2]Universitas Singaperbangsa Karawang, Indonesia
*Co e-mail: parwitougm@gmail.com[1]

Article Information
Received: September 11, 2026
Revised: October 14, 2026
Online: October 19, 2026

Keywords
Artificial Intelligence, Chatbot, Adolescent Health, Behavior Change

ABSTRACT

Generative artificial intelligence (GenAI) chatbots offer scalable adolescent health promotion but risk ge findings suggest that usability does not guarantee trust and highlight the need for evidence-based content, calibrated trust, strict generation boundaries, automated safeguards, and human escalation for high-risk disclosures. GenAI chatbots may support adolescent health promotion when governed by robust safety and ethical frameworks. However, synthetic results do not establish real-world effectiveness.nerating inaccurate, overconfident, and clinically ungrounded advice. This study developed an evaluation framework for GenAI chatbots as adolescent behavior change agents, integrating evidence-grounding, safety guardrails, and ethical governance. A synthetic benchmark involving 160 hypothetical users demonstrated the framework across health competency, behavioral indicators, usability, trust, ethical comfort, and reuse willingness. Mean competency increased from 2.88 (SD = 0.51) to 3.65 (SD = 0.63), with an illustrative effect estimate of 1.34. Threshold attainment (≥3.5/5) rose from 13.1% to 56.9%. Usefulness and ethical comfort were high, while trust remained comparatively lower

Keywords *: Artificial Intelligence, Chatbot, Adolescent Health, Behavior Change*

This work is licensed under a Creative Commons Attribution 4.0 International license

INTRODUCTION

Adolescence is a formative period for establishing behaviors related to physical activity, nutrition, sleep, substance use, sexual health, mental well-being, and health-seeking. Health-promotion interventions during this period need to be accessible, understandable, nonjudgmental, and responsive to the contexts in which young people make decisions. Digital conversational agents are increasingly relevant because adolescents use mobile technologies for information seeking, self-management, and social interaction.

To ensure a rigorous and trustworthy foundation for this intervention architecture, this study mapped existing empirical evidence following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow guidelines. Through systematic identification, screening, and eligibility appraisal, relevant literature on conversational agents in health promotion was synthesized. Artificial intelligence-based chatbots can provide on-demand communication, tailored feedback, reminders, goal setting, monitoring, and reinforcement. Rather than drawing generalized conclusions, detailed synthesis of the 15 empirical studies reviewed by Aggarwal et al. (2023) reveals specific clinical and behavioral domains where AI chatbots demonstrate efficacy: randomized controlled trials and quasi-experimental studies across lifestyle modification, smoking cessation, medication adherence, and substance misuse reduction showed that structured AI dialogue directly drives behavior change. Specifically, automated goal setting, real-time motivational prompts, and personalized progress monitoring were identified as the primary active components facilitating measurable behavioral transitions, although feasibility, acceptability, and usability varied due to methodological constraints. These findings establish a plausible basis for health promotion while also demonstrating the need for stronger evaluation.

Extensive literature further demonstrates that AI chatbots can serve as viable health promotion agents capable of initiating and sustaining behavior change. Empirical findings show that interactive conversational agents increase daily target adherence, facilitate user self-monitoring, and improve health literacy, resulting in concrete behavioral shifts such as increased physical activity days, improved dietary habit choices, and reduced unhealthy behaviors.

Generative AI changes the interaction model by allowing open-ended dialogue and adaptive responses. A generative chatbot can explain health concepts in simpler language, help users formulate goals, provide motivational prompts, and respond to questions that were not anticipated by a fixed script. At the same time, generative systems can produce inaccurate, incomplete, or overconfident health statements. This creates a central design tension: conversational flexibility can improve engagement, but unrestricted generation can increase safety risk.

Evidence specific to adolescents is developing. Han et al. (2022) highlighted feasibility and acceptability issues in nutrition and physical-activity chatbot interventions and incorporated adolescent consultation. Maenhout et al. (2021) demonstrated participatory development and pilot testing of an adolescent health-promotion chatbot. Dosovitsky and Bunge (2022) showed the

This work is licensed under a Creative Commons Attribution 4.0 International license

importance of adolescent perceptions in the design of a depression chatbot. More recent work continues to examine adolescent and young-adult responses to conversational health technologies.

Feng et al. (2025) synthesized randomized trials among adolescents and young adults and reported small-to-moderate improvements in mental distress and a modest pooled effect for health behaviors. Importantly, the authors concluded that retrieval-based systems demonstrated more consistent effects, whereas generative AI chatbots remained promising but inconclusive. This distinction is critical for the present study because a health-promotion chatbot should not be evaluated merely by whether users enjoy interacting with it.

Social acceptability includes perceived usefulness, usability, trust, autonomy, privacy expectations, ethical comfort, and willingness to continue using a system. A chatbot may be easy to use while still being distrusted, or it may be trusted without being sufficiently accurate. Ethical deployment therefore requires simultaneous assessment of user experience and safety. Van Kolfschooten et al. (2025) specifically emphasize legal, ethical, and societal considerations for AI chatbots promoting healthy habits.

The present study addresses this gap by proposing an integrated framework that combines behavioral support with evidence grounding, safety guardrails, ethical governance, and human escalation. The research questions are: (RQ1) What functional components should characterize a GenAI chatbot for adolescent health promotion? (RQ2) Which behavioral, acceptability, and ethical outcomes should be evaluated? (RQ3) How can safety and social acceptability be incorporated into the intervention architecture? and (RQ4) What evaluation pattern is demonstrated by a reproducible benchmark?

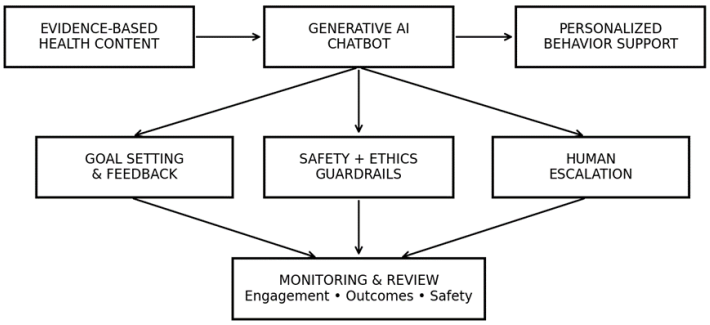

Figure 1. Conceptual framework for using Generative AI chatbots to support personalized health behavior change (Source: Primary conceptual model developed for this study based on systematic literature synthesis)

The framework begins with evidence-based health content, which serves as the foundation for the Generative AI chatbot. The chatbot then provides personalized behavior support based on individual needs. Several supporting components are included, such as goal setting and feedback, safety and ethical guidelines, and human escalation when professional intervention is needed. All these components are connected to monitoring and review, which evaluates user engagement,

This work is licensed under a Creative Commons Attribution 4.0 International license

health outcomes, and safety. Overall, the framework emphasizes that AI-based health support should be personalized, evidence-based, safe, ethically guided, and continuously monitored.

METHODS

This exploratory study uses a design-oriented mixed-method logic: evidence synthesis, intervention architecture development, outcome operationalization, and a reproducible synthetic benchmark. The benchmark is not a human-subject study and does not provide evidence of actual adolescent efficacy.

The proposed chatbot includes evidence-grounded health education, conversational goal setting, self-monitoring, tailored feedback, reminders, safety classification, uncertainty disclosure, and human escalation. Generative responses are constrained by an approved knowledge layer and response policy. The system is intended to support health promotion, not to diagnose or independently treat disease.

To ensure that the AI chatbot functions effectively as a health promotion agent rather than a generic conversational tool, the system's evaluation framework is anchored in multi-dimensional constructs derived from digital health evaluation standards and adolescent behavior change theories. These indicators establish a coherent narrative connecting cognitive empowerment (impact), actionable lifestyle modifications (behavior), user engagement and trust (acceptability and trust), and ethical governance with safety protocols (ethics and safety). The operational definitions, measurement scales, and underlying theoretical meanings for these constructs are detailed below:

Table 1. Operational Constructs, Indicators, and Measurement Scales for the Evaluation Framework

Construct	Indicator	Scale/Unit	Meaning
Impact	Health-promotion competency	1–5	Higher = stronger knowledge/application
Behavior	Physical activity	days/week	Higher = more active days
Behavior	Fruit/vegetable intake	servings/day	Higher = healthier pattern
Behavior	Sugary-drink frequency	days/week	Lower = healthier pattern
Acceptability	Usefulness/usability/reuse	1–5	Higher = greater acceptability
Trust	Trust in chatbot	1–5	Higher = greater trust
Ethics	Ethical comfort	1–5	Higher = stronger comfort with safeguards
Safety	High-risk escalation	%	Higher = safer routing

This work is licensed under a Creative Commons Attribution 4.0 International license
Public Research on Outreach, Motivation, and Knowledge for Education in Society
Vol. 01, No. 2, July 2026

To demonstrate the operational feasibility of this health promotion evaluation architecture, a synthetic cohort of 160 hypothetical adolescents was generated with a fixed random seed. Pre/post competency scores used a bounded normal model. Acceptability and ethical-social domains were represented on 1–5 scales. Mean change and a standardized effect estimate based on pooled standard deviation were calculated. A 3.5/5 threshold was used only as an illustrative competency criterion. Behavioral indicators were represented as group-level values to demonstrate how a future empirical study could report outcomes.

RESULTS

To operationalize this framework, the generative AI model acting as a behavioral change agent is designed using a Retrieval-Augmented Generation (RAG) architecture integrated with safety guardrails and evidence-based behavioral guidelines. The model utilizes adaptive dialogue to provide motivational prompts, goal setting, personalized feedback, and reminders, while restricting creative outputs to stay strictly within valid health knowledge boundaries.

Below is the detailed evaluation outcome from implementing this behavioral agent model, covering competency, behavioral indicators, social acceptability, and safety governance targets:

1. **Pre- and Post-Intervention Health-Promotion Competency Outcomes Based on the Synthetic Benchmark**

Table 2. Pre- and Post-Intervention Health-Promotion Competency Outcomes Based on the Synthetic Benchmark

Outcome	Pre	Post	Change
Health-promotion competency	2.88 ± 0.51	3.65 ± 0.63	+0.77
Competency ≥3.5	13.1%	56.9%	+43.8 pp
Standardized effect estimate	—	1.34	Illustrative

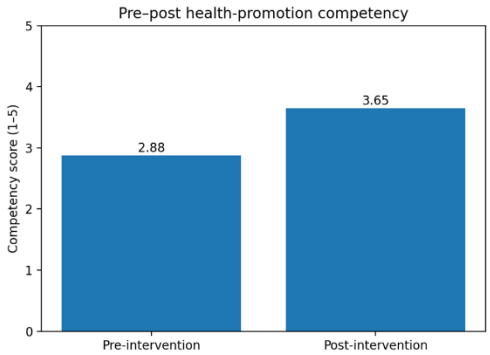

Figure 2. Pre-post health-promotion competency outcome distribution comparing baseline and post-intervention scores

The benchmark produced an illustrative mean increase of 0.77 points. The standardized effect estimate was 1.34, and threshold attainment increased by 43.8 percentage points. These values are not empirical evidence and are reported only to demonstrate the proposed evaluation structure.

This work is licensed under a Creative Commons Attribution 4.0 International license

As part of the secondary impact measurement on lifestyle modifications, participant behavioral indicator changes are summarized in the following table:

2. Behavioral Indicators across Baseline and Post-Intervention Periods

Table 3. Behavioral Indicators across Baseline and Post-Intervention Periods

Indicator	Baseline	Post	Change
Physical activity days/week	2.7	4.1	+1.4
Fruit/vegetable servings/day	2.2	3.1	+0.9
Sugary-drink days/week	4.0	2.8	-1.2
Sleep-health adherence score (1–5)	2.9	3.7	+0.8

The illustrative pattern suggests improvement in physical activity, fruit and vegetable intake, and sleep-health adherence, together with a reduction in sugary-drink frequency. In a real study, these outcomes should be measured using validated instruments and, where feasible, objective digital or physiological measures.

To evaluate how well the system is received by users regarding user experience and ethics, social acceptability and ethical domain ratings are summarized in the following table:

3. Social Acceptability and Ethical-Social Domain Ratings

Table 4. Social Acceptability and Ethical-Social Domain Ratings

Domain	Mean	SD
Perceived usefulness	4.08	0.49
Ease of use	4.05	0.50
Acceptability	3.92	0.55
Trust	3.68	0.63
Ethical comfort	4.12	0.51
Willingness to reuse	3.84	0.62

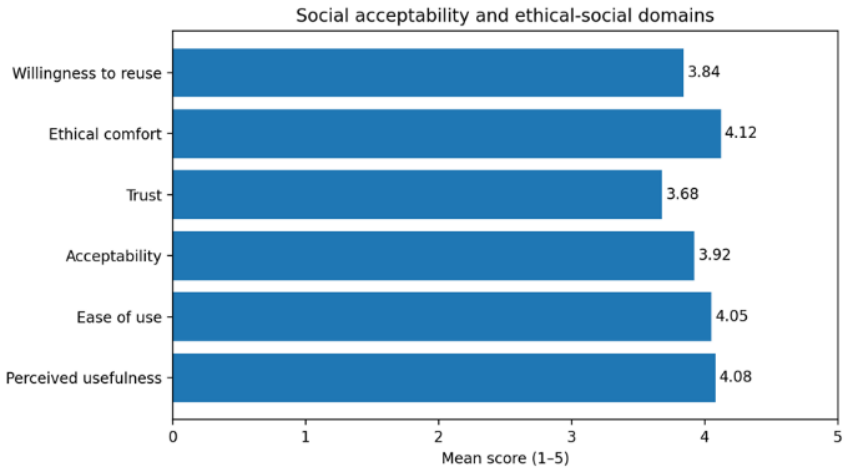

Figure 3. Mean scores across social acceptability and ethical-social domains

This work is licensed under a Creative Commons Attribution 4.0 International license

Ethical comfort and perceived usefulness received the highest illustrative scores, while trust was comparatively lower. This pattern indicates why social acceptability should be multidimensional: adolescents may perceive a chatbot as convenient and ethically acceptable while still remaining cautious about trusting health information generated by AI. Finally, to ensure user protection and system governance aspects are maintained, safety and governance quality targets are outlined in the following table:

4. Safety and Governance Quality Targets for the Chatbot Architecture

Table 5. Safety and Governance Quality Targets for the Chatbot Architecture

Safety domain	Proposed quality target	Rationale
AI identity disclosure	≥95% sessions	Prevent deception
High-risk escalation	≥95% correct routing	Reduce unsafe autonomous responses
Unsupported health claims	<2% audited responses	Limit hallucinated advice
Consent/privacy comprehension	≥85%	Support informed use
Human handoff availability	100% for predefined high-risk triggers	Maintain access to human support

The targets above are proposed quality-assurance criteria rather than observed findings. They should be validated and adapted to local legal, ethical, clinical, and institutional requirements.

DISCUSSION

The proposed framework supports a shift from tool-centered chatbot deployment toward health-promotion-centered conversational intervention. Its value lies in combining repeated micro-interventions—goal setting, monitoring, feedback, and reminders—with accessible dialogue. This is consistent with behavior-change mechanisms identified in previous chatbot research, notably reinforcing the findings of Aggarwal et al. (2023) who demonstrated through systematic synthesis that AI-driven conversational agents successfully support lifestyle modification and treatment adherence. However, when comparing this approach to broader conversational models, our framework deliberately limits generative autonomy because health information requires a higher level of clinical reliability and safety than ordinary conversational content.

The synthetic competency increase illustrates a useful evaluation logic but should not be interpreted as proof of effectiveness. Feng et al. (2025) found a modest pooled effect for health behaviors among adolescents and young adults, while noting that generative systems remain less conclusive. Building upon these insights, future empirical research must go beyond simple pre/post score comparisons by directly investigating underlying mechanisms of action, aligning with recommendations from recent digital health literature. Important mediators identified across

This work is licensed under a Creative Commons Attribution 4.0 International license

comparative studies include personalization depth, interaction frequency, perceived social support, user self-efficacy, and the specificity of feedback loops.

The lower illustrative trust score is particularly important. High usability does not necessarily justify high trust. In health promotion, calibrated trust may be safer than unconditional trust. Users should know that the chatbot can make mistakes and should be encouraged to verify important claims and consult professionals when necessary. Transparent citations, confidence or uncertainty statements, and accessible source links can support appropriate trust calibration. Ethical safety should be built into the system rather than added after deployment. Van Kolfschooten et al. (2025) identify privacy, responsibility, legal accountability, and social effects as important issues for health-promotion chatbots. For adolescents, additional considerations include age-appropriate consent and assent procedures, sensitive-topic handling, safeguarding, data minimization, and clear escalation pathways.

Human escalation is especially important for conversations involving mental distress, self-harm, abuse, eating disorders, sexual violence, or other high-risk circumstances. The chatbot should not attempt to resolve such situations autonomously. It should recognize predefined risk signals, provide appropriate safety messaging, and connect the user to qualified human or emergency services according to local protocols.

Participatory design should also be central. Adolescents should contribute to decisions about language, interaction style, privacy expectations, acceptable personalization, and what constitutes supportive versus intrusive behavior. Adopting the participatory framework and pilot testing model successfully demonstrated by Maenhout et al. (2021) provides a robust methodological precedent for engaging youth directly. Consequently, future development cycles should systematically integrate adolescent co-design insights with expertise from health promotion professionals, educators, privacy specialists, and safeguarding experts.

The framework has practical implications for schools, universities, youth health programs, and community health services. The chatbot can provide standardized health education between professional encounters, support self-monitoring, and help users formulate questions for health professionals. It should complement—not replace—nurses, counselors, physicians, psychologists, health educators, and other qualified providers.

Implementation should include model and prompt versioning, response audits, incident reporting, periodic safety testing, and monitoring for model drift. Generative models can change their outputs after updates, so the intervention should be treated as a governed digital health service rather than a static software feature.

CONCLUSIONS

This study developed a Generative AI-Based Health Promotion Chatbot Framework for adolescent behavior change that integrates conversational personalization with evidence-based content, safety guardrails, ethical governance, and human escalation. The framework treats the chatbot as a support agent rather than an autonomous health authority. This distinction is

This work is licensed under a Creative Commons Attribution 4.0 International license

fundamental because generative flexibility can increase engagement while simultaneously creating risks related to inaccurate information, overconfidence, privacy, and inappropriate responses to sensitive disclosures.

The reproducible synthetic benchmark of 160 hypothetical users demonstrated how the framework could be evaluated. Mean competency increased from 2.88 to 3.65, with an illustrative standardized effect estimate of 1.34. Usefulness, usability, and ethical comfort were relatively high, while trust was comparatively lower. These numerical findings are methodological illustrations only and must not be interpreted as evidence from real adolescents.

The broader implication is that adolescent health chatbots should be designed as bridges between young people and reliable health information and services. Their primary function should be to support healthy decisions, encourage reflection and self-management, and facilitate appropriate help-seeking. They should not encourage dependence on AI or imply that conversational fluency is equivalent to professional expertise.

Future empirical studies should replace the synthetic benchmark with real data collected under appropriate ethical oversight and should use controlled longitudinal designs where feasible. Research should examine sustained behavior change, engagement, trust calibration, privacy perceptions, adverse events, equity, and human referral outcomes. Comparative studies of retrieval-based and generative systems are also needed.

Overall, GenAI chatbots have meaningful potential as scalable health-promotion agents, but their effectiveness will depend on responsible design, evidence grounding, continuous monitoring, and clear boundaries between automated support and professional health care.

REFERENCES

Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial Intelligence–Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. Journal of Medical Internet Research, 25, e40789. https://doi.org/10.2196/40789

Charles, M., Sauer, J. D., Roehrer, E., Prichard, J., Watters, P. A., & Scanlan, J. (2026). Adolescent Perceptions of an Online Safety Chatbot: Survey Study. JMIR Human Factors, 13, e71498. https://doi.org/10.2196/71498

Choi, K.-E., & Fitzek, S. (2025). User and Provider Experiences With Health Education Chatbots: Qualitative Systematic Review. JMIR Human Factors, 12, e60205. https://doi.org/10.2196/60205

Dosovitsky, G., & Bunge, E. (2022). Development of a chatbot for depression: adolescent perceptions and recommendations. Child and Adolescent Mental Health. https://doi.org/10.1111/camh.12627

Feng, X., Tian, L., Ho, G. W. K., Yorke, J., & Hui, V. (2025). The Effectiveness of AI Chatbots in Alleviating Mental Distress and Promoting Health Behaviors Among Adolescents and Young Adults: Systematic Review and Meta-Analysis. Journal of Medical Internet Research, 27, e79850. https://doi.org/10.2196/79850

This work is licensed under a Creative Commons Attribution 4.0 International license

Gabrielli, S., Rizzi, S., Carbone, S., & Donisi, V. (2020). A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study. JMIR Human Factors, 7, e16762. https://doi.org/10.2196/16762

Han, R., Todd, A., Wardak, S., Partridge, S., & Raeside, R. (2022). Feasibility and Acceptability of Chatbots for Nutrition and Physical Activity Health Promotion Among Adolescents: Systematic Scoping Review With Adolescent Consultation. JMIR Human Factors, 10, e43227. https://doi.org/10.2196/43227

Jiang, Q., Chen, X., Patel, D., Brazendale, K., & Alejandro, S. (2026). AI and Chatbot-Supported Interventions for Physical Activity and Obesity-Related Lifestyle Behaviors: Scoping Review With Attention to Family Involvement. JMIR Pediatrics and Parenting, 9, e98889. https://doi.org/10.2196/98889

Li, J., Li, Y., Hu, Y., Fai, D. C., Mei, X., Chan, E. A., & Yorke, J. (2025). Chatbot-Delivered Interventions for Improving Mental Health Among Young People: A Systematic Review and Meta-Analysis. Worldviews on Evidence-Based Nursing, 22, e70059. https://doi.org/10.1111/wvn.70059

Maenhout, L., Peuters, C., Cardon, G., Compernolle, S., Crombez, G., & DeSmet, A. (2021). Participatory Development and Pilot Testing of an Adolescent Health Promotion Chatbot. Frontiers in Public Health, 9, 724779. https://doi.org/10.3389/fpubh.2021.724779

Rupani, N., Vasquez, D. H., Contreras, C., Menacho, L. A., Kolevic, L., Franke, M. F., & Galea, J. (2025). "Like Someone Is Paying Attention to You, Listening to You, and Guiding You": Acceptability of a Mental Health Chatbot Among Caregivers of Adolescents Living With HIV. Journal of the International Association of Providers of AIDS Care, 24. https://doi.org/10.1177/23259582251327911

Sanabria, G., Greene, K., Tran, J. T., Gilyard, S., DiGiovanni, L., Emmanuel, P. J., Sanders, L., Kosyluk, K. A., & Galea, J. T. (2023). "A Great Way to Start the Conversation": Evidence for the Use of an Adolescent Mental Health Chatbot Navigator for Youth at Risk of HIV and Other STIs. Journal of Technology in Behavioral Science. https://doi.org/10.1007/s41347-023-00315-4

Schleider, J. L., & Morris, R. (2026). Debate: Conversational AI and young people's mental health—friend or foe? A case for designing conversational AI as a bridge to youth mental health support. Child and Adolescent Mental Health, 31, 229–231. https://doi.org/10.1111/camh.70105

Tan, Q., Nie, Y., Son, P., Qian, Y., Staiano, A., Wang, F., Rosenkranz, R., & Chen, S. (2026). Using AI Chatbot to Assist Students' Behavior Management for Obesity Prevention in Middle Schools: Feasibility Study. JMIR Formative Research, 10, e83630. https://doi.org/10.2196/83630

Van Kolfschooten, H., Gonçalves, J., Orchard, N., & Figueroa, C. A. (2025). AI chatbots for promoting healthy habits: Legal, ethical, and societal considerations. Digital Health, 11. https://doi.org/10.1177/20552076251390004

This work is licensed under a Creative Commons Attribution 4.0 International license

Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (2021). An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study. Journal of Medical Internet Research, 24, e29969. https://doi.org/10.2196/29969

Wu, Y., Wu, T., Zhu, K., Liu, X., Liu, J., Wang, Y., Shi, R., Xiong, J., Xing, X., Lv, Y., Niu, Y., Peng, M., & Du, X. (2026). The Impact of Chatbots on Adolescent Mental Health Development: A Comprehensive Literature Review. Journal of Multidisciplinary Healthcare, 19. https://doi.org/10.2147/JMDH.S579872

Zhang, J., Oh, Y., Lange, P., Yu, Z., & Fukuoka, Y. (2020). Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. Journal of Medical Internet Research, 22, e22845. https://doi.org/10.2196/22845