

Supporting Youth Mental and Sexual Health Information Seeking in the Era of Artificial Intelligence (AI) based Conversational Agents: Current Landscape and Future Directions

JINKYUNG PARK, Vanderbilt University, USA

VIVEK SINGH, Rutgers University, USA

PAMELA WISNIEWSKI, Vanderbilt University, USA

Following the recent release of various Conversation Agents (CAs) (e.g., ChatGPT4), teens are increasingly using CAs for interactive knowledge discovery on sensitive topics, including mental and sexual health topics. Exploring such sensitive topics through online search has been an essential part of adolescent development, and CAs can support their knowledge discovery on such topics through human-like dialogues. Yet, never-before encountered risks have been documented with teens' interactions with AI-based CAs, such as being exposed to sexually inappropriate content and/or being given advice that is detrimental to their mental and physical wellbeing (e.g., to self-harm). We discuss some of the challenges related to ensuring the safety of youth when interacting with CAs regarding sexual and mental health topics. We call for discourse on how to build guardrails for the safe evolution of AI-based CAs for youth.

Additional Key Words and Phrases: Youth, Artificial Intelligence, Conversational Agents, Mental Health, Sexual Health, Information Seeking

1 INTRODUCTION

Conversational Agents (CAs) are systems enabled with the ability to interact with the users using natural human dialogue [32]. Recently, Microsoft made their Bing search engine available for conversational search [24], Google released their own chatbot Bard [12], and OpenAI released GPT-4 [27]. CAs (often called chatbots) can offer recommendations, answer questions, and converse with users. Now conversational agents are increasingly used by teens for learning and knowledge discovery [11] including mental and sexual health topics [34]. Exploring sexual health topics through online search has been an essential part of adolescent (ages 13-17) development [10, 31], and CAs can support their knowledge discovery on such topics through human-like dialogues. Yet, never-before encountered risks have been documented with teens' interactions with AI-based CAs, such as being exposed to sexually inappropriate content and/or being given advice that is detrimental to their mental and physical wellbeing (e.g., to self-harm) [7, 18]. For instance, weeks after Snapchat's new Large Language Model (LLM) based chatbot (i.e., *My AI*) was released, multiple claims have been reported about inappropriate content for teens that *My AI* generated including advice on how to have sex for the first time with an adult partner [7]. Such risks can arise due to generative AI models that are trained on presumably the entirety of the internet without any bounds, making CAs developmentally inappropriate for the uniquely social, yet vulnerable, stage of adolescence [17]. Considering the existing and potential harms that CAs can pose to youth, we need to start thinking about how to build guardrails to keep CAs safe for youth, especially related to sensitive topics, such as mental and sexual health. In this position paper, we present the current landscape of mental and sexual health CAs designed for youth, their potential benefits and challenges, and call for further research.

Authors' addresses: Jinkyung Park, Vanderbilt University, Nashville, USA, jinkyung.park@vanderbilt.edu; Vivek Singh, Rutgers University, New Brunswick, USA, v.singh@rutgers.edu; Pamela Wisniewski, Vanderbilt University, Nashville, USA, pamela.wisniewski@vanderbilt.edu.

2 CURRENT LANDSCAPE

Since the recent release of various CAs, the search has transformed into a more interactive process of chatting with conversational agents to be more about knowledge discovery. Now conversational agents are increasingly used by teens for learning and knowledge discovery [11] including mental and sexual health topics [34]. According to a recent survey, fifty percent of youth ages 12 to 18 reported having used ChatGPT for learning, and those who used ChatGPT are three times more likely to use it than search engines (e.g., Google) for future knowledge discovery [23]. As adolescents do not seek professional help for sensitive health problems due to perceived social stigma and embarrassment, confidentiality, and financial costs [29], CAs are applied in a variety of settings in the health domain.

2.1 Conversational Agents to Support Youth Mental Health

A plethora of work has been done to explore CAs to support adolescents' mental health including mental health interventions [1, 5, 6, 20], education/training for mental well-being [9, 13, 19, 26], and providing mental health-related information/resources [3, 21, 34]. For instance, educational content or relevant information based on Cognitive Behavioral Therapy (CBT) and positive psychology is delivered by CAs based on youths' inquiries related to mental health issues such as depression, anxiety, and stress [5, 8, 37]. In the majority of the existing work, user testing with the concepts and/or prototypes has been conducted to explore the acceptability, user engagement, and effectiveness of reducing mental health symptoms (e.g., depression, anxiety, alcohol abuse). The potential benefits of such mental health CAs include accessibility and serving as a middle ground for adolescents who are not comfortable with in-person conversations about their mental health needs. Adolescents in such situations benefit from access to resources [34] and emotional support by friendly and empathic, yet knowledgeable responses generated by CAs [15].

2.2 Conversational Agents to Support Youth Sexual Health

Compared to mental health, CAs to support youth's sexual health information seeking are under-studied. While youths suffer from various physical and mental challenges due to a lack of knowledge about sexual and reproductive health, they are reluctant to use specialized services and have limited awareness about their sexual and reproductive health issues [30]. CAs can be used for sexual health information seeking especially at-risk youth populations such as sexual minority adolescents, who are even less likely to seek professional care due to limited resources and social stigmatization [2, 14]. Recently, a few sexual health CAs for youth have been studied to facilitate adolescents' sexual information seeking [4, 21, 25, 30, 36]. The potential benefits of conversing with CAs for sexual health information seeking are their accessibility and ability to have non-judgemental conversation [28]. At the same time, one of the major barriers to using CAs was lacking human-like traits such as cognitive and affective empathy (e.g., the responses given by CAs were perceived as dry and generic) [25]. Users were also skeptical that CAs are capable of helping them when they felt anxious about their sexual health, specifically in the context of HIV, pregnancy, and other aspects of health that are perceived as highly consequential, potentially severe, and/or stigmatizing.

3 KEY CHALLENGES TO ADDRESS

Broadly, there are two technical approaches to building CAs: LLM-based and Rule-based. The major challenge of existing CAs in the mental and sexual health domain stemmed from the underlying technologies to develop the systems. We outline key challenges related to two different approaches that we would like to discuss with the workshop participants.

3.1 Challenges in Rule-based CAs: Restrictive and Less-Human Like Responses

Rule-based CAs are developed with pre-defined keywords and commands programmed by the developer. Most of the CAs in the healthcare domain were built upon pre-defined sets of responses based on domain-specific knowledge. For instance, 400 answers about the most frequently asked sexual health topics that the domain experts reviewed are encoded to CAs, and when users ask questions, CAs process users' input using natural language processing, and provide the most relevant answer from a set of 400 pre-defined answers by using rule-based decision tree techniques [4]. This means that the users are restricted to receiving predetermined answers to their questions, and there is little or no room for free responses. If a user enters a question or sentence without predetermined keywords, the conversational agents will be unable to understand the input and respond with a default message such as "Sorry, I did not understand." Despite these restrictions, simple conversational agents are frequently used in the sexual health domain [2]. As a result, interactions were perceived as limited in exploring individual issues and contexts, lacking sufficient depth to make clinical judgments and recommendations. CAs were considered restricted in offering personalized advice, as participants had doubts about the effectiveness of an algorithm or computer pattern being able to provide advice on sexual health and the complexities associated with lifestyles and activities [25]. The technology was seen as only providing advice about mainstream, easily accessible information, already available on the Internet. Subsequently, some struggled to understand the need for chatbots in sexual health [25]. Yet, the majority of the existing research [4, 21, 25, 30, 36] implemented rule-based approaches to develop CAs to provide mental and sexual health information for adolescents.

3.2 Challenges in LLM-based CAs: Unregulated, Offensive, and Inappropriate Responses

On the other hand, LLM-based CAs that are trained on presumably the entirety of web data [16] have the potential to tackle the above challenges. These conversational agents do not respond with pre-defined answers but rather generate coherent follow-up text to user inputs, trained on human-generated textual data such as Wikipedia content or social media posts [22]. Recent large language models (e.g., GPT-4) with a larger number of parameters have enabled a new paradigm of in-context learning, in which models understand input text written in human language (i.e., a prompt), and generate the following text that coherently follows the prompt [16]. With such capability, LLM-based CAs have the potential to undertake more complex tasks that involve greater interaction, reasoning, prediction, and accuracy [35].

Yet, research on LLM-based CAs is still sparse and in the early stage (almost nonexistent in the sexual health context). More critically, there are limitations and challenges already documented in designing LLM-based CAs, which are mainly resulting from the inherent characteristics of LLMs. Since LLMs have learned a vast amount of online text, there is a risk that the conversation flow can go beyond directions intended by the CAs designer [16]. Particularly, the risks inherent to LLM-based CAs can introduce adolescents to new types of risks such as being exposed to developmentally inappropriate and/or inaccurate content [18]. Such problems could be inherent in LLM-based conversational models as they generate responses with the highest probability of relevancy, making responses unfiltered and unregulated, and even offensive and inappropriate [17]. With human-like and authoritative responses from LLM-based CAs, it may be difficult for younger users (e.g., adolescents) to distinguish accurate information and fabricated answers [33]. Hence, designing developmentally appropriate and safer CAs for teens is pivotal for promoting their online safety. However, very little work has been done in the area of keeping teens safe with conversational agents. As a consequence, teens and parents are left without guidelines to set healthy boundaries, as well as policymakers have no regulations to require tech companies to abide by to keep younger users safe. Therefore, we call for more efforts to design and develop developmentally appropriate and safe LLM-based CAs for teens while fulfilling their information needs.

4 FUTURE DIRECTIONS

Our position is that we need to start thinking about how to build guardrails for the safe evolution of AI-based CAs for youth. Overall, safe CAs for youth need to be discussed at the intersection of AI technology, youth-centered design, and clinical support. Below are suggestions for future directions.

- In terms of AI technology, research on LLM-based CAs in general is still at the beginning and new approaches (e.g., prompt tuning) are needed to refine the models to generate developmentally appropriate content for youth.
- As early evidence showed that teens prefer friendly and empathic responses generated by CAs, research on the personality (e.g., older peer vs. health professional like) of CAs needs to be conducted.
- Finally, the effectiveness of CAs has mainly been examined by trials with concepts/prototypes with a small number of users. Therefore, more research efforts to evaluate fully functioning systems with a larger number of participants are needed.

5 CONCLUSION

Our research interests align with the purposes of the workshop to identify the major challenges for the new frontiers of generative AI with Generation Z themselves, activists, and industry practitioners. We hope the insights from the workshop will benefit us better understand the perspectives of the multi-stakeholders regarding the use of CAs for discovering knowledge about sensitive topics including sexual and mental health. Also, while we have identified some of the safety issues of LLM-based CAs to support adolescents' mental and sexual health information seeking, we hope that participating in the workshop will help us address some of the remaining challenges.

6 ABOUT THE AUTHORS

Jinkyung Park is a postdoctoral scholar in the Department of Computer Science at Vanderbilt University. Her research focuses on Human-Computer Interaction to promote online safety for youth and vulnerable populations.

Vivek Singh is an associate professor in the School of Communication and Information at Rutgers University. He designs AI systems that are responsive to human values and needs.

Pamela Wisniewski is an associate professor in the Department of Computer Science at Vanderbilt University. Her work lies at the intersection of Human-Computer Interaction, Social Computing, and Privacy. Her expertise helps her empower end users and teach students to understand the value of user-centered design and evaluation.

REFERENCES

- [1] Carlos Abreu and Pedro F Campos. 2022. Raising awareness of smartphone overuse among university students: a persuasive systems approach. In *Informatics*, Vol. 9. MDPI, 15.
- [2] Divyaa Balaji, Linwei He, Stefano Giani, Tibor Bosse, Reinout Wiers, and Gert-Jan de Bruijn. 2022. Effectiveness and acceptability of conversational agents for sexual health promotion: a systematic review and meta-analysis. *Sexual health* 19, 5 (2022), 391–405.
- [3] Francesca Beilharz, Suku Sukunesan, Susan L Rossell, Jayashri Kulkarni, Gemma Sharp, et al. 2021. Development of a positive body image chatbot (KIT) with young people and parents/carers: qualitative focus group study. *Journal of medical Internet research* 23, 6 (2021), e27807.
- [4] Erika Bonnevie, Tiffany D Lloyd, Sarah D Rosenberg, Kara Williams, Jaclyn Goldberg, and Joe Smyser. 2021. Layla's Got You: developing a tailored contraception chatbot for Black and Hispanic young women. *Health Education Journal* 80, 4 (2021), 413–424.
- [5] Gilly Dosovitsky and Eduardo Bunge. 2023. Development of a chatbot for depression: adolescent perceptions and recommendations. *Child and Adolescent Mental Health* 28, 1 (2023), 124–127.
- [6] Danielle Elmasri and Anthony Maeder. 2016. A conversational agent for an online mental health intervention. In *Brain Informatics and Health: International Conference, BIH 2016, Omaha, NE, USA, October 13-16, 2016 Proceedings*. Springer, 243–251.
- [7] Geoffrey A Fowler. 2023. Snapchat tried to make a safe AI. It chats with me about booze and sex. <https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/>

- [8] Silvia Gabrielli, Silvia Rizzi, Giulia Bassi, Sara Carbone, Rosa Maimone, Michele Marchesoni, and Stefano Forti. 2021. 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 and uHealth* 9, 5 (2021), e27965.
- [9] Silvia Gabrielli, Silvia Rizzi, Sara Carbone, Valeria Donisi, et al. 2020. A chatbot-based coaching intervention for adolescents to promote life skills: pilot study. *JMIR human factors* 7, 1 (2020), e16762.
- [10] Howard Gardner and Katie Davis. 2013. *The app generation: How today's youth navigate identity, intimacy, and imagination in a digital world*. Yale University Press.
- [11] Radhika Garg, Hua Cui, Spencer Seligson, Bo Zhang, Martin Porcheron, Leigh Clark, Benjamin R Cowan, and Erin Beneteau. 2022. The last decade of HCI research on children and voice-based conversational agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [12] Google. 2023. Bard. <https://bard.google.com/>
- [13] Christine Grové. 2021. Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in psychiatry* 11 (2021), 606041.
- [14] Christine YW Harb, Lauren E Pass, Isabella C De Soriano, Adelaide Zwick, and Paul A Gilbert. 2019. Motivators and barriers to accessing sexual health care services for transgender/genderqueer individuals assigned female sex at birth. *Transgender Health* 4, 1 (2019), 58–67.
- [15] Camilla Gudmundsen Høiland, Asbjørn Følstad, and Amela Karahasanovic. 2020. Hi, can I help? Exploring how to design a mental health chatbot for youths. *Human Technology* 16, 2 (2020), 139–169.
- [16] Eunkyung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [17] Samantha Murphy Kelly. 2023. The dark side of Bing's new AI chatbot. <https://www.cnn.com/2023/02/16/tech/bing-dark-side/index.html>
- [18] Samantha Murphy Kelly. 2023. Snapchat's new AI chatbot is already raising alarms among teens and parents. <https://www.cnn.com/2023/04/27/tech/snapchat-my-ai-concerns-wellness/index.html>
- [19] Theodora Koulouri, Robert D Macredie, and David Olakitan. 2022. Chatbots to support young adults' mental health: an exploratory study of acceptability. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 2 (2022), 1–39.
- [20] Florian Onur Kuhlmeier, Ulrich Gnewuch, Stefan Lüttke, Eva-Lotta Brakemeier, and Alexander Mädche. 2022. A Personalized Conversational Agent to Treat Depression in Youth and Young Adults—A Transdisciplinary Design Science Research Project. In *International Conference on Design Science Research in Information Systems and Technology*. Springer, 30–41.
- [21] Norma Leon Lescano, Eiriku Yamao, Elizabeth Xiomara Valladares Sánchez, and Miguel Angel Pablo Estrella Santillan. 2022. Iterative design and implementation of a chatbot for sexual and reproductive health counseling in Peru. In *2022 IEEE XXIX International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*. IEEE, 1–4.
- [22] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.
- [23] Common Sense Media. 2023. New Poll Finds Parents Lag Behind Kids on AI and Want Rules and Reliable Information to Help Them. <https://www.commonsensemedia.org/press-releases/new-poll-finds-parents-lag-behind-kids-on-ai>
- [24] Microsoft. 2023. Introducing the new Bing. <https://www.bing.com/?FORM=Z9FD1>
- [25] Tom Nadarzynski, Vannesa Puentes, Izabela Pawlak, Tania Mendes, Ian Montgomery, Jake Bayley, Damien Ridge, and Christy Newman. 2021. Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis. *Sexual health* 18, 5 (2021), 385–393.
- [26] Ginger Nicol, Ruoyun Wang, Sharon Graham, Sherry Dodd, and Jane Garbutt. 2022. Chatbot-delivered cognitive behavioral therapy in adolescents with depression and anxiety during the COVID-19 pandemic: feasibility and acceptability study. *JMIR Formative Research* 6, 11 (2022), e40242.
- [27] OpenAI. 2023. Introducing ChatGPT. <https://openai.com/blog/chatgpt>
- [28] Hyanghee Park and Joonhwan Lee. 2020. Can a Conversational Agent Lower Sexual Violence Victims' Burden of Self-Disclosure?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [29] Jerica Radez, Tessa Reardon, Cathy Creswell, Peter J Lawrence, Georgina Evdoka-Burton, and Polly Waite. 2021. 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* 30 (2021), 183–211.
- [30] Rifat Rahman, Md Rishadur Rahman, Nafis Irtiza Tripto, Mohammed Eunus Ali, Sajid Hasan Apon, and Rifat Shahriyar. 2021. AdolescentBot: Understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in Bangladesh. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [31] Afsaneh Razi, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2020. Let's talk about sext: How adolescents seek support and advice about their online sexual experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [32] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction* 37, 1 (2021), 81–96.
- [33] Melissa Rudy. 2023. Teens are turning to Snapchat's 'My AI' for mental health support — which doctors warn against. <https://www.foxnews.com/health/teens-turning-my-ai-mental-health-support-which-doctors-warn-against>

- [34] Gabriella Sanabria, Karah Y Greene, Jennifer T Tran, Shelton Gilyard, Lauren DiGiovanni, Patricia J Emmanuel, Lisa J Sanders, Kristin Kosyluk, and Jerome T Galea. 2023. "A Great Way to Start the Conversation": Evidence for the Use of an Adolescent Mental Health Chatbot Navigator for Youth at Risk of HIV and Other STIs. *Journal of Technology in Behavioral Science* (2023), 1–10.
- [35] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinakaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research* 22, 8 (2020), e17158.
- [36] Hua Wang, Sneha Gupta, Arvind Singhal, Poonam Muttreja, Sanghamitra Singh, Poorva Sharma, and Alice Piterova. 2022. An artificial intelligence chatbot for young people's sexual and reproductive health in india (snehai): Instrumental case study. *Journal of Medical Internet Research* 24, 1 (2022), e29969.
- [37] Ruth Williams, Sarah Hopkins, Chris Frampton, Chester Holt-Quick, Sally Nicola Merry, and Karolina Stasiak. 2021. 21-day stress detox: open trial of a universal well-being chatbot for young adults. *Social Sciences* 10, 11 (2021), 416.