

AI-Enabled Chatbot Interventions on Health Outcomes among People Living with HIV: A framework-guided Systematic Review

Atena Pasha

Department of Psychology and Sociology, Texas A&M University-Kingsville, TX, USA <https://orcid.org/0000-0001-5603-4961>

Mohammad Jahanaray

School of Education, Virginia Commonwealth University, VA, USA <https://orcid.org/0000-0001-8224-7645>

Abdul-Hanan Saani Inusah

Department of Health Promotion, Education and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA; South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

Xiaoming Li

Department of Health Promotion, Education and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA; South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

Shan Qiao

shanqiao@mailbox.sc.edu

Department of Health Promotion, Education and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA; South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

Systematic Review

Keywords: HIV Infections, Artificial intelligence, Chatbots, Digital health, Systematic review

Posted Date: February 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8935464/v1>

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

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Artificial intelligence (AI)–enabled chatbots are increasingly proposed as scalable tools to support linkage to HIV care, ART treatment adherence, and mental health among people living with HIV (PLWH). However, it remains unclear whether existing chatbot interventions are sufficiently developed, evaluated, or ethically governed to meaningfully improve outcomes for PLWH. Prior reviews have examined digital or mobile health tools broadly, but limited efforts have systematically assessed chatbot interventions through AI-specific implementation and governance frameworks.

Methods

We conducted a systematic review (PROSPERO: CRD420251271843) following the PRISMA 2020 guidelines across eight databases, covering publications from January 2005 to December 2025. Eligible studies examined the development, implementation, or evaluation of chatbot interventions designed to support health outcomes among PLWH. Data extraction and synthesis were guided by implementation and AI-specific frameworks, including SPIRIT-AI, CONSORT-AI, TEHAI, and WHO guidance on AI ethics and governance.

Results

Ten studies published between 2020 and 2025 met the inclusion criteria, representing 138 participants across diverse populations, including PLWH (adolescents and adults), caregivers, and healthcare providers, primarily from North and South America. Chatbots were designed to assist HIV management through ART adherence support, appointment reminders, resilience building, peer support promotion, healthcare provider access and connection, disclosure decision-making, and psychoeducation, with the majority being mobile- or web-based and using natural language processing or rule-based dialogue systems, with limited use of large language models. While usability, acceptability, and feasibility outcomes were consistently favorable, rigorous evaluation of clinical or mental health outcomes was largely absent. Framework-guided assessment revealed substantial gaps in reporting on potential harms, real-world integration, and adoption readiness, indicating limited alignment with established AI implementation and governance standards.

Conclusions

To the best of our knowledge, this is the first systematic review of AI-enabled chatbot interventions for PLWH, which highlights a critical gap between technological innovation and clinical impact. Despite growing enthusiasm for AI-enabled chatbots in HIV care, the current evidence base remains largely developmental and insufficient to support scale-up or policy adoption. Future research must move beyond usability testing toward ethically grounded, framework-aligned evaluations to translate promising digital innovations into scalable, ethical, and sustainable tools that can advance long-term HIV treatment outcomes.

INTRODUCTION

Human immunodeficiency virus (HIV) remains a major global public health challenge. In 2024, an estimated 40.8 million people were living with HIV worldwide, and approximately 630,000 individuals died from HIV-related causes, underscoring the persistent global burden of the epidemic. Despite these challenges, access to treatment has expanded substantially; by the end of 2024, nearly 31.6 million people living with HIV (PLWH) were receiving antiretroviral therapy (ART), corresponding to a global ART coverage rate of approximately 77% (1).

The widespread availability of ART has transformed HIV into a manageable chronic condition (2). There is increasing enthusiasm among various key stakeholders, such as PLWH, healthcare providers, and researchers, to enhance the quality of life for PLWH (3, 4). Alongside the "95-95-95" targets (5), which represent a critical objective for HIV treatment, care, and management, the ultimate aim is to help PLWH achieve both physical and psychological well-being (4). PLWH, despite achieving viral suppression, continue to encounter a variety of significant health challenges, including mental health disorders, substance use disorders, comorbidities related to HIV, accelerated aging, and various associated chronic diseases (6–8).

To address these challenges, PLWH need lifelong adherence to treatment, consistent engagement in care, and sustained self-management. These demands include attending routine clinical appointments, understanding and managing treatment regimens, coping with symptoms, and making daily health-related decisions (9). Unfortunately, effective self-care and disease management are frequently undermined by individual, social, and structural barriers, including treatment fatigue, disrupted daily routines, HIV-related stigma, and limited access to supportive services (10). These challenges highlight the need for acceptable, accessible, and scalable interventions that can improve both physical and psychological well-being through effective health interventions.

Technology-enhanced or enabled health interventions have increasingly been explored as strategies to address these challenges, a trend that was further accelerated by the COVID-19 pandemic (11, 12). Interventions such as Short Message Service (SMS) reminders, online counseling, and web-based education have demonstrated benefits for ART adherence and HIV self-management by enhancing privacy, reducing stigma-related barriers, and enabling remote access to reliable health information (13, 14). Within this broader landscape, chatbots, as conversational agents capable of simulating human dialogue through text or voice, have emerged as a promising class of digital health tools (11, 15). Chatbots have shown feasibility and acceptability across a range of health domains, including mental health support, chronic disease management, and medication adherence (11, 16, 17). Users frequently report favorable perceptions related to convenience, anonymity, and non-judgmental interaction (14, 16, 18).

In the context of HIV care, chatbot research is still at an early stage. A limited number of studies have evaluated chatbots designed specifically for PLWH, such as MARVIN, a bilingual chatbot developed to support ART adherence and HIV self-management, with early findings suggesting good usability and user satisfaction (19). Other chatbot interventions targeting adolescents living with HIV have shown promise for supporting mental health, facilitating

information-seeking, and improving communication with caregivers (13). However, the majority of HIV-related chatbot research to date has focused on prevention-oriented outcomes, such as HIV self-testing and pre-exposure prophylaxis (PrEP) information (18, 20–22). In contrast, relatively few studies have examined chatbot interventions aimed at improving health outcomes among PLWH (13, 19, 23), and no synthesis has systematically evaluated their effectiveness.

Recent advances in artificial intelligence (AI), particularly in natural language processing (NLP), machine learning, and large language models (LLMs), have substantially expanded the capabilities of chatbot technologies (11, 24). Modern AI-enabled chatbots can generate context-aware responses, personalize interactions, integrate multimodal data, and potentially learn from ongoing user engagement (20, 25). These features position AI-enabled chatbots as potentially valuable tools for supporting self-management, mental health, and clinical outcomes in HIV care (18, 19, 24). At the same time, the use of AI in health interventions raises important concerns related to algorithmic bias, data privacy, safety during mental health crises, and the limits of automated systems in providing empathic support (14, 16, 26). These risks underscore the importance of rigorous evaluation and transparent reporting.

Given that chatbots are AI-enabled interventions with potential safety-critical implications, their assessment should be informed by recognized frameworks, including Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) (27) and Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) (which provide guidance for reporting AI-based clinical trials) (28), Translational Evaluation of Healthcare AI (TEHAI) (which evaluates technical validity, clinical validity, and adoption readiness) (29), and the World Health Organization (WHO)'s guidance on the ethics and governance of AI for health, which emphasizes transparency, equity, accountability, and data governance (30). Evaluating alignment with these frameworks is essential to inform the safe, reliable, and equitable implementation of chatbot interventions in HIV care.

Despite growing interest in AI-driven digital health tools, no systematic review has focused specifically on chatbot-based interventions targeting health outcomes among PLWH. Existing reviews have examined either broad AI applications in HIV care without identifying chatbot interventions (24), mental health chatbots in general populations without attention to HIV-specific contexts (11, 31), or prevention-oriented HIV chatbots without focusing on treatment and clinical outcomes (32, 33). Furthermore, it remains unclear to what extent existing HIV chatbot studies align with established AI evaluation and reporting frameworks.

The purpose of this systematic review was to evaluate the effectiveness of chatbot-based interventions in improving health outcomes among PLWH. Specifically, this review aimed to:

1. Identify and evaluate studies using chatbots to support health outcomes (including both physical and mental health) among PLWH.
2. Describe key protocol characteristics, including chatbot type, AI capabilities, delivery format, target population, and integration into HIV care settings.
3. Assess methodological quality and alignment with existing AI-specific evaluation frameworks, including SPIRIT-AI, CONSORT-AI, TEHAI, and relevant WHO guidelines.

METHODS

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) (34) guidelines to ensure transparency, methodological rigor, and reproducibility. The review protocol was peer-reviewed and pre-registered in the PROSPERO database (ID: CRD420251271843). In addition to PRISMA guidance, the review was informed by AI-specific evaluation and reporting frameworks, including SPIRIT-AI (27), CONSORT-AI (28), TEHAI framework (29), and the WHO guidance on the ethics and governance of artificial intelligence for health (30). These frameworks guided the design of the data extraction process and the interpretation of findings.

Search Strategy

A comprehensive literature search was conducted across eight electronic databases, including PubMed, PsycINFO, Embase, Web of Science, IEEE Xplore, ClinicalTrials.gov, CINAHL, and Google Scholar. The search covered studies published between January 1, 2005, and December 1, 2025, reflecting the period during which chatbot and AI-enabled health interventions became increasingly prevalent. Search terms included combinations of keywords and controlled vocabulary related to HIV (e.g., “HIV,” “human immunodeficiency virus”), chatbots (e.g., “conversational agent,” “interactive agent”), and intervention domains of interest (e.g., “mental health,” “ART adherence,” “CD4 count”). Database-specific syntax and Boolean operators were applied as appropriate. The full search strategies for each database are provided in the Supplementary Materials. In addition to database searches, the reference lists of all included studies and relevant review articles were manually screened to identify additional eligible publications.

Eligibility Criteria

Studies were considered eligible if they were peer-reviewed publications written in English regarding the development, implementation, and assessment of a chatbot or conversational agent designed for PLWH to improve health wellbeing outcomes (e.g., mental health and/or clinical domains relevant to PLWH). Studies across intervention development and evaluation phases were included, including formative design or development studies, usability or acceptability evaluations, protocol papers describing planned evaluation, and other empirical studies in which a PLWH-focused chatbot was the core component of the work, regardless of whether health outcomes were empirically evaluated in the study.

Experimental, quasi-experimental, mixed-methods, and controlled observational study designs were considered. No restrictions were placed on participant age, gender, sexual orientation, or geographic location. Studies were excluded if they focused exclusively on HIV prevention without outcomes relevant to PLWH, did not include a chatbot as a core component, or were editorials, commentaries, narrative reviews, or conference abstracts.

Search Process and Study Selection

All records retrieved from the database searches were imported into Rayyan (35), a cloud-based systematic review management platform. Duplicate records were removed prior to screening. Titles and abstracts were independently screened by two reviewers to identify potentially eligible studies. Full texts of relevant articles were then retrieved and assessed for eligibility based on the predefined inclusion and exclusion criteria. Disagreements at any stage of the screening process were resolved through discussion, and when consensus could not be reached, a third reviewer was consulted. The study selection process is summarized using a PRISMA flow diagram (see Fig. 1).

Data Extraction and Synthesis

This systematic review was guided by established frameworks for the evaluation and reporting of AI-enabled health interventions and chatbot-focused studies in HIV care (27–30). Given the complexity of chatbot applications in HIV care and the importance of ethical and safety considerations, we adopted a multi-framework approach to comprehensively assess intervention characteristics, implementation considerations, and alignment with ethical and translational standards. Rather than applying each framework in its entirety, we adopted a selective and integrative approach, focusing on items most relevant to chatbot-based interventions and chatbot-focused studies targeting mental health and clinical outcomes among PLWH. Framework items were selected based on their applicability across diverse study designs, feasibility of extraction from published reports, and relevance to intervention safety, effectiveness, and implementation. Overlapping items across frameworks were harmonized to avoid redundancy and ensure coherence.

Specifically, we integrated guidance from SPIRIT-AI (27) and CONSORT-AI (28) to review study protocols and completed studies involving AI. These guides emphasize transparent reporting of AI system design, input data, human–AI interaction, error handling, and harms, which are critical for reproducibility and interpretability of AI-enabled interventions. In addition, we applied the TEHAI framework (29) to assess the readiness of chatbot interventions for real-world adoption across three domains: capability (technical performance and validation), utility (safety, usability, and clinical relevance), and adoption (integration, scalability, and sustainability). To ensure ethical and equity-oriented evaluation, we further drew on the WHO Guidance on the Ethics and Governance of AI for Health (30), which highlights principles of transparency, accountability, safety, inclusiveness, and human-centered design. These principles are particularly relevant for chatbot interventions targeting PLWH, a population that may face stigma, marginalization, and disparities in access to care.

Data extraction was conducted using a structured form informed by SPIRIT-AI, CONSORT-AI, TEHAI, and WHO guidance on AI ethics and governance (27–30). Extracted data included study characteristics, participant demographics, chatbot intervention features, AI components, delivery platforms, and integration into HIV care settings. Information related to mental health outcomes, clinical outcomes, feasibility, acceptability, usability, implementation challenges, ethical considerations, reported harms, limitations, and recommendations were also collected. Two reviewers (AI, MJ) independently extracted data from all included studies, and discrepancies were resolved through discussion to ensure accuracy and consistency.

A narrative synthesis was conducted due to heterogeneity in study designs, chatbot technologies, intervention targets, and outcome measures. Findings were synthesized across six key domains: 1) study and chatbot characteristics, 2) content development and user-centered design, 3) input output and privacy protections, 4) feasibility, acceptability, and usability, 5) preliminary efficacy, and 6) implementation challenges. Where sufficient similarity existed in outcome measures, quantitative results were summarized descriptively. Framework alignment was used to identify reporting gaps, methodological limitations, and implications for future research and implementation.

RESULTS

Characteristics of included studies

Among the 2,550 records retrieved from various databases, 10 papers that met the inclusion criteria were selected for the data extraction process (see Fig. 1). The included studies were published between 2020 and 2025, reflecting the relatively recent emergence of interest in the application of AI chatbots within HIV care. Three studies were conducted in the United States (36–38), and three in Peru (13, 23, 39) with two of the studies by the same research team. The remaining studies were conducted in Indonesia, Nigeria, Singapore, and Canada (19, 40–42).

Among the 10 included studies, four did not report data from human participants, as they focused primarily on chatbot development (40) or were feasibility studies with a planned trial (19). Among those reporting data from human participants, 138 participants were involved across these studies. One study sampled adults living with HIV, reporting a mean age of 40.2 years (SD = 11.5) (19) and five studies included youth living with HIV (13, 23, 38, 42, 43). Other participants included caregivers (14), while Comulada et al. (38) sampled staff (e.g., project directors, intervention coaches) and advisory board members, as shown in Table 1.

Table 1
Summary of included studies.

Study ID	Authors, Year	Country	Study design	Demographics	Sample size	Participants health condition
1	Ardiana et al., 2020	Indonesia	Chatbot development study with usability and accuracy testing	NR	30 users (randomly selected).	NR
2	Asaeikheybari et al., 2020	US	Technical chatbot system development with internal (alpha) evaluation	NR (Medical professionals and experts only for content validation)	NR (no human subjects)	NR (no human subjects)
3	Galea et al., 2024	Peru	Study protocol for chatbot development and pilot testing	<i>Planned:</i> Adolescents aged 10–19 years; sex and sexual/gender identity will be collected	<i>Planned:</i> up to 50 adolescents living with HIV for pilot testing	Adolescents living with HIV (ALWH) (acquired at or near birth, or during adolescence)
4	Ma et al., 2025	Canada	Mixed-methods usability evaluation study	Mean age 40.2 ± 11.5; men 82.1%, women 17.9%; diverse ethnic groups and sexual orientations reported	28(completed the 3-week chatbot usability study)	People living with HIV (PLWH)
5	Pande et al., 2023	Nigeria	Chatbot development and prototype design study	Age 18–25; other demographics NR	23 champions (design workshop participants)	Young people living with HIV (design informants)
6	Vasquez, et al. 2025	Peru	Human-centered chatbot development study	Adolescents living with HIV aged 11–19 years (Youth Advisory Board participants only)	6 adolescents	Adolescents living with HIV (ALWH)
7	Rupani et al., 2025	Peru	Qualitative acceptability study (in-depth interviews)	Age 30–70 among caregivers (60% aged 30–50, 40% aged 50–70); 20% male, 80% female	15	Caregivers of adolescents living with HIV (ALWH); caregivers' own mental health not directly assessed
8	Hightow-Weidman et al. 2022	Us	Multi-phase intervention development with usability testing and post-session qualitative feedback	Mean age 27.6 years; all participants were born male and identified as male; All were Black or African American	8	Young men who have sex with men (YMSM) living with HIV; all in care; all virally suppressed; all self-reported ≥ 90% ART adherence. Also 5 participants had disclosure stigma.
9	Koh et al., 2024	Singapore	Qualitative assessment of ChatGPT responses to predefined ART counselling questions	NR (no human subjects)	NR (no human subjects)	NR (no human subjects)
10	Comulada et al., 2024	US	Qualitative focus groups discussions	Age: young participants 20–24 years; staff/Advisory Board 21–38 years. Sex/Gender identity: predominantly cis gendered, majority identifying as gay, and predominantly Black	28 (13 research staff, 8 community advisory board members, 7 young people)	Young people living with HIV (on ART); HIV research staff and advisors (not living with HIV)
Footnotes:						
<i>Sample size reflects the number of individuals involved in the study phase reported (e.g., development, usability, or evaluation), as applicable.</i>						
NR = Not reported						

Characteristics of Chatbot Interventions

Chatbots in the reviewed studies were designed or proposed to address a range of physical or psychosocial needs among PLWH, shown in Table 2. Specifically, the chatbots have been designed to provide HIV/AIDS-related information and counseling support (40), enhance engagement in care and peer support (44), facilitate mental health education, self-help skills, and depression screening (23), support HIV and ART management or offer advice (19, 41), foster empathy and motivation and deliver general assistance such as scheduling appointments or medication reminders (42), offer psychoeducational coping support and linkage to care for adolescents to address depression and stigma (13, 14), and assist with HIV disclosure decision-making (37).

Table 2
AI Application Features (based on SPIRIT-AI items).

St ID	AI Type/Model (e.g., chatbot, LLM, rule-based, neural network)	Intended Use & Purpose *	Setting & Integration Requirements (onsite/offsite, devices, internet, EHR integration)	Implementation details (Intervention duration / sessions)	Version of AI /Algorithm Used	Input Data Type **	Handling of Poor-Quality/Missing Data	Human Interact
1	Mobile-based chatbot	To provide HIV/AIDS-related information and support counseling activities via a mobile chat interface	Android mobile app (requires Android OS ≥ 4.4)	App installation followed by usability evaluation in which users completed predefined chat tasks	AI engine based on AIML (Artificial Intelligence Markup Language) chatbot	User input as text questions	Incorrect responses are identified and reported to administrators for knowledge base updates.	End-user interact directly the cha
2	Natural Language Processing (NLP)-enhanced mobile app with integrated chatbot features	To support and inform young adults living with HIV via a social-media-style app, improving engagement in care and peer support.	Smartphone app (Android/iOS)	An alpha test of the NLP feature was conducted (details NR). For implementation, EasyESA semantic framework was used.	Explicit Semantic Analysis (ESA) (A NLP technique)	User-generated text data within the app (posts, queries, or messages)	NR	Users ir indeper with the feature; clinician support availabl elsewhere the app
3	Chatbot / conversational agent (platform-based)	Provide mental health education, self-help skills, depression screening support, and linkage to care	Mobile-based chatbot?NR	<i>Planned:</i> Three-phase implementation conducted over 2 years, including formative qualitative research and iterative chatbot development in Year 1, followed by pilot testing of the finalized chatbot with up to 2 weeks of user interaction and subsequent data analysis	Chatbot is programmed and deployed using the SmartBot360 platform through iterative versions (0.1–0.3 during development; 1.0 finalized prototype for pilot testing); AI model architecture not specified	User input is text-based conversation	Study mentions that data will be cleaned, and summary tables will be generated, but the procedures/methods of it were not specified.	Users d interact the cha
4	AI-based chatbot	Support HIV self-management; provide ART information and reminders	Individual mobile devices, Facebook Messenger required	Participants were instructed to complete ≥ 20 conversations within 3 weeks.	Rasa framework; intent classification; entity extraction; decision trees	User text input	When unable to understand the user's intent or reach a diagnosis- or treatment-related intent, it acknowledges its limits and encourages contact with a healthcare provider.	Primari unsupe user-cl interact (preced one-tim training session
5	Hybrid chatbot (rule-based, frame-based, and machine learning components)	Health coach chatbot to guide and support young people living with HIV across coaching needs (information, engagement, empathy, general assistance)	Accessible via WhatsApp on users' smartphones	Still in development phase – no formal intervention deployment yet	Voiceflow Dialog Manager; Rasa NLU; pre-trained sentiment and emotion detection models integrated in selected dialog acts.	User text input via WhatsApp chat; interactive questions and prompts.	Integrates hybrid dialog management, affect-aware responses, and human-in-the-loop escalation, which can be interpreted as mechanisms for handling sensitive and potentially problematic conversational situations.	Fully automa chatbot current prototy phase; unsupe user interact Human escalat describ

Notes:

* **Use and purpose:** what the AI is designed to do; patient, provider, or both

** **Input Data Type:** what the chatbot needs: text, voice, EHR data, etc.

*** **Human-AI Interaction:** Specifying whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users

******How Output Informs Decision-Making:** how it guides patient/provider behavior

NR = Not reported

St ID	AI Type/Model (e.g., chatbot, LLM, rule-based, neural network)	Intended Use & Purpose *	Setting & Integration Requirements (onsite/offsite, devices, internet, EHR integration)	Implementation details (Intervention duration / sessions)	Version of AI /Algorithm Used	Input Data Type **	Handling of Poor-Quality/Missing Data	Human Interact
								planned feature
6	Chatbot (text-based)	Psychoeducation and coping support for adolescents living with HIV (depression, stigma, coping skills). Addressing mild to moderate depression.	Web-based platform accessible via smartphone or web browser. Requires basic internet access.	No interventions implemented since the study is a development and design tutorial. However, co-design sessions over 5 months (monthly; each 90–120 min).	Smartbot360 platform (algorithmic details not specified). Wix platform was used after session 2 feedback.	User inputs via chat interface (text and predefined selections). Asking for name and age but no collection of sensitive identifiers (e.g., phone number, gender)	If the name is too short (≤ 2 characters) or numeric, EVA prompts the user to re-enter a valid name.	User interact directly with chatbot guided by prompts. Optional demand support through time chat a health profess
7	Chatbot	Mental health education, self-help skills, and linkage to care for adolescents living with HIV (ALWH).	Web-based chatbot accessed via an internet browser on participant devices.	Single session (2–3 hours) for caregiver study (20 min independent chatbot use + guided exploration of modules + post-interview)	Smartbot360 platform (algorithmic details not specified).	User text input and menu selections within chatbot	NR	Users interact with the chatbot independent study support present guidance emotional support of research procedu
8	AI-facilitated conversational role-play embedded in a mobile (mHealth) app	To support HIV status disclosure decision-making and communication skills practice among young men who have sex with men (YMSM) living with HIV	Mobile app on personal device; usability testing conducted in private sessions with study staff	Multi-phase development; final intervention includes 4 modules, 24 activities, and 8 AI-facilitated role-playing scenarios; usability sessions lasted 60–90 minutes. After usability feedback, the final time per module was reduced to 45 minutes	NR	User-typed inputs during disclosure conversations (Speech feature added in later versions); responses selected from a pre-developed dialogue database. System-collected paradata including activity completion, time spent, and chat logs.	NR	Direct conversational interaction between and virtual avatar, i version only but version spoken commu as well.
9	ChatGPT (AI natural language processing chatbot)	To provide ART counselling and advice for people living with HIV, addressing ART knowledge, initiation, side effects, adherence, and sexual health	Publicly accessible chatbot; online use;	ChatGPT instructed to answer predefined ART-related questions across three domains; no intervention duration or sessions reported	ChatGPT version 3.5	Text-based user prompts/questions	Quality of responses limited or generic when input detail was insufficient	Direct text-based interaction between research and chat

Notes:

* **Use and purpose:** what the AI is designed to do; patient, provider, or both

** **Input Data Type:** what the chatbot needs: text, voice, EHR data, etc.

*** **Human-AI Interaction:** Specifying whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users

******How Output Informs Decision-Making:** how it guides patient/provider behavior

NR = Not reported

St ID	AI Type/Model (e.g., chatbot, LLM, rule-based, neural network)	Intended Use & Purpose *	Setting & Integration Requirements (onsite/offsite, devices, internet, EHR integration)	Implementation details (Intervention duration / sessions)	Version of AI /Algorithm Used	Input Data Type **	Handling of Poor-Quality/Missing Data	Human Interact
10	Chatbot (conversational agent). Type or model NR	Proposed functions include scheduling appointments, referrals, reminders & 24/7 support	Text-based chatbot demonstrated; potential integration with websites and SMS discussed	No intervention implemented; a 5-minute chatbot demonstration was shown to participants during focus groups	NR	User text input	NR	Planner involves text-based interact users
Notes: * Use and purpose: what the AI is designed to do; patient, provider, or both ** Input Data Type: what the chatbot needs: text, voice, EHR data, etc. *** Human-AI Interaction: Specifying whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users ****How Output Informs Decision-Making: how it guides patient/provider behavior NR = Not reported								

All AI chatbots in the review have been either mobile- or web-based, allowing access on personal devices primarily through text-based communication (14, 37). Subsequent iterations of the chatbot created by Hightow-Weidman et al. (37) introduced the capability for spoken communication. These chatbots are built using various programming languages and commonly leverage NLP techniques, with platforms such as SmartBot360 frequently in use (13, 23, 41, 44).

NLP techniques, including Explicit Semantic Analysis (ESA), have been employed to integrate chatbot applications with resources, including the Positive Peers website, thereby enhancing user engagement and addressing users' concerns (44). For instance, Ardiana et al. (40) developed an Android-based chatbot focused on HIV counseling, utilizing Android Studio and the Java programming language. They implemented an Artificial Intelligence Markup Language (AIML) interpreter paired with a pattern-matching algorithm to effectively align user questions with suitable responses.

The chatbots also incorporate various algorithms, including intent classification and entity extraction (used for identifying elements such as time, drug names, or quantities), as well as decision trees for dialogue management, implemented through the Rasa framework, as seen in the MARVIN chatbot (19). Within these studies, ChatGPT was utilized in two distinct contexts: first, to provide counseling and guidance on ART to PLWH, addressing aspects like ART knowledge, initiation, side effects, adherence, and sexual health (41); and second, to examine perceptions of chatbots and potential integration of chatbots into HIV research studies (38).

Content Development and User-Centered Design

During the initial phase of AI chatbot development, several studies shared their response-generation and content-selection processes, which were influenced by user needs and requirements and employed a range of methodologies. For example, Galea et al. (23) planned to use qualitative semi-structured interviews to examine the thoughts, perspectives, and consequences of depression among adolescents living with HIV. Similarly, Comulada et al. (38) interviewed research staff, community advisory board members, and young people living with or at risk of HIV to evaluate perceptions of chatbots and their potential integration into HIV research.

Asaeikheybari et al. (44) utilized content that had been validated and approved by HIV practitioners and community advisory boards to support PLWH in multiple aspects of their lives. Other studies incorporated co-design strategies through patient and stakeholder engagement, featuring continuous communication with patient partners (19). In a similar vein, Vasquez et al. (13) a human-centered approach conducting interviews with healthcare professionals providing pediatric HIV services, and caregivers of adolescents living with HIV to inform the design of the EVA (Educación, Vinculación, y Autoayuda) chatbot. Lastly, Ardiana et al. (40) AIML chatbot design was informed primarily by interviews with HIV counselors.

Additionally, the initial phase of the Hightow-Weidman (43) study involved participants sharing their past disclosure experiences. They discussed disclosure strategies, including barriers and facilitators, collaborated in pairs to create realistic disclosure scenarios, and contributed to the development of a stand-alone interactive dialogue feature.

Inputs, Outputs, and Privacy Protections

In these studies, users primarily provided questions, symptoms, and situational details rather than formal personal data. Most systems deliberately minimized the collection of identifiable information, prioritizing user anonymity (e.g., Asaeikheybari et al., 2020; Vasquez et al., 2025). In some cases, users could delete all conversational records to address privacy concerns (19).

Users of the MARVIN system responded to inquiries regarding their preferred language as well as topics related to ART, including dosing, drug interactions, travel, and reminder requests (19). Conversely, EVA users discussed a range of mental health symptoms, such as anxiety, depression, isolation, suicidal

ideation, and stigma associated with both mental health and HIV. They also shared personal experiences, educational content, self-help strategies such as deep breathing exercises and emotion management, and topics related to linkage to care (13).

The test questions in the study conducted by Koh et al. (41) centered on ART initiation, missed doses, drug interactions, pregnancy and breastfeeding, and sexual health while on ART. Galea et al. (23) planned to administer a survey to collect data on sociodemographic factors (e.g., age, sex), HIV-related issues (such as acquisition route, current viral load, and frequency of missed HIV care visits), knowledge and history of depression, as well as prior experiences with chatbots.

Regarding outputs, these chatbots can provide information about HIV/AIDS to the general public or PLWH (40). They can guide users to relevant blog posts that closely match their queries and may address specific questions or concerns (44). They have also been shown to improve practical self-help coping skills (23), address issues related to time management, medication dosing, common drug interactions, medication storage, and identification (19), and enhance empathy and motivation, which may lead to improved ART adherence (42). Additionally, these chatbots can support users' emotional well-being (13), address internalized HIV stigma or fear of disclosing HIV status (37), and provide suggestions regarding ART initiation, adherence, and management of side effects (41).

Feasibility, acceptability, and usability

As shown in Table 3, five studies examined the outcomes of chatbot interventions, with an emphasis on usability, acceptability, and feasibility within their target populations. For instance, the chatbot developed by Ardiana et al. (40) achieved a user satisfaction score of 3.3 out of 4. Users also evaluated its learnability, the effectiveness of the AI-generated content in facilitating HIV/AIDS counseling, and the memorability of that content (40). Asaeikheybari et al. (44) assessed the answer ranking quality of their chatbot, the Precision HIV Health App, using metrics such as Normalized Discounted Cumulative Gain (showing how well the results are ordered by usefulness) and Precision@k (the proportion of top results that are correct), which effectively simulate human relevance judgments in recommendation scenarios. Their study also evaluated technical feasibility through alpha testing. Ma et al. (19) employed the Acceptability E-Scale (AES) and the Usability Metric for User Experience-lite (UMUX-lite), both of which indicated satisfactory levels of usability (Mean 69.9, in UMUX-lite) and acceptability (23.8/30 score in AES) for their chatbot, MARVIN.

Table 3
Key Findings regarding Implementation and preliminary efficacy.

Study ID	Feasibility (trial completion, recruitment, adherence to chatbot use, technical feasibility)	Acceptability (patient/provider satisfaction, trust, qualitative feedback)	User experience and engagement (User friendly and Usability)	Preliminary efficacy Health Outcomes	Implementation challenges	Strengths	Limitations /Concerns	Recommendations for Future Research
1	NR	User satisfaction score – 3.6/4.0	Overall positive. Usability testing – 3.30/4. 71% accurate of the answers given by the chatbot.	NR	NR	NR	Limited knowledge database leading to incorrect responses; accuracy dependent on expanding chatbot training data	Expand knowledge database training to improve response accuracy.
2	Technical feasibility via the NLP that process of free-text queries and generate semantic matches.	NR	NR	NR	NR	Design contribution: AI-powered mobile health app integrating NLP (ESA) to improve relevance of information retrieval and provide a compact, stigma-free platform for addressing user inquiries. Ability to chat with peers, clinical providers, and community forums. Chat out section for mindfulness videos.	NR	For the NLP system to be tested in future versions.
3	Plan to use the Feasibility of Intervention Measure	Plan to use the Acceptability of the Intervention Measure, the Intervention Appropriateness Measure.	Plan to assess usability and user experience.	NR	NR	Inexpensive, adaptable and highly scalable. Iterative human-centered design. It links users to real mental health care and services when needed (care linkage)	Single site; small convenience sample	Conduct efficacy trial with chatbot and expand to other settings.
4	Recruitment rate 74%; 70% retention	Mean Acceptability E-Scale (AES) score = 23.8/30; qualitative feedback highlighted reliable information, ease of access, and confidential interaction.	The Usability test, using Usability Metric for User Experience-lite (UMUX-lite) was good with a mean 69.9 (SD 14.2) surpassing the benchmark of 68.	NR	Limited understanding of user input; some users needed to rephrase questions to obtain relevant answers.	Reliable expert-validated information; easy access and real-time support; emotionally safe and confidential interaction	Small sample; single-site study; limited conversation topics and chatbot comprehension; use of Facebook Messenger as the deployment platform.	Further development including expanded improved language comprehension on additional platforms further evaluation.
5	NR (Planned: randomized controlled trial (RCT))	NR (Planned: qualitative feedback)	NR (Planned: RCT to measure use, usability, and user perceptions)	NR (Planned: adherence; psychological well-being)	Anticipated: large training data requirements for end-to-end dialog management; need for synthetic data generation with human verification.	NR	Anticipated ethical and safety concerns related to generative models; limited availability of real training data; need for human verification.	End-to-end human-in-the-loop evaluation; RCT evaluation

NR = Not reported

Study ID	Feasibility (trial completion, recruitment, adherence to chatbot use, technical feasibility)	Acceptability (patient/provider satisfaction, trust, qualitative feedback)	User experience and engagement (User friendly and Usability)	Preliminary efficacy Health Outcomes	Implementation challenges	Strengths	Limitations /Concerns	Recommendations for Future Research
					generative model risks.			
6	Feasibility assessed only during development resulting in successful recruitment of adolescents and clinicians; observed ability to use the chatbot as intended (~ 20 min/session); technical feasibility demonstrated through iterative testing and error correction; Planned formal trial completion or feasibility outcomes in future study.	During development sessions YAB indicated that EVA's content was informative, EVA's presentation was intuitive and the self-help module was engaging. Planned: Formal acceptability evaluation in future.	Adolescents reported increased ease of navigation and improved enjoyment over successive versions. Formal usability scales or metrics: NR	NR	Difficult to fulfill all adolescent board requests due to platform limitations and high user expectations; team had to simplify some desired features. During early versions, adolescents identified navigation errors, unintuitive flows, Content length, Internet and device access constraints, and excessive button use.	Human-centered iterative design approach, Multidisciplinary development team such as HIV health care professionals (4 physicians, 1 nurse, and 1 psychologist, service managers, and adolescents). AI development procedure transparency. Connect with a real mental health professional when needed.	Software limitations and early dependence on button-based navigation limited the ability to meet adolescents' preferences for more advanced, AI-like conversational interactions.	Future work: more advanced platforms formally evaluate feasibility and acceptability
7	Study procedures were feasible – 15/28 eligible caregivers enrolled and all completed the chatbot interaction and assessments	Acceptability reported across affective attitude, ethics, adoption, perceived effectiveness, barriers, and self-efficacy	Caregivers found the chatbot easy to use and engaging, supported by visual aids, but some reported emotional discomfort and navigation difficulties, especially among less tech-savvy users.	Exploratory pre–post increase in depression knowledge (ADKQ median 7/14 → 9/14); no statistical testing due to small sample.	Limited digital literacy, dependence on smartphone and internet access, emotional discomfort	NR	Small qualitative sample; not a formal feasibility study; no clinical or health outcomes assessed; caregivers' own mental health not directly evaluated.	Authors recommend larger and longitudinal studies as feasibility effective caregiver- and cultural adaptive components
8	Usability tests were conducted in phase 2 and 5 (Phase 4 had response refinement). All enrolled participants completed the usability sessions (60–90 minutes). Occasional technical issues (e.g., slowness or confusing avatar responses) Overall, intervention is described as feasible.	High satisfaction reported; 93% of participants rated activities positively; AI-facilitated scenarios perceived as relevant and representative. Provided feedback on length and asked for more activities about disclosing to family, dating, and additional information pertaining to HIV and state disclosure laws.	Generally positive user experience: app was easy to navigate, visually appealing, and provided new useful information; engaging content	NR	Typing reduced realism (participants preferred a voice-input option); emotional discomfort with negative AI responses; Activity length concerns, minor technical issues during AI-facilitated role-play, and uncertainty regarding optimal implementation setting and support needs	Theory-based (grounded in Social Cognitive Theory and the Disclosure Processes Model), multi-phase participatory development; AI-facilitated role-play grounded in lived experiences; high user satisfaction	Small sample (n = 8); single geographic location; all participants Black; usability study only; no efficacy outcomes	Evaluate effectiveness, acceptability, multisite feasibility, assess implementation context, support, and dialogue
9	Not assessed	Not assessed	Not assessed	Not assessed	Limitations of response specificity due	Described accurate and comprehensive	Single chatbot (ChatGPT-3.5) evaluated; non-	Evaluating chatbot vs assessing
NR = Not reported								

Study ID	Feasibility (trial completion, recruitment, adherence to chatbot use, technical feasibility)	Acceptability (patient/provider satisfaction, trust, qualitative feedback)	User experience and engagement (User friendly and Usability)	Preliminary efficacy Health Outcomes	Implementation challenges	Strengths	Limitations /Concerns	Recommendations for Future Research
					to lack of individualized context (e.g., pregnancy, geography)	responses and could recognize life-threatening scenarios (e.g., abacavir hypersensitivity)	exhaustive proof-of-concept question set; responses lacked contextual specificity; language and real-world use not assessed	perform across larger and diverse populations; study real-world use PLWH
10	NR	Participants expressed positivity perceived value of chatbots for scheduling and referral, but emphasized that they would not be acceptable for all users	Participants discussed anticipated convenience and raised concerns about potential misunderstandings	NR	Participants raised concerns about potential neglect, lack of personalization, impersonal communication, cultural and language mismatch, technological fallibility, and the need to retain human support	NR	Small stakeholder samples; chatbot was demonstrated but not used by participants; perceptions based on expectations rather than lived use	Co-design stakeholder involvement; tailor language and function; further evaluation of chatbot integration; and integrate with HIV services

NR = Not reported

Qualitative evaluations further confirmed the perceived acceptability of these chatbots. Semi-structured interviews conducted by Rupani et al. (14) yielded positive feedback regarding acceptability, addressing aspects such as affective attitudes, ethics, adoption, perceived effectiveness, barriers, and self-efficacy in using their chatbot, EVA. Similarly, Hightow-Weidman et al. (37) reported high user satisfaction, noting that their chatbot effectively facilitated relevant and representative scenarios while receiving commendable design feedback through qualitative interviews. Koh et al. (41) evaluated the responses of ChatGPT to predefined ART counseling questions. Finally, Galea et al. (23) and Pande et al. (42) planned future assessments of feasibility and acceptability using measures such as the Acceptability of the Intervention Measure, the Intervention Appropriateness Measure, and qualitative feedback from participants.

Several studies have identified important needs and considerations for PLWH when using AI chatbots. Participants expressed a preference for chatbots over human communication, particularly in situations with time constraints. Chatbots can reduce waiting times and assist with scheduling appointments for study assessments (38). Additionally, chatbots can provide referrals for housing and other support services. Other recommendations include shortening the modules to give users more control over accessing specific sections, simplifying the scripts, and ensuring that the information is presented clearly to enhance content retention (13). It has also been suggested to incorporate components for sentiment and emotion detection (42). Lastly, the design of chatbot applications should prioritize simplicity and compatibility with Android devices (40).

Preliminary Efficacy

Formal evaluations of clinical outcomes such as ART adherence, viral suppression, CD4 count, or mental health symptom reduction were limited across the included studies. While some interventions explicitly targeted mental health or depression-related outcomes, most chatbots were designed to increase HIV-related knowledge, support coping strategies, facilitate disclosure decision-making, and improve engagement with healthcare services, rather than deliver HIV or mental health treatment. Studies by Ardiana et al. (40), Vasquez et al. (13), Rupani et al. (14), and Koh et al. (41) provided counseling, psychoeducational content, or clinical advice to PLWH; however, none evaluated their effectiveness in reducing mental health disparities or producing measurable mental health treatment outcomes. This pattern aligns with the field's early-stage nature and highlights opportunities for more rigorous outcome evaluation and more consistent reporting of adoption- and ethics-related domains in future studies.

Hightow-Weidman et al. (37) developed an mHealth application designed to improve disclosure self-efficacy among young men who have sex with men through role-playing scenarios based on social cognitive theory, which resulted in high user satisfaction. Although not designed as a clinical outcome intervention, Rupani et al. (14) reported that use of their chatbot (i.e., EVA) led to increased depression knowledge, as measured by the adolescent depression knowledge questionnaire (ADKQ) (median score increased from 7/14 to 9/14), but did not assess its impact on depression clinical outcomes.

Similarly, Koh et al. (41) found that ChatGPT provided accurate and comprehensive responses, including recognition of potentially life-threatening scenarios such as abacavir hypersensitivity reaction and appropriate advice. However, in certain contexts, such as specific geographic locations or for pregnant individuals, the advice lacked sufficient specificity and may have been inadequate. The study did not examine whether ChatGPT improved ART outcomes among PLWH, as this was not listed as an objective.

Implementation Challenges

Implementation challenges were frequently reported and may affect effectiveness and scalability. These challenges included: (a) system errors in responding to user queries due to insufficient knowledge databases (40); (b) limited conversation topics, particularly regarding lifestyle and behavioral factors such as diet, updates on HIV treatment and related diseases, reproductive and sexual health including pregnancy, healthcare service support such as appointments, mental health support, and socioeconomic issues like immigration processes (19); (c) impersonal and artificial dialogue with restricted emotional range, navigation errors, and discomfort when collecting information about gender or phone number (13); (d) inability to address all requests from adolescent advisory boards due to platform constraints, concerns about activity length, and emotional discomfort resulting from negative AI responses (13, 37); (e) cultural and language mismatches, potential neglect, potential for personal communication to lead to errors in automated scheduling and referrals, and lack of personalization (38); and (f) limited digital literacy of users, reliance on smartphones and internet access, and emotional discomfort (14).

To mitigate these challenges, one study updated the chatbot's knowledge database to improve system intelligence and testing accuracy when incorrect answers were provided (40). Additionally, the integration of hybrid dialog management, affect-aware responses, and human-in-the-loop escalation enabled the management of sensitive or potentially problematic conversational scenarios (42). In another study, if a user enters a name that is too short (≤ 2 characters) or numeric, EVA prompts the user to provide a valid name (13). To support users' mental health, links to relevant services were incorporated to increase access to virtual and in-person emotional support resources (13, 23, 37).

Study report standards: Framework Alignment and Reporting Observations

Consistent with the framework-informed methods, the extracted findings map to key reporting and implementation domains emphasized across SPIRIT-AI/CONSORT-AI (intervention description, human–AI interaction, handling of inputs/outputs and potential errors), TEHAI (capability, utility, adoption), and WHO ethical principles (privacy, safety, and inclusiveness). Across the included studies, most reported substantial details on intended use, user-centered design processes, and usability/acceptability, while comparatively fewer studies reported evaluation of downstream health outcomes or provided evidence of real-world integration and sustained implementation. In particular, access and re-usability (SPIRIT-AI Item 29) were not discussed, and harms (SPIRIT-AI Item 22) were rarely examined. However, it should be noted that Ma et al. (19) explicitly reported their study results in accordance with the CONSORT-AI checklist.

Reporting on performance errors was limited and addressed explicitly in only a small number of studies (13, 40, 42). In contrast, Koh et al. (41) noted that, in certain contexts, such as specific geographic settings or among pregnant individuals, ChatGPT-generated advice lacked sufficient specificity to account for individuals' unique clinical circumstances and may therefore have been inadequate. Nevertheless, the system consistently redirected users to seek guidance from healthcare professionals to obtain individualized, context-appropriate advice.

In general, based on the findings of this review, Hightow-Weidman et al. (43), Rupani et al. (14), and Ma et al. (19) addressed a broad range of study checklist domains, particularly those related to user-centered design, usability, acceptability, and ethical considerations. However, evaluation of health outcomes and intervention effectiveness was largely absent, as improving clinical outcomes was not the primary objective of these studies.

DISCUSSION

Principal Findings

This review synthesized evidence on AI-enabled chatbot interventions supporting health outcomes among PLWH and examined their alignment with emerging AI evaluation frameworks. Across the ten included studies, AI-enabled chatbot studies demonstrated substantial heterogeneity in technological sophistication, target populations, and implementation contexts, ranging from rule-based counseling tools to generative AI systems integrated with HIV self-management support platforms.

A key finding was that across the included studies, chatbots were primarily designed to provide psychoeducation, behavioral self-management support, stigma reduction, engagement in care, and ART-related counseling rather than delivering or evaluating treatment-level interventions. When evaluated through SPIRIT-AI, CONSORT-AI, TEHAI and WHO framework domains, current chatbots demonstrate strong performance in usability and user-centered design but limited reporting of harms, performance monitoring, and real-world adoption readiness. Collectively, our findings indicate that chatbot interventions are evolving rapidly but remain heterogeneous in technological maturity and implementation readiness.

Mechanisms of Action

Our findings highlighted that AI-enabled chatbot interventions are increasingly being designed to target psychosocial determinants of HIV health outcomes, which remain major barriers to sustained treatment engagement and psychological well-being among PLWH. Chatbots may be particularly suited for addressing these determinants because they provide anonymous, non-judgmental, and continuously accessible communication environments. Such features may facilitate disclosure of sensitive experiences, including stigma and mental health symptoms, which are often underreported in traditional clinical encounters. However, improvements in clinical mental health outcomes were rarely evaluated, reflecting that most chatbots were designed to support education and self-management rather than deliver formal therapeutic treatment. For example, one study reported improvements in depression-related knowledge but did not assess clinical depression outcomes (14).

Similarly, despite promising engagement outcomes, few studies evaluated clinical outcomes such as ART adherence, viral suppression, and CD4 count. The limited evaluation of clinical outcomes indicates that chatbot research in HIV care remains largely developmental, with innovation advancing more rapidly than clinical validation; a pattern commonly observed during early translational phases of digital mental health technologies. This evidence gap suggests that the clinical impact of chatbot interventions on HIV-related health outcomes remains uncertain and might be through multiple behavioral and psychosocial mechanisms. As chronic disease chatbot studies suggest that current chatbot interventions may function primarily as educational or behavioral support tools rather than as fully evaluated clinical tools (45, 46).

Implementation Facilitators and Barriers

A notable strength across included studies was the consistent use of user-centered design approaches. Several chatbot studies involved PLWH, caregivers, healthcare providers, and community advisory boards in the development and refinement of chatbot content and functionality. Such participatory approaches are widely recognized as critical for improving digital health intervention engagement and acceptability. Privacy protection emerged as another important strength, where most chatbot interventions deliberately minimized the collection of identifiable data and, in some cases, allowed users to delete conversation records (19), which is particularly important for PLWH, who often face stigma and confidentiality concerns. This aligns with prior healthcare chatbot literature, which emphasizes data security and transparency as central determinants of user trust and adoption (33, 45, 47).

However, emerging AI technologies introduce new safety challenges as well. For example, one included study found that AI-generated ART counseling responses were generally accurate but occasionally lacked sufficient contextual specificity in certain clinical situations, including pregnancy-related counseling (41). Prior reviews similarly highlighted risks related to algorithmic hallucination, misinformation, and limited capacity to address complex clinical decision-making (46, 48). Technical limitations also included incomplete knowledge databases, restricted conversational scope, and navigation errors. These technical and conversational limitations highlight a critical implementation challenge: while chatbots can deliver scalable support, current systems may struggle to replicate the contextual understanding and emotional nuance required for sustained mental health engagement.

Additionally, users reported concerns related to impersonal or emotionally limited dialogues, highlighting the difficulty of replicating empathic human interaction through automated systems. Structural barriers were also reported, including digital literacy limitations, reliance on smartphone access, and internet connectivity challenges. Further, cultural and linguistic mismatches were identified as potential risks that could affect user engagement and intervention effectiveness. Together, these findings emphasize the need for culturally tailored chatbot design, inclusive development strategies, and attention to digital equity to support successful real-world implementation.

To address these concerns, several studies proposed mitigation strategies, including updating chatbot knowledge bases, integrating human-in-the-loop escalation pathways for sensitive conversations, and incorporating affect-aware response systems. In this regard, advanced AI models, particularly LLMs, offer opportunities for personalized and context-aware health communication (49). However, these models may also introduce risks related to inaccurate or contextually inappropriate responses, particularly in complex clinical scenarios.

In light of these findings, healthcare systems adopting AI-enabled chatbot interventions should prioritize clinical validation, performance monitoring, and human oversight mechanisms to ensure safe deployment, particularly when chatbots provide medication counseling or mental health guidance. In this regard, hybrid chatbot models that incorporate human-in-the-loop escalation pathways and affect-aware responses may help mitigate safety and personalization challenges (45, 50).

Alignment With AI Governance Frameworks

Our review identified variability in reporting transparency and safety evaluation across included studies. Most studies provided detailed descriptions of intervention purpose, design processes, and usability outcomes, aligning with key domains outlined in SPIRIT-AI and CONSORT-AI reporting frameworks. However, several critical domains were inconsistently reported. Notably, few studies evaluated harms or unintended consequences associated with chatbot use, and access to intervention algorithms or reusable technical components was rarely described. The limited reporting of harms suggests that safety evaluation has not yet kept pace with technological innovation, emphasizing the need for standardized governance frameworks to guide the responsible deployment of AI-supported healthcare tools.

Additionally, evidence supporting long-term clinical implementation and real-world integration into HIV care systems was limited, indicating incomplete alignment with adoption domains of the TEHAI framework. From an ethical perspective and alignment with the WHO framework, most interventions demonstrated attention to privacy protection and inclusiveness. However, evaluation of algorithmic bias, safety during mental health crises, and clinical oversight remained limited. Although some studies involved healthcare professionals, these safeguards were not systematically assessed, highlighting important areas requiring further investigation before widespread clinical deployment.

One possible explanation for the inconsistent reporting of risks and unintended consequences is that several included studies were designed primarily as development studies or protocols for future implementation rather than as evaluations of fully deployed interventions. As a result, many studies prioritized chatbot design and functionality over systematic assessment of safety, implementation outcomes, or potential harms. This suggests that chatbot research in HIV care remains in an early translational phase, where feasibility has preceded comprehensive implementation evaluation. Although few studies reported risks, the available findings provide important insights to inform safer and more responsible future development. Building upon this, policymakers and research funders should prioritize the adoption of standardized AI reporting and evaluation frameworks to ensure transparency, accountability, and equitable implementation of chatbot technologies.

Strengths and Limitations of This Review

This review represents one of the first systematic syntheses specifically evaluating AI-enabled chatbot interventions targeting health outcomes among PLWH. The integration of multiple AI evaluation frameworks enabled a comprehensive assessment of technical, clinical, ethical, and implementation domains. Additionally, inclusion of studies across multiple countries and populations provided a multidisciplinary and global perspective. However, several limitations should be acknowledged. The number of eligible studies was small, varied substantially in design, target outcomes, and evaluation methodologies, limiting direct comparison. Many studies were formative, pilot, or usability-focused and therefore did not aim to evaluate clinical effectiveness. Moreover, rapid technological evolution in AI and chatbot development may limit the long-term generalizability of current findings. Furthermore, the inclusion of only English-language papers may have led to the omission of valuable insights that could have further enriched the findings.

Future Directions

The findings of this review highlight several critical priorities for future research. First, rigorous evaluation of clinical outcomes is needed to determine whether chatbot interventions can improve ART adherence, viral suppression, and psychological well-being among PLWH. To address this, longitudinal and randomized controlled trial designs will be essential to establish effectiveness and cost-effectiveness. Second, future research should incorporate standardized AI reporting guidelines to improve transparency, reproducibility, and safety evaluation. Greater attention to algorithm performance monitoring, harm reporting, and bias mitigation will be necessary to ensure equitable deployment.

Moreover, implementation science approaches are needed to evaluate integration into routine HIV care, including sustainability, scalability, cost-effectiveness, workforce impact, and equity implications across diverse healthcare settings. Additionally, clinicians may play a critical role in the development and implementation of chatbot interventions through content validation, cultural adaptation, and patient-centered design processes. Engaging healthcare providers in chatbot development and developing solid knowledge base may improve clinical accuracy, increase patient trust, and facilitate integration into routine HIV care workflows.

Clinical and Public Health Implications

HIV represents a chronic condition requiring lifelong treatment adherence, ongoing self-management, and sustained behavioral engagement (51). Several reviewed chatbot interventions were designed to provide medication counseling, self-management support, and behavioral reinforcement strategies that align with the long-term care trajectory experienced by PLWH. This is consistent with broader chronic disease chatbot literature demonstrating that conversational agents can support long-term self-monitoring, behavior modification, and patient engagement in disease management (45–47, 50). Accordingly, chatbots may provide a personalized and private communication environment that facilitates disclosure of stigmatized experiences and supports individuals who may otherwise avoid seeking care or discussing mental health concerns (33, 52). These preferences are highly relevant within the context of HIV care, where stigma and disclosure-related stress remain significant barriers to engagement in treatment and psychosocial support (6).

Chatbots may also support healthcare workforce capacity by delivering low-intensity behavioral interventions, facilitating patient triage, and supporting routine counseling tasks, while allowing healthcare providers to focus on complex clinical care needs, which aligns with prior chronic disease chatbot research (33, 50). At a systems level, chatbot interventions may offer scalable solutions to improve access to HIV and mental health services, particularly in resource-limited settings. By providing continuous, stigma-sensitive support, chatbot technologies may help address disparities in healthcare access, including barriers related to geography, cost, and limited availability of healthcare specialists. Policymakers should also consider digital equity when implementing chatbot interventions. Additionally, several studies identified barriers related to digital literacy, smartphone access, and internet connectivity, suggesting that without targeted inclusion strategies, chatbot technologies may inadvertently widen healthcare disparities. Supporting multilingual, culturally tailored, and accessible chatbot design will be essential to ensure equitable implementation across diverse populations affected by HIV.

However, policymakers and healthcare organizations should approach chatbot implementation cautiously and prioritize regulatory oversight, safety monitoring, and ethical governance. It is also noteworthy that clinicians should consider chatbot interventions as supportive tools rather than replacements for human care. Several studies highlighted limitations in chatbot emotional responsiveness and contextual clinical decision-making, indicating that human oversight remains essential, particularly when addressing mental health crises or complex ART-related counseling. Integrating chatbot systems within collaborative, hybrid care models, where automated systems provide initial support and escalate high-risk cases to healthcare professionals, may enhance both safety and scalability.

CONCLUSION

AI-enabled chatbot interventions are a promising and rapidly evolving approach to supporting PLWH health outcomes. Our findings showed that these chatbots are feasible and acceptable, with the potential to address psychosocial and self-management needs, which suggests that chatbot-based interventions may address key engagement barriers in HIV care, particularly stigma and reluctance to discuss sensitive mental health concerns in traditional clinical settings. However, there is still limited robust evidence demonstrating improvements in clinical or mental health outcomes. Our review supports the idea that chatbot interventions should serve as supplemental tools rather than substitutes for clinical care. Future research should focus on rigorous outcome evaluations, ethical governance, and implementation science to ensure the safe and equitable integration of chatbot technologies into HIV care.

Declarations

Acknowledgements: We would like to acknowledge Zoey Babyak for her valuable contributions during the screening process of this review study.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding sources: This work is funded by the National Institute of Health (NIH/NIAID) under the award numbers R01174892 (PI: Shan Qiao) and R56AI174897 (PI: Xiaoming Li).

Author Contributions: **AP** (Conceptualization, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review and editing), **MJ** (Investigation, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review and editing), **AI** (Data curation, Formal analysis), **XL** (Supervision, Writing – review and editing), **SQ** (Conceptualization, Supervision, Project administration, Methodology, Writing – review and editing).

Supplementary Information: The manuscript contains supplementary materials that are available in the supplementary document.

Ethics approval: No ethics approval is needed. This article does not contain any studies with human participants performed by any of the authors.

Informed consent: Informed consent is not required since this work is a systematic review.

References

1. UNAIDS, Global HIV & AIDS statistics — Fact sheet 2025 2025 [Available from: <https://www.unaids.org/en/resources/fact-sheet>
2. Wandeler G, Johnson LF, Egger M (2016) Trends in life expectancy of HIV-positive adults on antiretroviral therapy across the globe: comparisons with general population. *Curr Opin HIV AIDS* 11(5):492–500
3. Lazarus JV, Safreed-Harmon K, Kamarulzaman A, Anderson J, Leite RB, Behrens G et al (2021) Consensus statement on the role of health systems in advancing the long-term well-being of people living with HIV. *Nat Commun* 12(1):4450
4. Lazarus JV, Wohl DA, Cascio M, Guaraldi G, Rockstroh J, Hodson M et al (2023) Long-term success for people living with HIV: a framework to guide practice. *HIV Med* 24:8–19
5. UNAIDS Understanding measures of progress towards the 95-95-95 HIV testing, treatment and viral suppression targets 2025 [Available from: https://www.unaids.org/sites/default/files/media_asset/progress-towards-95-95-95_en.pdf
6. Pasha A, Qiao S, Zhang J, Cai R, He B, Yang X et al (2025) Changes in mental health care utilisation before and during the COVID-19 pandemic among people living with HIV in the USA: a retrospective cohort study using the All of Us dataset. *BMJ Public Health* 3(2). <https://doi.org/10.1136/bmjph-2024-002173>
7. Eke AC, Eke UA (2025) Beyond Viral Suppression: Navigating Structural Barriers, Aging and Frailty, Drug Resistance, Therapeutic Innovations, and Reproductive Health Challenges in the Global HIV/AIDS Epidemic. *J AIDS HIV Treat* 7(1):108
8. Tam CC, Harrison SE, Benotsch EG, Litwin AH, Zhou Y, Shen Z et al (2024) Psychological distress mediates the relationship between HIV-related stigma and prescription opioid misuse among Chinese people living with HIV. *AIDS Behav* 28(5):1673–1683
9. Pasha A, Jahanaray M, Li X, Qiao S (2025) Body Image and Its Associated Factors among People Living with HIV: A Scoping Review and Implications for Integrated Care. *AIDS Behav*. <https://doi.org/10.1007/s10461-025-04991-6>
10. Engler K, Lenart A, Lessard D, Toupin I, Lebouche B (2018) Barriers to antiretroviral therapy adherence in developed countries: a qualitative synthesis to develop a conceptual framework for a new patient-reported outcome measure. *AIDS Care* 30(sup1):17–28
11. Casu M, Triscari S, Battiato S, Guarnera L, Caponnetto P (2024) AI chatbots for mental health: a scoping review of effectiveness, feasibility, and applications. *Appl Sci* 14(13):5889
12. Li J (2023) Digital technologies for mental health improvements in the COVID-19 pandemic: a scoping review. *BMC Public Health* 23(1):413
13. Vasquez DH, Rupani N, Contreras C, Kolevic L, Franke MF, Galea JT Building EVA: Human-Centered Development of a Mental Health Chatbot for Adolescents Living with HIV: Tutorial
14. Rupani N, Vasquez DH, Contreras C, Menacho L, Kolevic L, Franke MF et al (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. *J Int Association Providers AIDS Care (JIAPAC)* 24:23259582251327911
15. Sanabria G, Greene KY, Tran JT, Gilyard S, DiGiovanni L, Emmanuel PJ et al (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. *J Technol Behav Sci* 8(4):382–391
16. MacNeill AL, Doucet S, Luke A (2024) Effectiveness of a mental health chatbot for people with chronic diseases: randomized controlled trial. *JMIR Formative Res* 8:e50025
17. Inkster B, Sarda S, Subramanian V (2018) An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth uHealth* 6(11):e12106
18. Cheah MH, Gan YN, Altice FL, Wickersham JA, Shrestha R, Salleh NAM et al (2024) Testing the feasibility and acceptability of using an artificial intelligence Chatbot to promote HIV testing and pre-exposure prophylaxis in Malaysia: mixed methods study. *JMIR Hum Factors* 11:e52055
19. Ma Y, Achiche S, Tu G, Vicente S, Lessard D, Engler K et al (2025) The first AI-based Chatbot to promote HIV self-management: A mixed methods usability study. *HIV Med* 26(2):184–206
20. Chen S, Zhang Q, Chan C-k, Yu F-y, Chidgey A, Fang Y et al (2023) Evaluating an innovative HIV self-testing service with web-based, real-time counseling provided by an artificial intelligence chatbot (HIVST-Chatbot) in increasing HIV self-testing use among Chinese men who have sex with men: protocol for a noninferiority randomized controlled trial. *JMIR Res Protocols* 12(1):e48447
21. Massa P, de Souza Ferraz DA, Magno L, Silva AP, Greco M, Dourado I et al (2023) A transgender chatbot (Amanda Selfie) to create pre-exposure prophylaxis demand among adolescents in Brazil: assessment of acceptability, functionality, usability, and results. *J Med Internet Res* 25:e41881
22. Ni Z, Oh S, Saifi R, Azwa I, Altice FL (2025) Evaluating the Usability of an HIV Prevention Artificial Intelligence Chatbot in Malaysia: National Observational Study. *JMIR Hum factors* 12:e70034
23. Galea JT, Vasquez DH, Rupani N, Gordon MB, Tapia M, Greene KY et al (2024) Development and Pilot-Testing of an Optimized Conversational Agent or Chatbot for Peruvian Adolescents Living With HIV to Facilitate Mental Health Screening, Education, Self-Help, and Linkage to Care: Protocol for a Mixed Methods, Community-Engaged Study. *JMIR Res Protoc* 13:e55559

24. Ngcobo S, Mntla EM, Shock J, Louw M, Mbonambi L, Serite T et al (2025) Artificial intelligence for HIV care: a global systematic review of current studies and emerging trends. *J Int AIDS Soc* 28(10):e70045
25. Fitzpatrick KK, Darcy A, Vierhile M (2017) Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4(2):e7785
26. Vilaza GN, McCashin D (2021) Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. *Front Digit Health* 3:689736
27. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ, Ashrafian H et al (2020) Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2(10):e549–e60
28. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H et al (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2(10):e537–e48
29. Reddy S, Rogers W, Makinen V-P, Coiera E, Brown P, Wenzel M et al (2021) Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ health care Inf* 28(1):e100444
30. Organization WH Ethics and governance of artificial intelligence for health 2021 [Available from: <https://www.who.int/publications/i/item/9789240029200>]
31. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB (2019) Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 64(7):456–464
32. Marcus JL, Sewell WC, Balzer LB, Krakower DS (2020) Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. *Curr HIV/AIDS Rep* 17(3):171–179
33. Van Heerden A, Bosman S, Swendeman D, Comulada WS (2023) Chatbots for HIV prevention and care: a narrative review. *Curr HIV/AIDS Rep* 20(6):481–486
34. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst reviews* 4(1):1
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. *Syst reviews* 5(1):210
36. Asaeikheybari G, Hughart C, Gupta D, Avery A, Step MM, Smith JM et al (eds) (2020) Precision HIV Health App, Positive Peers, Powered by Data Harnessing, AI, and Learning. 2020 Second International Conference on Transdisciplinary AI (TransAI); 21–23 Sept. 2020
37. Hightow-Weidman LB, Muessig K, Soberano Z, Rosso MT, Currie A, Adams Larsen M et al (2022) Tough Talks Virtual Simulation HIV Disclosure Intervention for Young Men Who Have Sex With Men: Development and Usability Testing. *JMIR Form Res* 6(9):e38354
38. Comulada WS, Rezai R, Sumstine S, Flores DD, Kerin T, Ocasio MA et al (2024) A necessary conversation to develop chatbots for HIV studies: qualitative findings from research staff, community advisory board members, and study participants. *AIDS Care* 36(4):463–471
39. Rupani N, Vasquez DH, Contreras C, Menacho L, Kolevic L, Franke MF et al (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. *J Int Assoc Provid AIDS Care* 24:23259582251327911
40. Ardiana D, Joni I B, Udayana D I, (eds) (2020) Mobile based chatbot application for HIV/AIDS counseling using artificial intelligence markup language approach. *Journal of Physics: Conference Series*; : IOP Publishing
41. Koh MCY, Ngiam JN, Yong J, Tambyah PA, Archuleta S (2024) The role of an artificial intelligence model in antiretroviral therapy counselling and advice for people living with HIV. *HIV Med* 25(4):504–508
42. Pande C, Martin A, Pimmer C (2023) Towards hybrid dialog management strategies for a health coach chatbot
43. Hightow-Weidman LB, Muessig K, Soberano Z, Rosso MT, Currie A, Larsen MA et al (2022) Tough Talks Virtual Simulation HIV Disclosure Intervention for Young Men Who Have Sex With Men: Development and Usability Testing. *JMIR FORMATIVE Res.* ;6(9)
44. Asaeikheybari G, Hughart C, Gupta D, Avery A, Step MM, Smith JM et al (eds) (2020) Precision hiv health app, positive peers, powered by data harnessing, ai, and learning. Second International Conference on Transdisciplinary AI (TransAI); 2020: IEEE
45. Wah JNK (2025) Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. *Front Public Health* 13:1530799
46. Aggarwal A, Tam CC, Wu D, Li X, Qiao S (2023) Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res* 25:e40789
47. Grassini E, Buzzi M, Leporini B, Vozna A (2025) A systematic review of chatbots in inclusive healthcare: insights from the last 5 years. *Univ Access Inf Soc* 24(1):195–203
48. Kim H-K (ed) (2024) The effects of artificial intelligence chatbots on women's health: A systematic review and meta-analysis. MDPI, Healthcare
49. Maity S, Saikia MJ (2025) Large Language Models in Healthcare and Medical Applications: A Review. *Bioengineering* 12(6):631
50. Kurniawan MH, Handiyani H, Nuraini T, Hariyati RTS, Sutrisno S (2024) A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. *Ann Med* 56(1):2302980
51. Qiao S, Aggarwal A, Garrett C, N'Diaye A, Pasha A, Esu I et al (2025) Sleep quality, emotional moods, and cognitive outcomes among people living with HIV: An ecological momentary assessment study. *PLoS ONE* 20(9):e0329399. <https://doi.org/10.1371/journal.pone.0329399>
52. Barreda M, Cantarero-Prieto D, Coca D, Delgado A, Lanza-León P, Lera J et al (2025) Transforming healthcare with chatbots: Uses and applications—A scoping review. *Digit Health* 11:20552076251319174

Figures

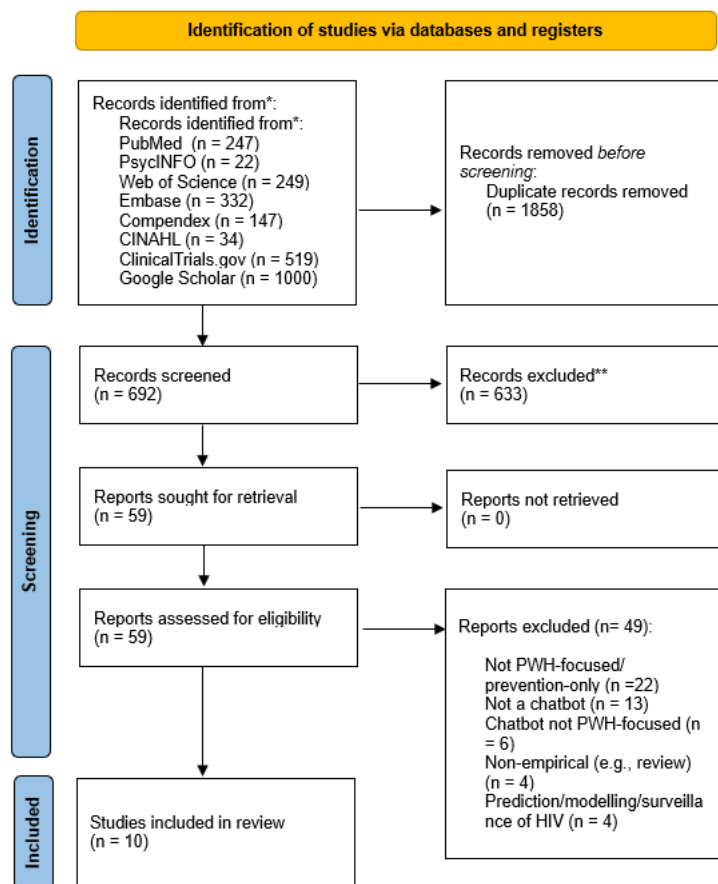


Figure 1

Flow diagram of the literature search and articles selection (adapted from PRISMA 2020 guidelines for systematic reviews).

Supplementary Files

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

- [SupplementaryFile.docx](#)
- [APPENDIX.docx](#)