Current Landscape and Future Directions for Mental Health Conversational Agents (CAs) for Youth: Scoping Review

Jinkyung Katie Park, Vivek Singh, Pamela Wisniewski

Submitted to: JMIR Medical Informatics
on: May 30, 2024

Table of Contents

Current Landscape and Future Directions for Mental Health Conversational Agents (CAs) for Youth: Scoping Review

Jinkyung Katie Park[1] PhD; Vivek Singh[2] PhD; Pamela Wisniewski[1] PhD

[1]Vanderbilt University Nashville US
[2]Rutgers University New Brunswick US

Corresponding Author:
Jinkyung Katie Park PhD
Vanderbilt University
2201 West End Ave
Nashville
US

Abstract

Background: Conversational Agents (CAs, chatbots) are systems enabled with the ability to interact with the users using natural human dialogue. They are increasingly used to support interactive knowledge discovery of sensitive topics such as mental health topics. While much of the research on CAs for mental health has focused on adult populations, the insights from such research may not apply to CAs for youth.

Objective: The aim of the study was to comprehensively evaluate the state-of-the-art research on mental health CAs for youth.

Methods: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we synthesized 39 peer-reviewed studies specific to mental health CAs designed for youth. We conducted a scoping review of the literature to evaluate the characteristics of research on mental health CAs designed for youth, the design/computational considerations of mental health CAs for youth, and the evaluation outcomes reported in the research on mental health CAs for youth.

Results: We found that most mental health CAs were designed as older peers to provide therapeutic and/or educational content to promote youth mental well-being. Most of the CAs were designed based on expert knowledge, with a few that incorporated inputs from youth. The technical maturity of CAs was in its infancy, focusing on building prototypes with rule-based models to deliver pre-written content, with limited safety features to respond to imminent risk. Research findings suggest that while youth appreciate the 24/7 availability of friendly/empathetic conversation on sensitive topics with CAs, they found the content provided by CAs to be limited. Finally, we found a concerning trend that most of the reviewed studies did not address the ethical aspects of mental health CAs while youth were concerned about the privacy and confidentiality of their sensitive conversation data.

Conclusions: Our study highlights the need for researchers to continue to work together to align evidence-based research on mental health CAs for youth with lessons learned on how to best deliver these technologies to youth. Our review brings to light mental health CAs needing further development and evaluation. The new trend of large language models (LLMs) based CAs can make such technologies more feasible. Yet, the privacy and safety of the systems should be prioritized. Although preliminary evidence shows positive trends in mental health CAs, long-term evaluative research with larger sample sizes and robust research designs is needed to validate their efficacy. More importantly, collaborative efforts with youth and clinical experts are needed from early design phases to summative evaluation stages to build safe, effective, and youth-centered mental health CAs. Finally, best practices for risk mitigation and ethical development of CAs with and for youth are needed to promote their mental well-being. Clinical Trial: N/A

(JMIR Preprints 30/05/2024:62758)
DOI: https://doi.org/10.2196/preprints.62758

Preprint Settings

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

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

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

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Review Paper

Current Landscape and Future Directions for Mental Health Conversational Agents (CAs) for Youth: Scoping Review

ABSTRACT

Introduction: Conversational Agents (CAs, chatbots) are systems enabled with the ability to interact with the users using natural human dialogue. They are increasingly used to support interactive knowledge discovery of sensitive topics such as mental health topics. While much of the research on CAs for mental health has focused on adult populations, the insights from such research may not apply to CAs for youth. The aim of the study was to comprehensively evaluate the state-of-the-art research on mental health CAs for youth.

Methods: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we synthesized 39 peer-reviewed studies specific to mental health CAs designed for youth. We conducted a scoping review of the literature to evaluate the characteristics of research on mental health CAs designed for youth, the design/computational considerations of mental health CAs for youth, and the evaluation outcomes reported in the research on mental health CAs for youth.

Results: We found that most mental health CAs were designed as older peers to provide therapeutic and/or educational content to promote youth mental well-being. Most of the CAs were designed based on expert knowledge, with a few that incorporated inputs from youth. The technical maturity of CAs was in its infancy, focusing on building prototypes with rule-based models to deliver pre-written content, with limited safety features to respond to imminent risk. Research findings suggest that while youth appreciate the 24/7 availability of friendly/empathetic conversation on sensitive topics with CAs, they found the content provided by CAs to be limited. Finally, we found a concerning trend that most of the reviewed studies did not address the ethical aspects of mental health CAs while youth were concerned about the privacy and confidentiality of their sensitive conversation data.

Discussion: Our study highlights the need for researchers to continue to work together to align evidence-based research on mental health CAs for youth with lessons learned on how to best deliver these technologies to youth. Our review brings to light mental health CAs needing further development and evaluation. The new trend of large language models (LLMs) based CAs can make such technologies more feasible. Yet, the privacy and safety of the systems should be prioritized. Although preliminary evidence shows positive trends in mental health CAs, long-term evaluative research with larger sample sizes and robust research designs is needed to validate their efficacy. More importantly, collaborative efforts with youth and clinical experts are needed from early design phases to summative evaluation stages to build safe, effective, and youth-centered mental health CAs. Finally, best practices for risk mitigation and ethical development of CAs with and for youth are needed to promote their mental well-being.

Keywords: conversational agent; chatbot; mental health; youth; adolescent; scoping review, PRISMA

Introduction

Background

Conversational Agents (CAs, often called chatbots) are systems with the ability to interact with the users using natural human dialogue [1]. Examples of CAs range from customer service chatbots that

are available on commercial websites and social media platforms to open domain text-based chatbots such as GPT4 and Bing, and voice assistants such as Apple's Siri and Amazon's Alexa. Driven by advances in the underlying language models, conversational agents enable two-way interactive communication with the user and have been applied in multiple domains including healthcare [2]. Particularly, CAs are seen as an innovative digital medium to communicate information and resources with younger users, given their high digital literacy and familiarity with chat applications [3]. Now, CAs are increasingly used by youth for interactive knowledge discovery on sensitive topics, including mental health topics [4].

Youth are in a unique transitional phase between childhood and adulthood.[1] Recent reports show that youth are increasingly experiencing mental health issues these days. For instance, from 2009 to 2019, the proportion of high school students reporting persistent feelings of sadness or hopelessness increased by 40%; between 2007 and 2018, suicide rates among youth ages 10-24 in the US increased by 57% [5]. Yet, they are hesitant to seek professional help on mental health topics due to societal views towards the topics themselves and perceived public stigma and embarrassment associated with help-seeking in those topics [6]. With the ability for humanlike interactions with the user, mental health CAs have been developed to support their unique informational, educational, and therapeutic needs related to mental health topics.

Existing systematic reviews on mental health CAs shed light on the potential of CAs to provide relevant information and resources via interactive communication. For instance, the review of 53 studies found the most common use of mental health CAs to be delivering pre-written therapeutic and training content for people with depression and autism [7]. Another systematic review of 13 studies on the outcome of mental health treatment delivered by mental health CAs found reductions in psychological distress after interacting with the mental health CAs [8]. Yet, a meta-analysis of 12 studies on mental health CAs found conflicting results regarding the effect of chatbots on the severity of anxiety and positive and negative affect [9]. A more recent analysis of 32 randomized controlled trials confirmed that CA-based mental health interventions are effective in improving various mental health conditions in the short term, while significant long-term effects were not observed [10]. Another meta-analysis of 35 studies on Artificial Intelligence (AI)-based mental health CAs showed a significant reduction in symptoms of depression and distress, with no significant improvement in overall psychological well-being [11]. As such, the trends of the clinical effectiveness of mental health CAs in the existing literature are inconclusive.
A scoping review of 37 studies focused on user perceptions toward mental health CAs showed overall positive opinions toward the CAs such as usefulness and ease of use. At the same time, they found conversations with CAs to be shallow, confusing, or brief [12]. There are some systematic reviews on mental health CAs to support people with specific mental health conditions such as substance use disorder [13], depression and anxiety [14], and serious mental illness [15]. For instance, a systematic review of 7 studies involving CAs for assessing serious mental health illness (i.e., major depressive disorder, schizophrenia spectrum disorder, bipolar disorder, and anxiety disorder) found generally positive outcomes regarding conversational agents' diagnostic quality, therapeutic efficacy, and acceptability. Yet, they showed that there continues to be a lack of standard measures for evaluating conversational agents as well as lacked representation of the pediatric population [15]. As such, reviews on mental health CAs are well-established and provide important insights into the benefits and pitfalls of existing research on mental health CAs. *However, little work has been done to explore trends in research on mental health CAs for younger populations such as*

[1]The United Nation (UN) defines youth as those between the ages of 15 and 24 years, which comprise of 18% of the total world population. The UN Secretariat uses the terms youth and young people interchangeable with the understanding that member states and other entities use different definitions. Following the definition provided by the UN, in this review, we use the term "youth" to refer to adolescents and young adults aged between 15-24.

adolescents and youth.

Recently, Balan et al. conducted a scoping review of 25 studies on CAs designed to improve the emotional components of mental health (e.g., depression, anxiety) of the young population [16] and found that while feasibility/usability outcomes are optimistic, the clinical effectiveness of CA-based mental health interventions remains inconclusive. While trends in therapeutic CAs to improve the emotions of young populations were studied, a comprehensive trend in research on mental health CAs with various goals (e.g., informational, assessment) or social, behavioral, or cognitive aspects of mental health has not been studied in prior work. In addition, while trends in the technology and clinical efficacy of mental health CAs for the young population were addressed in prior work, trends in design aspects (e.g., CA role/characteristics) and evaluation outcomes beyond efficacy (e.g., strength/weaknesses of CAs, ethics) were not addressed. Therefore, to fill the gap in the literature, we conducted a scoping review of 39 studies that focused on CAs to promote the mental health of youth. In this paper, we address the following research questions:

RQ1: What are the characteristics of empirical research on mental health CAs designed for youth?
RQ2: What are the (a) design and (b) computational considerations for mental health CAs for youth?
RQ3: What are the evaluation outcomes reported in empirical research on mental health CAs for youth?

Objectives

The objective of this study was to synthesize the current literature on mental health CAs designed for youth to understand the trends in research, the design and computational aspects of the CAs, and the strengths and weaknesses of the current mental health CAs for youth. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement guidelines [17], we conducted a comprehensive review of existing literature on mental health CAs for youth. We describe this process in the sections that follow.

Methods

Systematic Literature Search Process

We initially searched the literature with the terms (“conversational agent” OR “chatbot”) to explore synonyms of those terms used in the literature. Our initial search informed us of the various alternative terms used to describe conversational agents or chatbots, which allowed for a more inclusive and thorough search. Our final search string consisted of the keywords: (“conversational agent” OR “chatbot” OR “virtual agent” OR “virtual assistant” OR “AI assistant” OR “AI bot” OR “social bot”) AND (teen OR adolescent OR youth OR young) AND (mental). Then, we identified four relevant and cross-disciplinary databases that included research on conversational agents in the healthcare domain, including ProQuest, Scopus, Web of Science, and PubMed. The same search string was used to retrieve articles across three databases. The searches were limited to journal articles, conference papers, and book chapters written in English. The publication date was not specified. The initial search resulted in retrieving 315 articles from three databases (Web of Science: 59, ProQuest: 31, Scopus: 82, and PubMed: 49) in February 2024.

Inclusion and Exclusion Criteria

The purpose of this work was to review mental health CAs designed for youth. Therefore, we included full research papers that 1) were peer-reviewed (Journal articles and refereed conference

proceedings were both included), 2) discussed CAs for providing information/resources/support on mental health topics, 3) discussed CAs designed for youth, adolescents, or young adults, 4) described CAs that permitted two-way interactions that were fully automated (without human mediation), 5) included empirical results.

We excluded papers that 1) were non-full or non-reviewed such as works in progress, extended abstracts, reports, reviews, and meta-analyses, 2) did not include mental health topics (e.g., physical health), 3) included forms of one-way communication and human-mediated communication, 4) did not consider youth population, 5) did not primarily focus on the CAs (e.g., explored CAs as one feature of the mobile health apps), 6) did not focus on natural human dialogue as a primary communication mode for two-way interaction (e.g., embodied CAs, facial recognition, etc.), 7) were purely theoretical analysis or a review of existing studies.

Data Screening

We first removed the duplicate entries from 224 articles. After removing 94 duplicates, we had 132 unique entries. Screening of articles for inclusion was performed in two stages. First, we screen the articles by reviewing titles, abstracts, and keywords. Next, we conducted relevancy coding by reviewing full texts based on the above criteria. The initial screening using titles and abstracts led to the removal of 46 articles. With the 86 remaining articles, we proceeded with the relevancy coding of the full texts. Through this relevancy coding process, 42 articles were removed and a set of 34 articles were processed for cross-reference. To identify additional relevant papers that were not identified in our initial search, we cross-referenced the citations of 34 articles. Through this cross-reference process, we identified 5 relevant articles to include in our analysis. After one more iteration of the cross-reference process, no additional relevant papers were identified, which suggests that we reached a saturation point. The final number of articles that were included in our literature review was 39. Figure 1 presents the data screening process following PRISMA.

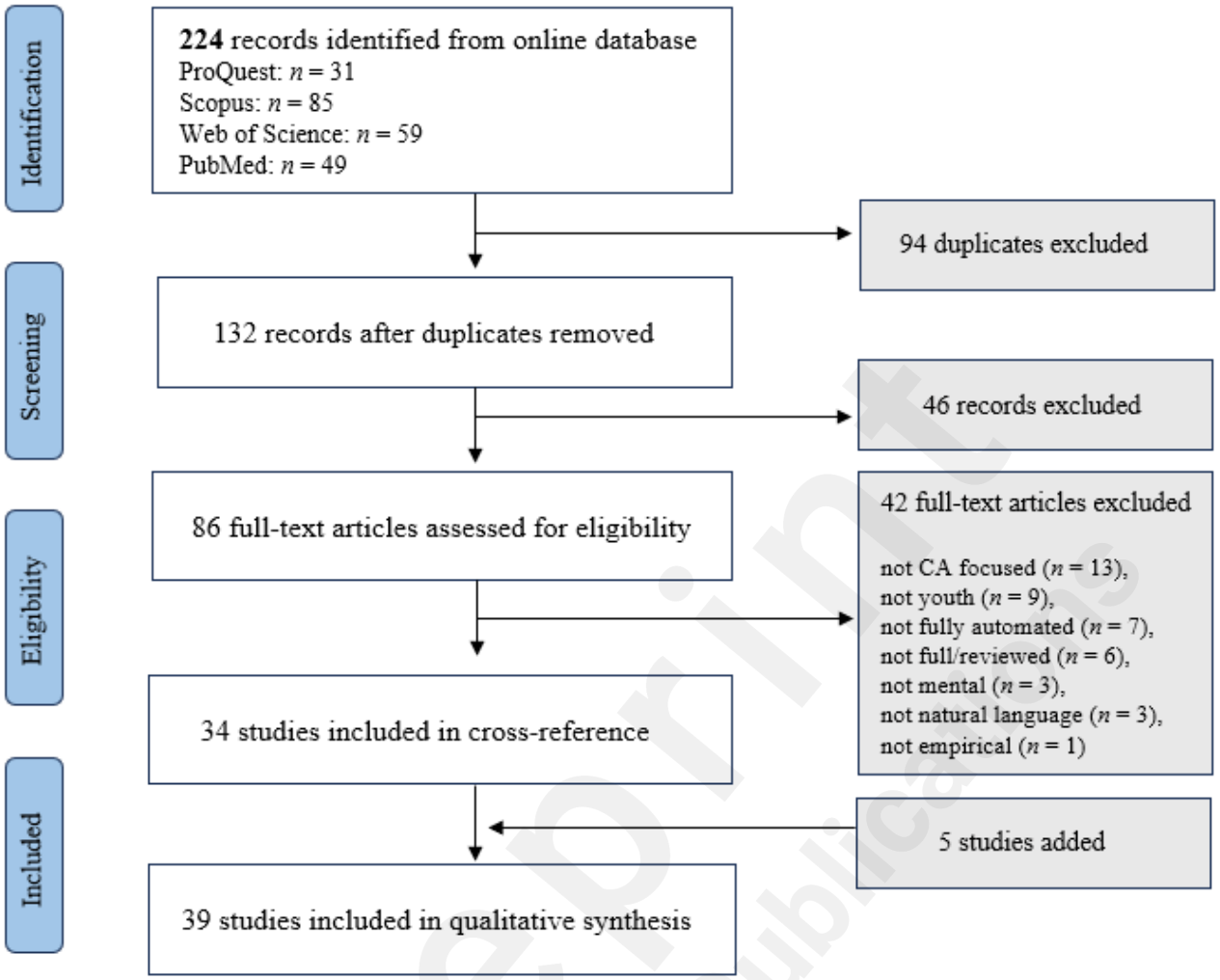

Figure 1: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the study

Data Analysis Approaches

We conducted a thematic analysis to identify major themes and trends in our dataset. We leveraged an iterative approach to our thematic analysis which involved refining the codes as we gained a deeper understanding of the data. The first author carried out the coding, supported by frequent check-ins with the last author for additional expertise on the subject matter. Furthermore, to ensure face validity, the second author quality-checked the coding during the writing of the results. For grounded thematic analysis, we analyzed the papers to identify codes for different dimensions aligning with our research questions. We familiarized ourselves with the literature identified and generated the initial codes. With the initial codes, we coded 20% of the dataset (n = 8) and reviewed the codes to ensure that they were representative of our dataset. Once we finalized the codes, we coded the entire dataset. Multiple codes were sometimes assigned to the same paper where necessary. Through the grounded thematic analysis, we identified major dimensions for the characteristics of empirical research on mental health CAs for youth (RQ1), design and computational considerations of mental health CAs for youth (RQ2), and the evaluation outcomes of empirical research on mental health CAs for youth (RQ3). Figure 2 is an overview of the framework including research questions, dimensions, and codes that were analyzed in this review to understand the trends in research on mental health CAs for youth.

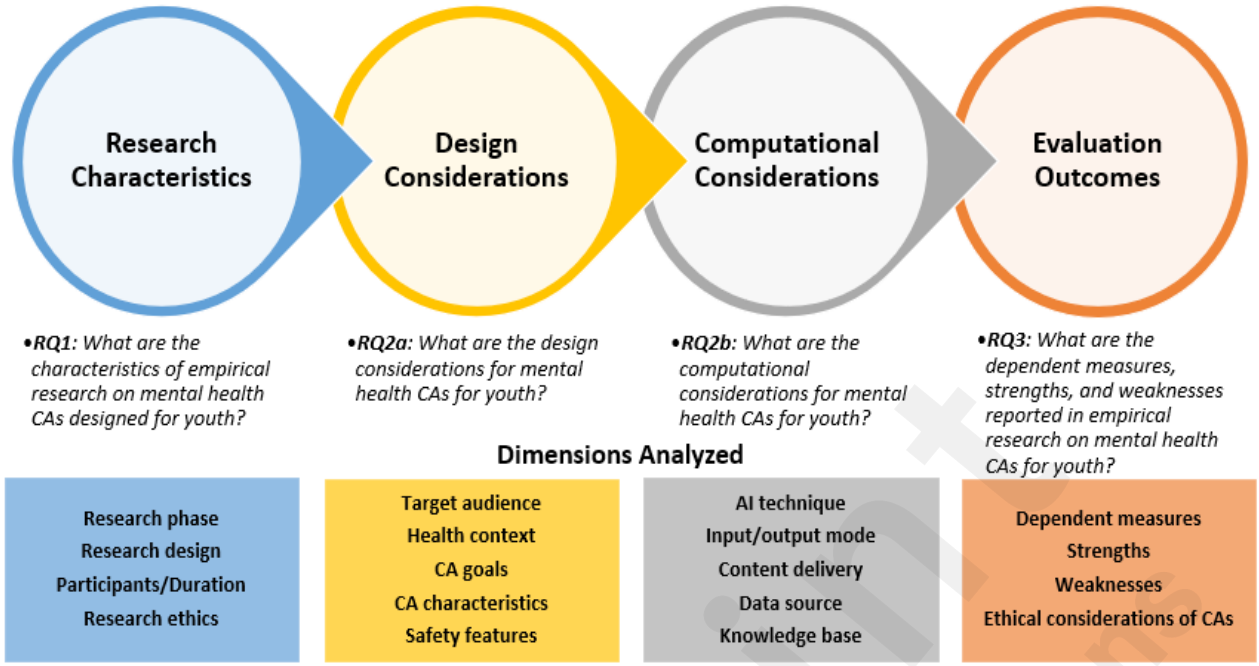

Figure 2: Overview of the analytical framework

Results

Overall, we included 39 studies in this scoping review [4, 18-55]. The reviewed studies were published between 2011-2024 (Feb.). The majority of the studies (72%, 28/39) were published between 2021 and 2024, with a few published before 2021 (28%, 11/39). Figure 3 shows the trend in the number of publications over time.

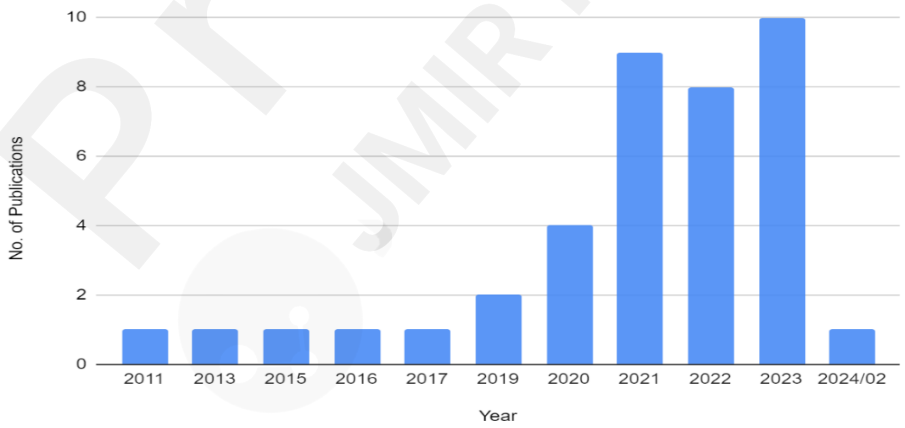

Figure 3: Number of publications over time (n = 39)

The reviewed studies were conducted in various regions including the US (21%, 8/39), New Zealand (13%, 5/39), Australia, (10%, 4/39), UK, China, and Brazil (8%, 3/39, respectively), Germany, Italy, Norway (5%, 2/39, respectively), Argentina, Bangladesh, Belgium, Canada, Netherlands, Philippines, Portugal (3%, 1/39, respectively).

Characteristics of empirical research on mental health conversational agents for youth

Most of the empirical research involved development and summative evaluation; less work involved formative evaluation to (re)design mental health conversational agents with and for youth

The majority of the reviewed studies included summative evaluation (92%, 36/39), followed by development (74%, 29/39), design (38%, 15/39), and formative evaluation (38%, 15/39) (Table 1). In most of the studies (78%, 36/46), multiple research phases were conducted, while 10 studies (22%) focused on a single research phase (i.e., summative evaluation of existing systems). All empirical research included an evaluative phase to assess the mental health CAs for youth. The majority of the studies included summative evaluation (92%, 36/39) and formative evaluation (38%, 15/39). In 12 studies (31%), both formative and summative evaluations were conducted. The term formative and summative evaluations originally referred to methodological frameworks for evaluating instructional materials [56]. Then the terms have been adopted in the Human-Computer Interaction literature. In HCI research, summative evaluation is the evaluation of a design after it is nearly complete, often used during field testing. Formative evaluation is conducted during user interface development, often iteratively to achieve the desired usability level. [57].

In the formative evaluation work that we reviewed, researchers interacted with youth populations to gain early feedback on the initial CA design through interviews, focus groups, and design workshops. The purpose of such formative evaluation work was to include voices of youth from the system design process to build youth-centered conversational systems. In those papers, researchers used participatory design approaches such as co-design [31], conceptual design [32], and human-centered design [49]. In most of the studies that included formative evaluation, summative evaluations were followed after the iterative design process. For instance, Ludin et al. conducted formative user testing with school guidance counselors, clinicians, and youth to learn about the issues youth faced during the COVID-19 pandemic and lockdowns. The youth's opinions from the formative evaluations were used to inform the ongoing content iteration of the prototypes. Once the prototype was developed, the authors conducted summative user testing to assess the usability and acceptability of the chatbot with 127 youth participants [33].

Table 1: Characteristics of empirical research on mental health CAs for youth (*n* = 39).

Dimensions/Codes	Percentage, Count	Key Trends
Research Phase		
Design	38%, 15	Most of the empirical research involved development and summative evaluation; less work involved formative evaluation to (re)design mental health CAs with youth
Development	74%, 29	
Formative Evaluation	38%, 15	
Summative Evaluation	92%, 36	
Research Design		
User testing	46%, 18	Most of the research involved user testing or experiments to evaluate mental health CAs
Experiment	46%, 18	
Research Participants/Duration		
Participants age		Most commonly, the research on mental health CAs was conducted with:
Adolescents (12-18)	41%, 16	

Young adults (18-30)	36%, 14	✓ less than 100 participants who are ✓ adolescents and young adults ✓ without mental health conditions ✓ for less than 4 weeks
Youth (14-25)	23%, 9	
Participants group		
General population	54%, 21	
At-risk population	46%, 18	
Number of participants		
<100	78%, 28	
100-200	15%, 6	
>200	5%, 2	
Duration of CA interaction		
One time	15%, 6	
<4weeks	31%, 12	
4-8 weeks	13%, 5	
9-16 weeks	3%, 1	
Research Ethics		
Consent	67%, 26	Research ethics such as institutional review and consent were addressed in the majority of research, yet only a few research addressed data privacy and safety of youth participants
Institutional Review	67%, 26	
Data Confidentiality	21%, 8	
Safety	10%, 4	

Research on mental health conversational agents was conducted with small numbers of adolescents and young adults without mental health conditions for less than 4 weeks

Participant group: The majority of the participants in the reviewed studies were adolescents (41%, 16/39) aged between 12-18, and young adults (36%, 14/39) over 18. In fewer studies, the participants were youth (23%, 9/39) aged between 14-25. Overall, in a slightly higher proportion of the empirical studies, the participants were general youth (54%, 21/39) compared to at-risk youth (46%, 18/39) such as those who are experiencing depression [51], anxiety [18], body image concerns [23], and alcohol abuse [20].

The number of participants: The number of youth participants ranged from a minimum of 3 young adults with autism to explore the acceptability of Cognitive Behavioral Therapy (CBT)-based mental health CA through in-depth interviews [53] to a maximum of 798 adolescents to evaluate the effectiveness and user engagement through survey [48]. In the majority of the research, the number of participants was less than 100 (78%, 28/39) with few studies having participants between 100-200 (15%, 6/39) and more than 200 (5%, 2/39).

Duration of CA interaction: Youth participants interacted with mental health CAs for a minimum of 30 mins [20, 25] to a maximum of 16 weeks [47]. In most of the reviewed studies, participants interacted with CAs for less than 4 weeks (31%, 12/39) or once (15%, 6/39). In fewer studies, participants interacted with CAs for 4-8 weeks (13%, 5/39) or 9-16 weeks (3%, 1/39). In general, short-term engagement with CAs was for user testing to assess user experience and acceptability. All of the long-term engagement with CAs was for the experiments to assess the effectiveness of CAs in reducing mental health conditions.

Most of the research involved user testing or experiments to evaluate mental health conversational agents

User testing: The most prevalent research design used in the empirical studies was user testing (46%, 18/39) to assess user engagement and user experience of the CAs. For instance, in a study by Beilharz et al., 17 adolescents and 8 parents/caregivers in Australia participated in a user study to evaluate the "KIT" prototype, a CA designed to support people with concerns about body image and eating issues. After interacting with KIT, participants took part in focus group interviews and shared their experiences with KIT such as acceptability, ease of use, and CA design [23]. Sometimes, user testing lasts for a longer period. For instance, Gabrielli et al. conducted a user test with 21 adolescents over 4 weeks in Italy to assess the user experience and perceived value of content delivered by mental health CAs. Surveys with Likert scale measures and open-ended questions were used to evaluate overall usefulness, ease of use, value of the content, and suggestions for improvement [37].

Experiment: Another prevalent research design was an experiment (46%, 18/39), all of which was to explore the effectiveness of CAs in reducing mental health conditions. Broadly, there were two types of experimental design: randomized controlled design and non-controlled design. With the randomized controlled design, participants are randomly assigned to one of the experimental groups and those in different groups receive different types of intervention for comparison (e.g., chatbot vs online search). For instance, young adults who completed active cancer treatment in 5 years were randomly assigned to either immediate access to mental health CA (experimental group) or access to only daily emotion ratings and access to full chatbot content after 4 weeks (control). After 4 weeks, participants in the experimental group reported an average reduction in anxiety, while the control group reported an increase in anxiety [43]. With non-controlled experiments, all participants receive the same intervention, and the comparison takes place before and after the intervention. For instance, 105 young adults interacted with the same mental health CAs for 15 days. The participants completed the surveys asking about their mental health conditions (i.e., Patient Health Questionnaire, PHQ) before and after the 15-day intervention period. The comparison of the average survey scores before and after the intervention confirmed that the overall scores decreased after interacting with the mental health chatbot [36].

Research ethics such as institutional review and consent were addressed in the majority of research, yet only a few research considered data privacy and safety of youth participants

In the majority of the reviewed studies, authors explicitly stated that they acquired participant consent (67%, 26/39) and that the studies were approved by their institutional review board (67%, 26/39). Beyond institutional review and participant consent which are mandatory in many institutions, in some studies, authors addressed considerations for privacy and/or confidentiality of youth' digital trace data related to mental health (21%, 8/39). In only a small portion of the reviewed studies (10%, 4/39), the authors provided support to promote the safety of youth participants in empirical research in the mental health context. For instance, in Nicol et al., there was safety monitoring provided by the study team by tracking the digital PHQ-9 assessment over 12 weeks. When the PHQ-9 assessment results indicated suicidal ideation, an email alert was sent to the study team, and the principal investigator contacted the participant's primary care providers to share information and discuss the next steps regarding the assessment of suicide risk. By the end of the 12-week study, 10 out of 18 participants (56%) triggered at least 1 alarm to assess for suicidal ideation [22]. In another study by Liu et al., participants were informed that professionals would intervene by telephone (i.e., Mental Assistance Hotline) when the participants reported that they needed

emergency psychological assistance. Participants were informed that the professional would remain in contact with them until they were ensured to be safe [47]. In 10 papers, research ethics were not mentioned. None of the reviewed papers mentioned risk mitigation plans and mandated reporting or provided critical reflections on best practices for working with youth in the mental health context.

Design considerations of mental health conversational agents for youth

Most of the mental health conversational agents were designed for older and general youth populations

Target Audience: In the majority of the studies, the target audience of the CAs were youth/young adults (56%, 22/39), adolescents (31%, 12/39), and college students (15%, 6/39). In most of the studies, CAs were designed for general youth populations (69%, 27/39). In a small proportion of the reviewed studies (31%, 12/39), the CAs were designed for at-risk youth populations (Table 2).

Table 2: Design considerations of mental health CAs for youth (*n* = 39).

Dimensions/Codes	Percentage, Count	Key Trends
Target Audience		
Age Group		Most of the mental health conversational agents were designed for older and general youth populations
Youth/Young adults	56%, 22	
Adolescent	31%, 12	
College Students	15%, 6	
Health Condition		
General youth	69%, 27	
At-risk youth	31%, 12	
CA Goals/Health Context		
Health Context		Most mental health conversational agents for youth are designed to provide: ✓ educational/informational support to promote the general mental well-being ✓ therapeutic support to alleviate mental health conditions such as depression, anxiety, and stress
Mental well-being	38%, 15	
Depression	31%, 12	
Anxiety	23%, 9	
Stress	21%, 8	
Substance Use	7%, 3	
Body Image	5%, 2	
Phone Addiction	3%, 1	
CA Goals		
Treatment	64%, 25	
Education/Training	56%, 22	
Informational	38%, 15	
Assessment	26%, 10	
Monitoring	5%, 2	
Behavioral Change	5%, 2	
CA Role/Characteristic		
CA Role		Most mental health conversational agents for youth were designed to be: ✓ life coaches or older peer mentors
Coach/Peer	28%, 11	
Healthcare	18%, 7	

Professional		✓ in friendly and empathetic tones ✓ with few options to personalize agent characters
CA Characteristics		
Friendly	41%, 16	
Empathetic	28%, 11	
Culture-specific	10%, 4	
Gender-specific	8%, 3	
Simple and Factual	5%, 2	
Personalization		
CA character	13%, 5	
App Appearance	3%, 1	
User Avatar	3%, 1	
Safety Features		
Reminder	31%, 12	The majority of the mental health conversational agents did not have the safety features in place
Emergency Contact	26%, 10	
Alert to adults/experts	5%, 2	

Within those studies, 12 studies focused on mental health CAs, out of which 5 studies focused on designing mental health CAs for young adults with depressive symptoms [18, 19, 29, 51, 52]. Few studies focused on CAs to promote the mental health of adolescents with Type 1 Diabetes [26], adolescents with body image and eating issues [23], young adults being treated for cancer [43], youth at risk of HIV and STIs [30], young adults from immigrant and refugee communities [49], LGBTQ+ youth [50], and young adults with autism [53]. There was a trend in which the number of mental health CAs designed for at-risk youth increased, while the number of CAs for general youth decreased (Figure 4). Overall, the majority of the mental health CAs were designed for older and general youth populations, with the current trend toward designing CAs for youth who are at risk of mental health conditions.

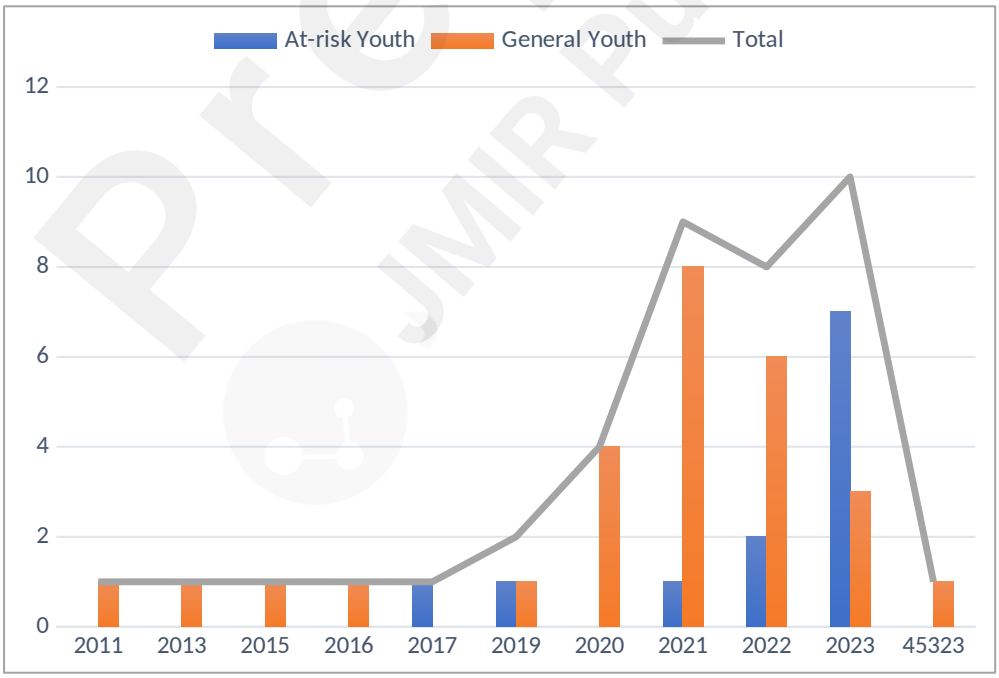

Figure 4: Target audience of mental health CAs by year (n = 39)

Mental health conversational agents are designed to promote the general mental well-being of youth as well as help alleviate mental health symptoms of depression, anxiety, and stress

Health context: Within the studies that focused on mental health CAs for youth, most of the CAs were designed to promote the general mental health/well-being of youth (38%, 15/39) (e.g., by promoting life skills and resilience), followed by those aimed to reduce depression (31%, 12/39), anxiety (23%, 9/39) and stress (21%, 8/39). There are a few papers that focused on mental health CAs to address the issues of substance use (i.e., alcohol and drug abuse) [20, 28, 42], body image [23, 48], and phone addiction [21]. Overall, a large proportion of the existing research on mental health CAs focused on supporting youth with symptoms of depression, anxiety, and stress.

CA goals: The majority of CAs for mental health was designed to help alleviate mental health symptoms (64%, 25/39) such as depression and anxiety, followed by educational/training content (56%, 22/39) to promote mental health and informational support on mental health topics (38%, 15/39). Some mental health CAs were designed to assess mental health symptoms (26%, 10/39) and/ or monitor mental health conditions (5%, 2/39). Two mental health CAs were designed to promote behavior changes such as reduced screen time [21] and reduced alcohol consumption [41]. Overall, in most of the work on mental health CAs, the CAs were designed to help alleviate youth's mental health symptoms or to provide educational intervention to promote youth's mental well-being, while less work has been done on CAs for informational support or the assessment/ monitoring of youth' mental health condition.

Mental health conversational agents were designed to be life coaches or older peer mentors with friendly and empathetic tones, with few options to personalize agent characters

CA role: In 20 studies (51%), the CA role was mentioned. In many of those studies (28%, 11/39), the mental health CAs were designed as older peers and/or younger coaches to guide youth with therapeutic modules and provide youth with information to promote mental health. For instance, the mental health CA was designed to be a young person who messages the user once a day and guides them through a brief daily activity to help reduce stress. The interaction was designed to be a brief conversation with a friend who checks in and has a helpful tip or an anecdote/story to share [38]. In other studies, mental health CAs were designed to provide formal health professional-like support (18%, 7/39). For instance, in a study by Wrightson-Hester et al., the mental health CA emulates therapists who encourage patients to explore their problems by asking questions to provide personalized mental health support [52].

CA characteristics: Meanwhile, the majority of mental health CAs were designed to provide friendly (41%, 16/39) and/or empathetic (28%, 11/39) responses. For instance, in response to loneliness reflected in users' input, the chatbot replied "*I'm so sorry you're feeling lonely. I guess we all feel a little lonely sometimes*" or if participants showed excitement in their input, the chatbot replied, "*Yay, always good to hear that!*" [18]. In four studies (10%), mental health CAs were designed to be culture-specific. For instance, in a study by Ludin et al., researchers collaborated with two groups of indigenous people (Māori and Pākehā) in New Zealand to co-create mental health CA called "Aroha," meaning "caring and kind" in their native language. The mental health strategies and activities were designed for and targeted Māori youth and their family [33]. A few mental health CAs were gender-specific (8%, 3/39) or designed to provide short and simple answers (5%, 2/39). In about one-third of the reviewed studies (28%, 11/39), the role or personality of mental health CAs

was not specified.

CA characters/avatars: In the majority of the studies, options to personalize mental health CA characters were not addressed (85%, 33/39). In a few studies (15%, 6/39), features to support personalization of CA characters were discussed. For instance, for the CA to promote youth resilience, the avatars were designed as older peers with users' choice of gender and ethnicity representative of the target population [32]. In another study, the authors designed the body image CA with an option to choose between male and female versions of the CA avatar [48]. Some studies designed mental health CAs in which users can customize their avatars [52] and app appearance [53]. In one study, personalization of the mental health CA characters was the major focus of summative evaluation in which youth preferred to choose from 3-4 characters with variations of gender, avatar, age, and social role (health professional vs. younger coachlike) instead of personalizing each aspect separately [19].

The majority of the mental health conversational agents did not have the safety features in place

The majority of the studies did not address the safety features of the designed mental health CAs (54%, 21/39). In 18 studies (46%), safety features were discussed including a reminder that users are interacting with chatbots, not human experts, and chatbots are not replacements for healthcare providers or places for seeking help (31%, 12/39). There were safety features to provide emergency contact (26%, 10/39) such as crisis hotlines and those that refer to health professionals. In two studies, alert features were implemented to notify a trusting adult or providers when a youth was identified as a risk to themselves or others [22, 31]. For instance, if trigger words such as self-harm, suicide, death, dead, kill, die were used by youth, the alert feature to notify adults is activated. The alert recipient is a Primary Support Contact who was identified and confirmed during the initial consent process [31]. Overall, safety features of mental health CAs were discussed in less than half of the reviewed articles.

Computational considerations of mental health conversational agents for youth

Most of the mental health conversational agents were prototypes built upon existing mobile chat applications

Maturity of Device: In the majority of studies, mental health CAs were prototypes (67%, 26/39). In 7 studies (18%), CAs were developed as fully functioning systems, and in 8 studies (21%), researchers evaluated existing systems developed in other studies. In one study [24], CAs were explored as concepts without having actual systems (Table 3).

Delivery channel: Mental health CAs were delivered via diverse channels. The majority of CAs were delivered through mobile apps (41%, 16/39), followed by web applications (28%, 11/39), and desktop applications (15%, 6/39). In 11 studies (28%), CAs were made available on more than one channel. In many studies (41%, 16/39), CAs were delivered using existing chat applications such as Facebook Messenger (e.g., [32, 33, 36]), WhatsApp [51], Windows Live Messenger [42], WeChat [47], and Telegram [27]. In some papers, the delivery channel of CAs was not specified (21%, 8/39).

Table 3: Computational considerations of mental health CAs for youth (n = 39).

Dimensions/Codes	Percentage, Count	Key Trends
System Characteristics		
Maturity of Device		Most of the mental health conversational agents were prototypes built upon existing mobile chat applications
Prototype	67%, 26	
Fully Functioning System	18%, 7	
Existing System	21%, 8	
Concept	3%, 1	
Delivery Channel		
Mobile	41%, 16	
Web	28%, 11	
Desktop	15%, 6	
Communication Mode		
Input Mode		All conversational agents supported text-based input (free text along with quick options) with a few that supported voice input; all supported textual output with many of them supporting visualized output such as image and video
Text	100%, 39	
Text + Speech	5%, 2	
Free text + options	49%, 19	
Free text	33%, 13	
Quick options	13%, 5	
Output Mode		
Text	100%, 39	
Image	54%, 21	
Video	31%, 12	
Audio	18%, 7	
Game	15%, 6	
AI Technique		
NLP	49%, 19	Free textual inputs were processed via NLP, while pre-written content was delivered via rule-based programming
Rule-based	38%, 15	
ML/Deep learning	8%, 3	
Content Delivery		
Flexible	64%, 25	Most mental health content was delivered in flexible ways; in some cases, pre-written content was delivered in a structured manner
Structured	21%, 8	
Semi-structured	8%, 3	
Knowledge base		
CBT	49%, 19	Mental health content was built upon evidence-based expert knowledge of cognitive and behavioral therapy and
Positive psychology	18%, 7	
Other therapeutic	18%, 7	

	content		positive psychology
	Clinical knowledge	expert 15%, 6	

All conversational agents supported text-based input (free text along with quick options) with a few that supported voice input; most conversational agents supported multimedia output

Input modality: In all of the reviewed studies, mental health CAs supported text as the primary mode of input (100%, 39/39). In two studies (5%), mental health CAs supported audio input along with textual input [47, 51]. For text-based CAs, by extracting keywords related to mental health such as depression and anxiety from user inputs, CAs extract necessary features to develop different classification models or assess its sentiment [46, 52]. For voice-based CAs, the key components include the use of automatic speech recognition (ASR) and Natural Language Understanding to comprehend users' input [47]. Text-based and voice-based chatbots are similar in their design, yet text-based CAs do not include ASR components and instead, rely on text as the primary modality of input and output. In the majority of the reviewed studies, CAs supported free text along with quick options (e.g., "yes" or "no") as user input (49%, 19/39), rather than free text only (33%, 13/39) or quick options only (13%, 5/39). There was a trend in which the proportion of CAs that supported free text along with quick options as user input increased, while the proportion of CAs that supported free text only as use input decreased over time (Figure 5).

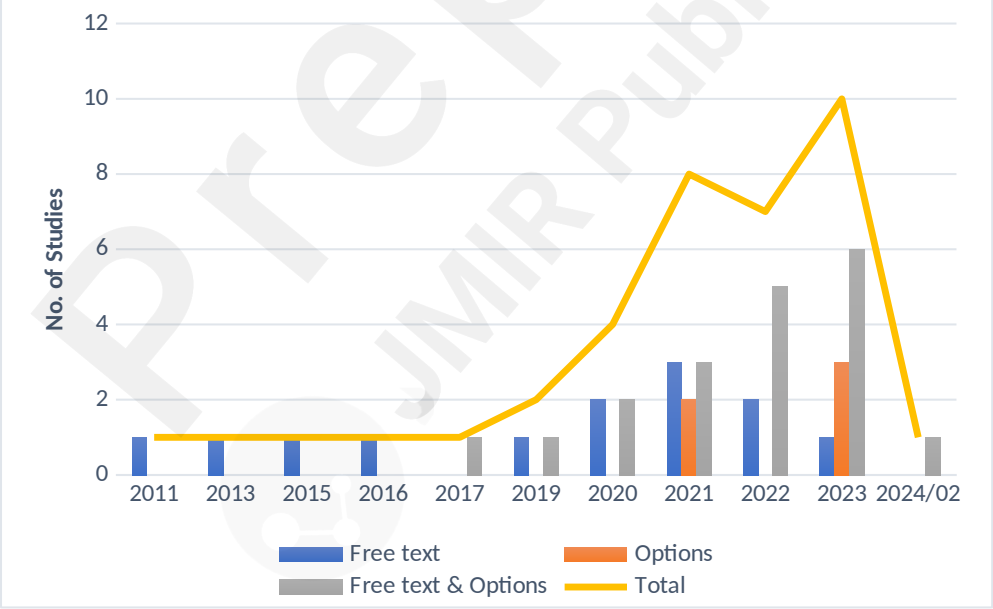

Figure 5: Input mode of mental health CAs by year (n = 39)

Output Modality: All of the CAs supported text as the primary mode of output (100%, 39/39). In 25 studies (64%), it was indicated that CAs support multi-media outputs. In addition to textual output, the majority of CAs supported image output (54%, 21/39) in the form of an emoji, moving images in Graphics Interchange Format, infographics, and animated CA avatars. CAs also provided outputs in the forms of video (31%, 12/39), audio (18%, 7/39), and game (15%, 6/39). In many of the reviewed studies (31%, 12/39), CAs supported more than two types of output (e.g., text, image, video).

Many conversational agents employed multiple AI techniques: free textual inputs were processed via NLP, while pre-written content was delivered via rule-based programming

In many studies that we reviewed, the CAs employed multiple AI-based computational methods depending on the specific section or feature. Free-textual input from users is processed using Natural Language Processing (NLP) which processes natural language datasets, such as text and voice, using statistical and machine learning models to recognize, understand, and generate text and speech [58]. Accordingly, the AI approaches that were most frequently referred to in the reviewed papers was NLP (49%, 19/39). Meanwhile, quick options are implemented using rule-based methods. Therefore, following the NLP, the second most frequently referenced technique in the reviewed papers was rule-based (38%, 15/39) approaches. Mostly, the CAs deliver pre-written conversational lessons via rule-based programming. Yet, they use techniques such as NLP at certain points in the tree to determine routing to subsequent conversational nodes (e.g., [18, 26, 38, 44, 54, 55]). For instance, NLP is used to process users' input so that CAs can trigger pre-written therapeutic content [47] or further crises and mental health support services [26] to youth. There were a few papers in which CAs relied solely on decisions made by decision trees or rule-based programs (e.g., [23, 25, 50, 51]). With rule-based CAs, once users choose among the given options (e.g., yes or no, choosing numbered options) the CAs trigger and provide the most relevant pre-written response. With the CAs that are built solely upon rule-based and/or decision tree techniques, there is no space for users to type free text to have conversations.

Most mental health content was delivered in flexible ways, considering user preferences and autonomy. In some cases, pre-written therapeutic content was delivered in a structured manner to help alleviate mental health symptoms for youth with mental health conditions.

We found three types of CA content delivery approaches in the reviewed studies: structured, semi-structured, and flexible. In the majority of the studies, mental health CAs implemented a flexible model (64%, 25/39). With a flexible model, CA contents are delivered depending on users' input or users have options to choose which content to navigate. For instance, Huang et al. designed a mental health chatbot that first detects stress levels from users' input, then based on the stress detection result as well as the users' chatting sentence type, the CA chooses an appropriate answer from the knowledge base database [44]. In another study by Palma et al., users have options to choose what kind of CBT-based therapy they would like to perform [53].

In about 21% of the reviewed studies (8/39), CA content was delivered in a planned and structured way. With the structured models, CA contents are pre-written and automatically delivered to users via rule-based programming for a pre-defined period (i.e., from a few days to a few weeks). For instance, Williams et al. designed a mental health CA in which daily conversations and modules are pre-written for 21 days, and the CA guides them through daily activity [38]. Particularly, in half of the papers, the structured CA content was designed to treat mental health conditions of at-risk youth including adolescent/young adults with depressive symptoms [29, 51], youth treated for cancer [43], and adolescents with Type 1 Diabetes [26].

In a small portion of the reviewed studies (8%, 3/39), CA content was delivered in a semi-structured way [32, 48, 49]. For instance, the mental health CA guided the user through 10 daily sessions that introduced the psychological content modules designed to be completed in 10-15 minutes. The same core psychological concepts (e.g., psychoeducation, emotional regulation, problem-solving) are

present for all users while users can specify their preferences for the content that they feel is most relevant to them [49]. In another study, the mental health content spanning 21 days was structurally designed for users to engage with the CA once a day. After going through the structured content, users end their daily interaction with CA with a choice of a motivational quote, joke, or an entry into a gratitude journal [32].

Mental health content was built upon evidence-based expert knowledge of mental health

In all studies, primary data sources were expert knowledge (100%, 39/39). In almost half of the reviewed studies on mental health CAs, the CA content was developed based on Cognitive Behavioral Therapy (CBT) (49%, 19/39), one of the most extensively studied evidence-based psychotherapy for a wide range of mental health issues such as depression, stress, and general mental well-being [59]. For instance, building upon the principles of CBT, He et al. designed 7 modules of mental health CA including cognitive distortions, self-esteem, mindfulness meditation, mental energy, natural connection, self-help, and loneliness. In their study, youth participants were asked to complete one module per day in sequence during the 1-week intervention period [29]. Besides CBT, mental health CA content was based on positive psychology (18%, 7/39) and other therapeutic content (18%, 7/39) such as Interpersonal Psychotherapy [19, 22], Acceptance Commitment Therapy [23], Perceptual Control Therapy [46, 52], self-compassion program [26], and micro intervention to improve body image [48]. Some mental health CA content was built upon inputs from school health service experts or counselors (15%, 6/39) and positive answers collected from the online forum [44]. In four studies, the source of mental health CA content was not specified (10%, 4/39).

In summary, we found that most mental health CAs were designed as older peers to provide therapeutic and/or educational content to promote youth mental well-being. Most of the CAs were built upon multiple computational methods to flexibly deliver pre-written and evidence-based mental health content. Text was the primary communication mode, with the majority supporting multimedia output. Table 4 presents a summary of the key trends in the design and computational aspects of mental health CAs for youth.

Table 4: Key trends of design/computational aspects of mental health CAs for youth

Aspects	Key Trends
Design	• **Target Audience**: designed for older and general youth populations • **Health Context/Goals**: designed to help alleviate mental health symptoms or to promote general mental well-being

	• **CA Role/Characteristics**: designed to be life coaches/older peer mentors with friendly and empathetic tones • **Personalization**: personalization of CA characters was not supported in most of the studies • **Safety Features**: Safety features were discussed in less than half of the studies
Computational	• **AI Technique**: multiple computational methods (e.g., NLP, rule-based) depending on the specific features • **Input Modality**: all CAs were text-based with a couple of them supporting audio input along with textual input. Most CAs supported free text along with quick options as user input • **Output Modality**: all CAs supported text as the primary mode of output with the majority supporting multi-media outputs • **Content Delivery**: CA content was delivered in flexible ways, considering user preferences and autonomy. In some cases, pre-written therapeutic content was delivered in a structured manner • **Data Source**: CA content was based on well-established evidence-based expert knowledge

Evaluation outcomes reported in research on mental health conversational agents for youth

Most studies focused on evaluating the effectiveness and acceptability of conversational agents

Effectiveness: In the majority of the evaluative research, the outcome variable was the effectiveness of mental health CAs (44%, 17/39) (Table 5). In those studies, pre-validated scales such as the Patient Health Questionnaire (PHQ) and Generalized Anxiety Disorder Scale (GAD) were used to measure the mental health conditions of the youth before and after the use of mental health CAs. For instance, to evaluate the effectiveness of mental health CA "Tess," Klos et al. conducted a randomized control trial with 181 Argentinian college students [45]. The PHQ-9 [60], a questionnaire comprising 9 items that evaluate the frequency and severity of depressive symptoms, and the GAD-7 [61], a 7-item scale that evaluates the frequency and severity of anxious thoughts and behaviors were used to measure the changes in participants' mental health [45].

Table 5: Evaluation outcomes of mental health CAs for youth (n = 39).

Dimensions/Codes	Percentage, Count	Key Trends
Dependent Measures		
Effectiveness	44%, 17	Most studies focused on evaluating the effectiveness and acceptability of conversational agents
Acceptability	44%, 17	
Usability	36%, 14	

User Engagement	26%, 10	
Validity/Accuracy	5%, 2	
Personalization	3%, 1	
Strengths		
Accessibility	36%, 14	Easy access to useful mental health information communicated through friendly, empathetic, and human-like responses were the major strengths of mental health conversational agents
Useful content	36%, 14	
Interactivity/Engaging	31%, 12	
Anonymity/ Confidentiality	18%, 7	
Empathetic/Friendly Responses	15%, 6	
Easy to navigate	10%, 4	
Human-like interaction	5%, 2	
Non-judgmental	5%, 2	
Weaknesses		
Limited responses/content	33%, 13	Limited/repetitive content, lack of human language understanding, and robotic responses were the critical weaknesses of mental health conversational agents
Lack of natural language understanding	21%, 8	
Lack of personalized content	18%, 7	
Non-human like traits	18%, 7	
Textual information with jargon	13%, 5	
Inaccurate responses	13%, 5	
Data privacy and confidentiality	13%, 5	
Ethical Considerations		
Privacy and confidentiality	10%, 4	Ethical considerations such as privacy, confidentiality, and safety of CAs were addressed in a few studies
Safety	3%, 1	

Acceptability: Another most evaluated outcome of mental health CAs was acceptability (44%, 17/39). Overall, current mental health CAs were perceived as beneficial and acceptable by not only youth but also health professionals [4]. For instance, semi-structured interviews with three young adults with autism showed that they considered the mental health chatbot novel and they would be interested in using such systems [53]. In a study by Mariamo et al., question type predicted the likelihood of response, such that yes/no questions were associated with a lower likelihood of response compared to multiple-response choice questions and a higher likelihood of response compared to open-ended questions [24].

Usability. Next, the usability of mental health CAs was evaluated in 14 studies (36%). For instance, Williams et al. conducted an experiment with 124 young adults in New Zealand to evaluate mental health CAs with 21-day stress detox modules. The results showed that the participants appreciated the interactivity, accessibility, and chatbot design, particularly due to visualized content as it is engaging and easier to relate to [38]. Similarly, a week-long exploratory study with 20 rural-living LGBTQ+ youth confirmed that youth appreciated the colorful design with multi-media content of "REALbot" as well as the conversational flow to first ask users' preferred names [50].

User Engagement: A more objective evaluation of CAs was also conducted such as user engagement

(26%, 10/39) through analyzing log data of human-CA interaction. For instance, Matheson et al. evaluated user engagement with the body image chatbot "Topity" which was designed to deliver micro-interventions for adolescents. The results from randomized control trials with 1,715 Brazilian adolescents show that 79% of the participants completed the minimum intervention dose of one micro-intervention technique. In addition, most participants chose to receive guidance from a woman avatar of Topity, compared to a male avatar [48].

Accuracy/Personalization: In a few studies, the accuracy/validity of mental health assessment (5%, 2/39) as well as personalization (3%, 1/39) were also explored. Huang et al. conducted a user study to evaluate the accuracy of a mental health chatbot designed for sensing and treating adolescents' stress. A month-long user testing showed that the chatbot's stress detection module achieved a precision rate of 78.34% and a recall rate of 76.12% [44]. In terms of personalization of mental health CA, while both experts and youth emphasized the need for autonomy to flexibly choose or change a module instead of a fixed schedule, experts emphasized the importance of planned therapeutic modules in a fixed sequence [19].

Easy access to useful mental health information communicated through friendly, empathetic, and human-like responses were the major strengths of mental health conversational agents

In 67% of the reviewed studies (26/39), the strengths of current mental health CAs for youth were addressed. The most frequently mentioned strengths of mental health CAs were 24/7 availability/accessibility (36%, 14/39), followed by useful information and therapeutic content (36%, 14/39), and the ability to have interactive and engaging conversations with CAs (31%, 12/39). In some studies, participants pointed to the strengths of mental health CAs such as perceived anonymity/confidentiality (18%, 7/39), empathetic and friendly responses generated by CAs (15%, 6/39), and easy-to-navigate (10%, 4/39). In fewer studies, participants liked non-judgmental responses generated by CAs [23, 43] delivered in multimedia output [27, 54]. They also liked the sense of caring and support [34, 40, 41] provided by CAs. Some participants appreciated human-like interaction with CAs [27, 49], perceived CAs as smart and trustworthy friends [37], and sometimes formed connections/friendships with bots [38]. They also liked the availability of personalized support [4, 29] delivered by personalized CA avatars [19].

Limited/repetitive content, lack of human language understanding, and robotic responses were the critical weaknesses of mental health conversational agents

In about 62% of the reviewed studies (24/39), the weaknesses of the current mental health CAs were addressed. The most frequently addressed limitations in the reviewed studies are limited/repetitive content provided by CAs (33%, 13/39), lack of understanding of human input (21%, 8/39), and lack of personalized content (18%, 7/39). Some of the limitations of mental health CAs were related to non-human-like traits (18%, 7/39) such as too-fast responses generated by CAs [30, 41], lack of empathy in their responses [37, 39], lack of trust toward non-human agents [35], a feeling of loneliness or disconnect when interacting with bot [38], and CAs being robotic [50, 54] and not smart enough [50]. Some youth found CA content with too much textual information with jargon or inaccurate/unclear responses generated by CAs hard to understand, while some shared concerns for confidentiality and privacy of sensitive information (13%, 5/39, respectively). In some studies, participants expressed difficulties with the free-text input mode as they found it hard to express their feelings [52] and utterances [20] rather than speak naturally. In a few studies, technical limitations

such as overall technical immaturity [29, 41] and susceptibility to changes in platform policies and bugs [51, 54] were addressed. Some of the potential safety concerns addressed in the reviewed studies were increased screen time [22], over-reliance on machines over human support [4, 22], risk of missing imminent risk [22, 51], and age-appropriateness of CA content [25]. Table 6 shows a summary of the strengths and weaknesses of mental health CAs reported in the reviewed studies.

Table 6: Mapping of strengths and weaknesses of mental health CAs for youth by dimension.

Dimensions	Strengths	Weaknesses
CA Characteristic	Empathetic and friendly responses	Too fast responses without empathy
CA Content	Useful information /therapeutic content	Inaccurate/unclear responses
AI Technique	Interactive and engaging conversations	Limited/repetitive content, lack of understanding of human input
Input Mode	Easy to start a conversation	Difficulties with terming the queries
Output Mode	Multimedia output	Too much textual information with jargon
Personalization	Availability of personalized support	Lack of personalized content
Privacy and Confidentiality	Perceived anonymity/confidentiality	Concerns for the privacy of sensitive information
Safety	Sense of caring and support via human-like interaction with CAs	Potential for over-reliance on CAs over human support and risk of missing imminent risk

Ethical considerations such as privacy, confidentiality, and safety of mental health conversational agents were addressed in a few studies

The majority of the reviewed studies (90%, 35/39) did not address the ethical aspects of mental health CAs. In four studies (10%), ethical considerations such as privacy, confidentiality, and safety of CAs were addressed [4, 22, 30, 35] as part of their empirical findings. For instance, safety was one of the outcome variables in the 12-week randomized control trial to assess the feasibility of the mental health CA designed for adolescents with depression and anxiety during the COVID-19 pandemic. Safety was assessed at 2, 4, 8, and 12 weeks by parents' reports on any hospitalizations or emergency department visits made by their child for depression/anxiety-related problems. By the end of the 12-week experiment, one parent from the intervention group reported that their teen was seen in an emergency department and discharged to home [22], indicating the potential safety concerns and the need for features to ensure the safety of youth. Meanwhile, in one study, ethical issues (e.g., privacy and confidentiality, efficacy, and safety of CAs) related to mental health CAs were discussed as the primary focus of the entire study. Through the group discussions with 14-18-year-old youth in the UK, the authors highlighted youth' concerns about mental health CAs related to their personal information. Their recommendations for designing ethical mental health CAs include 1) clear and transparent communication about the systems' privacy arrangement and limitations, 2) informing users of the extent to which the chatbots are evidence-based and empirically tested, and 3) the automated chatbots should have systems in place to prevent over-reliance and encourage users to seek human support [39]. Overall, more discussion on ethics standards and critical reflections on mental health CAs for the youth is needed.

Discussion

In this scoping review, we identified 39 studies that focused on CAs designed to support the mental health of youth. Below, we unpack the implications of our findings, followed by directions for future research. Finally, we provide recommendations for designing youth-centered, effective, and safe CA systems to support the mental health of youth.

Principal Findings

Mental health CAs that can support diverse youth with a variety of mental health issues are needed.

In the majority of the reviewed studies, the target audience of mental health CAs was general youth, with a recent trend toward designing mental health CAs for at-risk youth populations. According to 2021 statistics, growing numbers of youth are at risk of poor mental health outcomes. For instance, nearly half (45%) of lesbian, gay, bisexual, transgender, queer or questioning (LGBTQ) students seriously considered attempting suicide—far more than heterosexual students [62]; youth involved in the juvenile justice, or child welfare systems, as well as runaway youth and youth experiencing homelessness, are at higher risk of mental health challenges compared to other youth groups [5]. Accordingly, the US Surgeon General set an agenda to prioritize promoting the mental health of at-risk youth populations, such as racial, ethnic, sexual, and gender minority youth, individuals from lower socioeconomic backgrounds, youth with disabilities, youth involved in the juvenile justice system, and other groups [5]. Therefore, future work is needed to design and implement CAs to support the mental health of vulnerable youth. This will help improve understanding of disparities in mental health risks and designing proper interventions to promote mental health for youth. Additionally, we found that most mental health CAs were designed to reduce symptoms of depression, anxiety, and stress, while only a few were designed to provide support for body image, phone addiction, substance use, and other mental health issues. A recent report shows that besides depression and anxiety, youth are increasingly experiencing diverse mental health issues including attention-deficit/hyperactivity disorder (ADHD), eating disorders, body image, suicide, and self-harm [5, 62, 63]. Therefore, mental health CAs that can support youth with a variety of mental health issues are needed.

Multimodal input and output is a key to designing mental health conversational agents that are inclusive of youth with diverse communication needs.

In terms of user input, all mental health CAs supported textual input, with very few supported voice-based output. In the majority of the reviewed studies, mental health CAs supported free text along with quick options as user input, with a few supported quick options only. Prior research showed that rule-based CAs with quick options are perceived as restricted in offering personalized advice, leading to low trust in the effectiveness of CAs in providing advice on sensitive topics [64]. Therefore, providing options to freely type queries could benefit youth to explore diverse mental health topics. At the same time, it would still be useful to have quick options to choose from or auto-fill features as some of the health topics are difficult to term the queries from scratch. In addition, an option for voice-based input methods could be beneficial for supporting youth with diverse communication needs. When it comes to output mode, in almost two-thirds of studies, mental health CAs provided content in the form of images, audio, video, or games along with textual information. Multi-media content provided by mental and health CAs is important as evaluative research confirmed that some youth found lengthy texts hard to understand and prefer multi-media content. Hence, along with textual information, providing health information in a multi-media format is

needed for designing engaging conversational agents for youth.

More advanced AI technologies (e.g., LLMs) are needed to provide interactive and engaging mental health support for youth

Overall, mental health CAs for youth are in their infancy as many of the systems are developed as prototypes and are being evaluated for improvement. One of the major limitations found in evaluative research was limited content and/or responses provided by CAs as well as a *lack of personalized content.* This is because many of the mental health CAs were built upon a rule-based approach in which pre-defined sets of responses are based on domain-specific knowledge. Early evidence showed that rule-based CAs were seen as *only providing advice about mainstream*, easily accessible information, already available on the internet [64]. The evaluative studies we reviewed also demonstrated that youth perceived mental health content provided by rule-based CAs to be *repetitive* (e.g., [29, 31, 38, 49]. Yet, many of the existing research implemented rule-based approaches to provide mental health information for youth. Another major technical limitation found in the evaluative research was the lack of human language understanding, followed by inaccurate responses from CAs. Taken together, our findings signify the need for implementing more sophisticated language models in mental health CA development. A recent review study found that CAs enhanced by advanced AI technologies outperformed rule-based CAs in managing psychological distress, suggesting that human-like conversation may be key to determining the efficacy of interactive systems [11]. Further research is warranted to explore the potential benefits of implementing advance AI technology in mental health CAs for youth.

Safety should be prioritized when designing and implementing mental health conversational agents for youth

Recent advancements in large language models (LLMs) are promising in improving the technical immaturity of mental health CAs with the ability to understand input text written in human language in prompts and generate responses [65]. Early evidence demonstrated the effectiveness of the LLMs in generating coherent and relevant answers to psychological questions [66] or detecting mental health conditions [67]. Yet, when applying LLMs in mental health CAs for youth, the safety aspects of the information provided by those models should be rigorously considered given recent documentation of age-inappropriate and inaccurate content for youth generated by LLMs [68]. In addition, there is a risk that the conversation flow can go beyond the directions intended by the CA designer [65]. Therefore, chatbot development and implementations should undergo a robust validation process to establish a reliable and expert-informed evidence base for safety, particularly, with extra care when designing chatbots for vulnerable youth (e.g., those with mental health conditions). Meanwhile, safety concerns addressed in evaluative research included increased screen time, over-reliance on machines over human support, risk of missing imminent risk, and age-appropriateness. Although research findings are conflicting, the impact of over-reliance on machines and increased screen time on youth's well-being is one of the important safety concerns [69, 70]. Therefore, safety features to help track screen time and nudge youth about their CA use could be considered. Safety features to alert mental health professionals for imminent risk (e.g., nudges with sophisticated AI tech like deep learning) are also critical for mental health CAs.

Long-term, large-scale, and rigorous evaluation is needed to ensure the efficacy and safety of mental health conversational agents

Overall, most of the empirical research on mental health CAs was conducted with less than 100 older

youth populations in the short term (less than 4 weeks). This trend was significant as many of them involved user testing or clinical trials to assess the acceptability and feasibility of the prototypes. Although preliminary evidence shows positive trends in the effectiveness and acceptability of mental CAs, long-term evaluative research with larger sample sizes and more robust research designs is needed to validate their efficacy before their widespread adoption and use. In addition, we noted a lack of established methods for evaluating the safety of mental health CAs for unintended adverse effects. For instance, we identified that only a few studies reported evaluation outcomes related to the accuracy and validity of the mental health assessment modules in CAs. The trend was consistent with the mental health CAs for general adults [71]. However, evaluation of the accuracy of mental health content is critical to ensure the safety of youth interacting with the mental health CAs. While empirical research showed that age-appropriateness of mental health content was one of the weaknesses of mental health CAs, none of the reviewed studies evaluated whether the mental health content is developmentally appropriate for youth. Therefore, the accuracy and age appropriateness of mental health content provided by CAs should be further explored to ensure the efficacy and safety of mental health CAs for youth.

Collaborative efforts with youth and clinical experts are needed to design safe, effective, and youth-centered mental health conversational agents

The evaluation outcomes of the reviewed studies raised a few open questions to design safe, effective, and youth-centered mental health CAs. One is on the human-like traits of CAs; while youth prefer human-like traits of CAs, existing research documented safety concerns toward the humanness of CAs. For instance, prior research shows that younger youth may lose vital human contact [72] or unintentionally share personal information with CAs [73] if they become too attached to human-like CAs. Therefore, in some studies we reviewed, CAs were designed with non-human-like avatars (e.g., [23, 25, 30, 34]) or there were safety features to clearly state that the CAs are not human agents (e.g., [4, 27, 33, 35, 43, 47, 54]). As such, how to balance human-like and non-humna-like traits of mental health CAs for youth is an important open question to address in collaborative research with youth and clinical experts.

Another open question is on the CA role as evaluative research showed conflicting perceptions from youth and clinical experts; youth prefer peer/coach-like roles while experts were cautious about such roles for clinical purposes [19]. Therefore, careful consideration is needed when designing social roles of mental health CAs for youth taking into account specific health context, purpose, and target audience. Additionally, as evaluative research showed, providing an option to choose a preferred CA avatar can help youth feel comfortable when sharing confidential issues on sensitive topics [37]. Yet, we found that the current empirical research on mental health CAs rarely explored how personalizing the roles and characteristics of mental health CAs impact the effectiveness and/or user experience in both positive and negative ways. Therefore, further empirical research with youth and clinical experts is needed to understand how different permutations of CA roles and personalities play a role in supporting youth mental health.

In terms of personalized CA content delivery, we observed a trade-off between flexibility and structured planning in therapeutic content. Experts emphasized that planned modules could provide users with certainty and transparency, which makes mental health CAs more reliable and the treatment goals more visible; at the same time, they can reduce motivation, and user engagement, and thus lead to dropout [19]. Therefore, balancing between personalizing therapeutic content flexibly and maintaining a structured program is an open-ended question to address in future

research. Finally, in the majority of reviewed studies, the content of mental health CAs was based on well-established evidence-based expert knowledge. When considering vulnerable youth such as those with severe depressive symptoms and substance use disorder, evidence-based expert input is essential to safely leverage CA systems. However, to understand the mental health needs of youth, it is equally critical to work with youth from the early stages of CA design. Yet, very little work has been done to consider inputs from youth in content generation. Taken together, more collaborative research efforts with youth, caregivers, and domain experts need to be made to address the above open questions and build effective, safe, and youth-centered mental health CAs.

Ethics standards and best practices to design and develop mental health conversational agents for youth are needed.

We found a concerning trend in the existing literature on mental health CAs for youth: the majority of the papers did not address the ethical aspects of mental health CAs. Many of the existing review studies pointed to the lack of ethical considerations in the development of CAs in healthcare and called for further research to address ethical challenges in the process [2, 74, 75]. Personal health data is collected and processed in mental health CAs, yet the majority did not provide information regarding security and privacy aspects [75]. Our findings show that the trend is consistent for mental health CAs designed for youth. As mental health topics are sensitive, particularly for youth [6, 30], more discussion on ethics standards and critical reflections on best practices to develop mental health CAs for the youth population is needed. Another concerning trend we found was that the ethical implications and best practices of involving vulnerable youth in evaluative research on mental health CAs are rarely discussed. In a 12-week evaluative research with adolescents with depression and anxiety, more than half of the participants triggered at least 1 alarm for suicidal ideation, out of which two were referred to a child psychiatrist for ongoing clinical management [22]. This observation is a strong message that safety standards and best practices are pivotal when working with vulnerable populations. Hence, further research is needed to establish ethical standards for working with youth, particularly vulnerable youth, to ensure that participating in research does not harm already vulnerable populations. Below, we summarize the open questions that need to be addressed in future research (Textbox 1).

Textbox1: Open questions related to mental health CAs to be addressed in future research.

- **CA role**: How do different CA roles and characteristics impact youth's interaction with mental health CAs?
- **CA characteristics**: How can we balance human-like vs non-human-like traits of mental health CAs for youth?
- **CA content**: What are the age-appropriate, inclusive, and accurate mental health information for youth?
- **Content delivery**: What are the effective and safe delivery modes (structured vs flexible) of mental CA content for youth?
- **Safety**: How can we design CA features to promote the safety (e.g., monitoring imminent risk and screen time) of mental health CAs?
- **Data privacy and confidentiality**: How can we design AI-based systems that ensure privacy and confidentiality of youth data on sensitive topics?
- **Research ethics**: What are the ethics standards and best practices to design and develop mental health CAs for and with youth?

Design Guidelines

Based on our findings and broader implications, we provide the following guidelines for designing youth-centered mental health CAs (Textbox 2).

Textbox 2. Design guidelines for mental health CAs for youth

- **CA role and characteristics**: The CA role should be carefully designed considering the purpose and primary target audience. Regardless of roles, empathetic and friendly characteristics of mental health CAs are important.
- **CA content**: CA content should be age-appropriate, inclusive, and accurate. The content should be based on evidence-based expert knowledge along with inputs from youth.
- **AI technique**: More advanced language models (e.g., LLMs) are needed to provide diverse, context-aware, and personalized content for youth.
- **Input mode**: Free-textual inputs with auto-complete/quick options can help youth formulate questions that require domain knowledge. Along with textual input, an option for voice-based input can support youth with diverse communication needs.
- **Output mode**: Less textual content and more multimedia content can help youth understand mental health information.
- **Personalization**: It is important to give youth control over the personalization of CA content and avatars. Information on what is being personalized and how it is done should be clear and transparent.
- **Safety**: It should be clearly communicated to youth that CAs are not human, along with information on the capability and limitations of CAs. Information on confidentiality and data privacy should be clear and transparent. Safety features such as emergency contacts for imminent risk should be provided upfront and available 24/7. Additional safety features to track screen time could help reduce over-reliance on CAs and increased screen time.
- **Ethics**: Safety standards and critical reflections on best practices to co-design CAs with the youth population are needed.

Strength, Limitations, and Future Work

This scoping review has several strengths. First, we conducted a comprehensive literature search of multiple databases. We used holistic search terms rather than specific ones to capture the various representations of CAs used in mental health for youth. Second, we analyzed trends in empirical research on mental health CAs for youth as well as design and computational considerations of the mental health CAs studied in empirical research to provide a holistic mapping of the current landscape. This study, therefore, showcased the possible framework of mental health CAs that can be referenced by other researchers in this field.

Our study, however, has some limitations. First, given the novelty and multidisciplinary nature of the field, some unpublished literature presented at niche conferences and meetings may have been omitted. Second, during the data extraction process, we identified the design and computational approaches of the mental health CAs based on the descriptions reported in the reviewed studies. Hence, some of the design and computational aspects of mental health CAs that were considered yet not reported in the reviewed studies may have not been captured in this paper. As our findings suggest, the maturity of mental health CAs is still in its infancy and further review with more in-

depth analysis is needed as research in the field matures.

Conclusions

Conversational agents are increasingly used by youth for sensitive topics such as mental health topics. Trends in research on mental health CAs designed for youth have been under-explored. In this review paper, we fill an important gap by synthesizing 39 studies on mental health CAs designed for youth over the last 14 years. Our scoping review highlights that research on mental health CAs is in its infancy and early evidence shows both strengths and weaknesses in existing systems. Most commonly the research on mental health CAs was conducted with less than 100 participants who are adolescents and young adults without mental health conditions for less than 4 weeks. Further, less than a quarter of the studies considered data privacy and safety of youth in such research as tracked outcomes. We call attention to important open questions that researchers should address to move forward. When designing CAs for youth, a one-size-fits-all approach does not apply. With careful consideration of the health context and needs of specific target groups, mental health CAs can benefit youth. This can be only achieved when engaging with youth from the early design phases to summative evaluation of the systems. In this regard, we call for further investigation of best practices for risk mitigation strategies and ethical development of CAs with and for youth to promote their mental well-being.

Acknowledgments
Will be provided upon manuscript acceptance.

Conflicts of Interest
None declared.

Abbreviations

AI: artificial intelligence
ASR: automatic speech recognition
CA: conversational agent
CBT: cognitive behavioral therapy
LLM: large language model
NLP: natural language processing
PHQ: patient health questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

References

1. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: Trust-building factors and implications for conversational agent design. International Journal of Human–Computer Interaction. 2021;37(1):81-96.

2. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y-L, et al. Conversational agents in health care: scoping review and conceptual analysis. Journal of medical Internet research. 2020;22(8):e17158.
3. Balaji D, He L, Giani S, Bosse T, Wiers R, de Bruijn G-J. Effectiveness and acceptability of conversational agents for sexual health promotion: a systematic review and meta-analysis. Sexual health. 2022;19(5):391-405.
4. Koulouri T, Macredie RD, Olakitan D. Chatbots to support young adults' mental health: An exploratory study of acceptability. ACM Transactions on Interactive Intelligent Systems (TiiS). 2022;12(2):1-39.
5. Office of the Surgeon General. Protecting youth mental health: the US surgeon general's advisory. 2021. [cited 2024 4/11]; Available from: https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf
6. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. European child & adolescent psychiatry. 2021;30(2):183-211.
7. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. International journal of medical informatics. 2019;132:103978.
8. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR mental health. 2019;6(10):e14166.
9. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. Journal of medical Internet research. 2020;22(7):e16021.
10. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. Journal of Medical Internet Research. 2023;25:e43862.
11. Li H, Zhang R, Lee Y-C, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digital Medicine. 2023;6(1):236.
12. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. Journal of medical Internet research. 2021;23(1):e17828.
13. Ogilvie L, Prescott J, Carson J. The use of chatbots as supportive agents for people seeking help with substance use disorder: a systematic review. European Addiction Research. 2022;28(6):405-18.
14. Lim SM, Shiau CWC, Cheng LJ, Lau Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behavior Therapy. 2022;53(2):334-47.
15. Vaidyam AN, Linggonegoro D, Torous J. Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness: Changements du paysage psychiatrique des chatbots: une revue systématique des agents conversationnels dans la maladie mentale sérieuse. The Canadian Journal of Psychiatry. 2021;66(4):339-48.
16. Balan R, Dobrean A, Poetar CR. Use of automated conversational agents in improving young population mental health: a scoping review. NPJ Digital Medicine. 2024;7(1):75.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. International journal of surgery. 2010;8(5):336-

41.
18. Fitzpatrick KK, Darcy A, Vierhile M. 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 Ment Health. 2017 Jun 06;4(2):e19. PMID: 28588005. doi: 10.2196/mental.7785.
19. Kuhlmeier FO, Gnewuch U, Lüttke S, Brakemeier E-L, Mädche A. A Personalized Conversational Agent to Treat Depression in Youth and Young Adults – A Transdisciplinary Design Science Research Project. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2022.
20. Elmasri D, Maeder A, editors. A conversational agent for an online mental health intervention. Brain Informatics and Health: International Conference, BIH 2016, Omaha, NE, USA, October 13-16, 2016 Proceedings; 2016: Springer.
21. Abreu C, Campos PF, editors. Raising awareness of smartphone overuse among university students: a persuasive systems approach. Informatics; 2022: MDPI.
22. Nicol G, Wang R, Graham S, Dodd S, Garbutt J. Chatbot-Delivered Cognitive Behavioral Therapy in Adolescents With Depression and Anxiety During the COVID-19 Pandemic: Feasibility and Acceptability Study. JMIR Form Res. 2022 Nov 22;6(11):e40242. PMID: 36413390. doi: 10.2196/40242.
23. Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G. Development of a Positive Body Image Chatbot (KIT) With Young People and Parents/Carers: Qualitative Focus Group Study. J Med Internet Res. 2021 Jun 16;23(6):e27807. PMID: 34132644. doi: 10.2196/27807.
24. Mariamo A, Temcheff CE, Léger PM, Senecal S, Lau MA. Emotional Reactions and Likelihood of Response to Questions Designed for a Mental Health Chatbot Among Adolescents: Experimental Study. JMIR Hum Factors. 2021 Mar 18;8(1):e24343. PMID: 33734089. doi: 10.2196/24343.
25. Dosovitsky G, Bunge E. Development of a chatbot for depression: adolescent perceptions and recommendations. Child Adolesc Ment Health. 2023 Feb;28(1):124-7. PMID: 36507594. doi: 10.1111/camh.12627.
26. Boggiss A, Consedine N, Hopkins S, Silvester C, Jefferies C, Hofman P, et al. Improving the Well-being of Adolescents With Type 1 Diabetes During the COVID-19 Pandemic: Qualitative Study Exploring Acceptability and Clinical Usability of a Self-compassion Chatbot. JMIR Diabetes. 2023 May 05;8:e40641. PMID: 36939680. doi: 10.2196/40641.
27. Gabrielli S, Rizzi S, Bassi G, Carbone S, Maimone R, Marchesoni M, et al. Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study. JMIR Mhealth Uhealth. 2021 May 28;9(5):e27965. PMID: 33950849. doi: 10.2196/27965.
28. Schick A, Feine J, Morana S, Maedche A, Reininghaus U. Validity of Chatbot Use for Mental Health Assessment: Experimental Study. JMIR Mhealth Uhealth. 2022 Oct 31;10(10):e28082. PMID: 36315228. doi: 10.2196/28082.
29. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, et al. Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial. J Med Internet Res. 2022 Nov 21;24(11):e40719. PMID: 36355633. doi: 10.2196/40719.
30. Sanabria G, Greene KY, Tran JT, Gilyard S, DiGiovanni L, Emmanuel PJ, et al. "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. 2023 May 11:1-10. PMID: 37362063. doi: 10.1007/s41347-023-00315-4.
31. Grové C. Co-developing a Mental Health and Wellbeing Chatbot With and for Young People.

Front Psychiatry. 2020;11:606041. PMID: 33597898. doi: 10.3389/fpsyt.2020.606041.
32. Holt-Quick C, Warren J, Stasiak K, Williams R, Christie G, Hetrick S, et al. A Chatbot Architecture for Promoting Youth Resilience. Healthier Lives, Digitally Enabled: IOS Press; 2021. p. 99-105.
33. Ludin N, Holt-Quick C, Hopkins S, Stasiak K, Hetrick S, Warren J, et al. A Chatbot to Support Young People During the COVID-19 Pandemic in New Zealand: Evaluation of the Real-World Rollout of an Open Trial. J Med Internet Res. 2022 Nov 04;24(11):e38743. PMID: 36219754. doi: 10.2196/38743.
34. Høiland CG, Følstad A, Karahasanovic A. Hi, can I help? Exploring how to design a mental health chatbot for youths. Human Technology. 2020;16(2):139-69.
35. Brandtzæg PB, Skjuve M, Kristoffer Dysthe KK, Følstad A, editors. When the social becomes non-human: young people's perception of social support in chatbots. Proceedings of the 2021 CHI conference on human factors in computing systems; 2021.
36. Oliveira ALS, Matos LN, Junior MC, Delabrida ZNC, editors. An Initial Assessment of a Chatbot for Rumination-Focused Cognitive Behavioral Therapy (RFCBT) in College Students. Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part VI 21; 2021: Springer.
37. Gabrielli S, Rizzi S, Carbone S, Donisi V. A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study. JMIR Hum Factors. 2020 Feb 14;7(1):e16762. PMID: 32130128. doi: 10.2196/16762.
38. Williams R, Hopkins S, Frampton C, Holt-Quick C, Merry SN, Stasiak K. 21-day stress detox: open trial of a universal well-being chatbot for young adults. Social Sciences. 2021;10(11):416.
39. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, Group NYPsA. Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. Biomed Inform Insights. 2019;11:1178222619829083. PMID: 30858710. doi: 10.1177/1178222619829083.
40. De Nieva JO, Joaquin JA, Tan CB, Marc Te RK, Ong E, editors. Investigating students' use of a mental health chatbot to alleviate academic stress. 6th International ACM In-Cooperation HCI and UX Conference; 2020.
41. Maenhout L, Peuters C, Cardon G, Compernolle S, Crombez G, DeSmet A. Participatory Development and Pilot Testing of an Adolescent Health Promotion Chatbot. Front Public Health. 2021;9:724779. PMID: 34858919. doi: 10.3389/fpubh.2021.724779.
42. Crutzen R, Peters GJ, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health. 2011 May;48(5):514-9. PMID: 21501812. doi: 10.1016/j.jadohealth.2010.09.002.
43. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth. 2019 Oct 31;7(10):e15018. PMID: 31674920. doi: 10.2196/15018.
44. Huang J, Li Q, Xue Y, Cheng T, Xu S, Jia J, et al., editors. Teenchat: a chatterbot system for sensing and releasing adolescents' stress. Health Information Science: 4th International Conference, HIS 2015, Melbourne, Australia, May 28-30, 2015, Proceedings 4; 2015: Springer.
45. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. JMIR Form Res. 2021 Aug 12;5(8):e20678. PMID: 34092548. doi: 10.2196/20678.
46. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn

Psychother. 2014 Nov;42(6):731-46. PMID: 23899405. doi: 10.1017/S135246581300060X.
47. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. Internet Interv. 2022 Mar;27:100495. PMID: 35059305. doi: 10.1016/j.invent.2022.100495.
48. Matheson EL, Smith HG, Amaral ACS, Meireles JFF, Almeida MC, Linardon J, et al. Using Chatbot Technology to Improve Brazilian Adolescents' Body Image and Mental Health at Scale: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2023 Jun 19;11:e39934. PMID: 37335604. doi: 10.2196/39934.
49. Fabian KE, Foster KT, Chwastiak L, Turner M, Wagenaar BH. Adapting a transdiagnostic digital mental health intervention for use among immigrant and refugee youth in Seattle: a human-centered design approach. Transl Behav Med. 2023 Nov 05;13(11):867-75. PMID: 37418614. doi: 10.1093/tbm/ibad041.
50. Escobar-Viera CG, Porta G, Coulter RWS, Martina J, Goldbach J, Rollman BL. A chatbot-delivered intervention for optimizing social media use and reducing perceived isolation among rural-living LGBTQ+ youth: Development, acceptability, usability, satisfaction, and utility. Internet Interv. 2023 Dec;34:100668. PMID: 37746640. doi: 10.1016/j.invent.2023.100668.
51. Viduani A, Cosenza V, Fisher HL, Buchweitz C, Piccin J, Pereira R, et al. Assessing Mood With the Identifying Depression Early in Adolescence Chatbot (IDEABot): Development and Implementation Study. JMIR Hum Factors. 2023 Aug 07;10:e44388. PMID: 37548996. doi: 10.2196/44388.
52. Wrightson-Hester AR, Anderson G, Dunstan J, McEvoy PM, Sutton CJ, Myers B, et al. An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series. JMIR Hum Factors. 2023 Jul 21;10:e46849. PMID: 37477969. doi: 10.2196/46849.
53. Palma R, Lam HC, Shrivastava A, Karlinsey E, Nguyen K, Deol P, et al., editors. “Monday Feels Like Friday!”-Towards Overcoming Anxiety and Stress of Autistic Young Adults During Times of Isolation. International Conference on Information; 2023: Springer.
54. Kang A, Hetrick S, Cargo T, Hopkins S, Ludin N, Bodmer S, et al. Exploring Young Adults' Views About Aroha, a Chatbot for Stress Associated With the COVID-19 Pandemic: Interview Study Among Students. JMIR Form Res. 2023 Oct 12;7:e44556. PMID: 37527545. doi: 10.2196/44556.
55. Afrin Z, Farid DM, Mamun KAA, editors. A Cloud-Based Intelligent Virtual Assistant for Adolescents. International Conference on Intelligent Systems and Data Science; 2023: Springer.
56. Allal L. Assessment and the regulation of learning. International encyclopedia of education. 2010;3(1):348-52.
57. Hix D, Hartson HR. Formative Evaluation: Ensuring Usability in User Interfaces. Department of Computer Science, Virginia Polytechnic Institute & State ..., 1992.
58. Chowdhary KR. Natural language processing. Fundamentals of artificial intelligence. 2020:603-49.
59. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: A review of meta-analyses. Clinical psychology review. 2006;26(1):17-31.
60. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine. 2001;16(9):606-13.
61. Newman MG, Zuellig AR, Kachin KE, Constantino MJ, Przeworski A, Erickson T, et al. Preliminary reliability and validity of the Generalized Anxiety Disorder Questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. Behavior Therapy. 2002;33(2):215-33.
62. Center for Disease Control and Prevention. Mental Health: Poor Mental Health Impacts Adolescent Well-being. 2023 [cited 2024 4/11]; Available from: https://www.cdc.gov/healthyyouth/

mental-health/index.htm.
63. Office of Population Affairs. Mental Health for Adolescents. [cited 2024 4/11]; Available from: https://opa.hhs.gov/adolescent-health/mental-health-adolescents.
64. Nadarzynski T, Puentes V, Pawlak I, Mendes T, Montgomery I, Bayley J, et al. Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis. Sex Health. 2021 Nov;18(5):385-93. PMID: 34782055. doi: 10.1071/SH21123.
65. Jo E, Epstein DA, Jung H, Kim Y-H, editors. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; 2023.
66. Lai T, Shi Y, Du Z, Wu J, Fu K, Dou Y, et al. Supporting the Demand on Mental Health Services with AI-Based Conversational Large Language Models (LLMs). BioMedInformatics. 2023;4(1):8-33.
67. Xu X, Yao B, Dong Y, Gabriel S, Yu H, Hendler J, et al. Mental-llm: Leveraging large language models for mental health prediction via online text data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2024;8(1):1-32.
68. Kelly SM. Snapchat's new AI chatbot is already raising alarms among teens and parents. CNN; 2023 [cited 4/11 2024]; Available from: https://www.cnn.com/2023/04/27/tech/snapchat-my-ai-concerns-wellness/index.html.
69. Muppalla SK, Vuppalapati S, Pulliahgaru AR, Sreenivasulu H, kumar Muppalla S. Effects of excessive screen time on child development: an updated review and strategies for management. Cureus. 2023;15(6).
70. Joshi SV, Stubbe D, Li S-TT, Hilty DM. The use of technology by youth: Implications for psychiatric educators. Academic Psychiatry. 2019;43:101-9.
71. Jabir AI, Martinengo L, Lin X, Torous J, Subramaniam M, Tudor Car L. Evaluating conversational agents for mental health: scoping review of outcomes and outcome measurement instruments. Journal of Medical Internet Research. 2023;25:e44548.
72. Law T, Chita-Tegmark M, Rabb N, Scheutz M. Examining attachment to robots: benefits, challenges, and alternatives. ACM Transactions on Human-Robot Interaction (THRI). 2022;11(4):1-18.
73. Garg R, Cui H, Seligson S, Zhang B, Porcheron M, Clark L, et al., editors. The last decade of HCI research on children and voice-based conversational agents. Proceedings of the 2022 CHI conference on human factors in computing systems; 2022.
74. Omarov B, Narynov S, Zhumanov Z. Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review. Computers, Materials & Continua. 2023;74(3).
75. May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. Informatics for Health and Social Care. 2022;47(2):194-210.