The uses of chatbots in the context of children and teenagers bullying: a systematic literature review

Laura Iseut Lafrance St-Martin & Stéphane Villeneuve

Published online: 09 Feb 2024.

Submit your article to this journal

Article views: 838

View related articles

View Crossmark data

cogent

INFORMATION & COMMUNICATIONS TECHNOLOGY IN EDUCATION | REVIEW ARTICLE

🔓 OPEN ACCESS | Check for updates

The uses of chatbots in the context of children and teenagers bullying: a systematic literature review

Laura Iseut Lafrance St-Martin[a] (iD) and Stéphane Villeneuve[b] (iD)

[a]École des arts numériques, de l'animation et du design, Université du Québec à Chicoutimi, Montréal, Canada; [b]Département de didactique, Université du Québec à Montréal, Montréal, Canada

ABSTRACT

This paper presents the results of a systematic literature review concerning the actual and future uses of chatbots (artificial conversation agents) in youth bullying cases. The study found that while artificial intelligence is highly regarded as an interesting tool by researchers, the technology is not yet good enough to intervene in crisis cases. Indeed, the studies show that chatbots are still inaccurate regarding emotion detection, their language is not adapted to children's way of speaking and writing, and they are too predictable. Despite those limitations, the most promising use of chatbots regarding bullying is as a way to raise awareness about it among children and teenagers. Chatbots could then become a tool in preventing bullying cases. This paper also presents a review of researchers' rhetoric about chatbots, the reasons why youths want to interact with chatbots, and how the use of chatbots in intervention is going to affect the service providers' work.

SHORT SUMMARY

Artificial intelligence is currently seen as a promising solution to many problems. This article is part of a partnership research, funded by the Social Sciences and Humanities Research Council of Canada, with a technology development company. This research aims to develop a chatbot to facilitate the reporting and management of cases of school bullying. Specifically, this systematic literature review identifies the main potential uses of this technology in youth bullying intervention. Laura Iseut Lafrance St-Martin has been conducting critical research on the use of AI for several years and Stéphane Villeneuve has been working on bullying, cyberbullying, and the integration of technologies in education.

Introduction

Bullying is a major problem for children and teenagers (Biswas et al., 2020). It does not only affect mental health and social skills, but it can also have consequences lasting even after becoming an adult (Man et al., 2022). In more extreme cases, it can lead to suicidal thoughts and, ultimately, to death (Zhu et al., 2022). The severity of bullying effects on youths (i.e. underage persons, the age can change depending on countries) makes it an intervention priority for several researchers. How can we reduce the prevalence of bullying and its effects on children and teenagers? Can we develop new ways to intervene or equip service providers (i.e. a person who intervenes with the victims or witnesses, usually psychologists, social workers, or specialized educators) with new tools? Within those research interests, new technologies are often seen as promising.

In the last few years, artificial intelligence and, more precisely, chatbots have been studied to intervene in bullying contexts. Considering the recent nature of this topic, only one systematic literature review was done on the use of technologies in the context of bullying (Iivari et al., 2021). The research

involved the initial sorting of search results by relevance and the subsequent inclusion of papers into an Excel sheet based on predefined criteria, collaboratively categorizing papers into general themes, and focusing on research methods used to study bullying and the application of design and technology to address bullying. The authors found in their literature review that there were many limitations such as (1) many studies are mainly concentrated in North America and Europe, (2) pre-school children are less represented in research (3) the cultural context can play a major role. Nonetheless, this study had a broad scope and did not specifically address the use of chatbots.

This article reviews recent research on chatbots and bullying, highlighting their potential role in addressing this issue. Since it is still a new area of research, most papers were in the ideation or designing phase. Nonetheless, a literature review can help grasp how chatbots can help regarding youth bullying. This article presents the results of various studies concerning: (1) chatbots' general objectives, (2) researchers' rhetoric, (3) youth's relation to chatbots, (4) effects on service providers, and (5) the intervention's effectiveness.

Research method

This research was guided by the PRISMA model (Liberati et al., 2009) which helps to determine inclusion/exclusion criteria, the search strategy, the study selection, and data extraction processes (presented in Appendix 1). This research was guided by the following question: Are chatbots a promising avenue for reducing the prevalence of bullying among children and adolescents? As many researchers have shown, bullying is not a problem easy to fix. Therefore, there is no simple way for researchers to assess if chatbots can be a possible way to reduce the prevalence of bullying and cyberbullying among children and teenagers.

We divided the principal question into three questions. As we wanted to study the reasons why chatbots are seen as a promising avenue, we asked (RQ1) Why use a chatbot in the context of bullying? In other words, what is the researchers' rhetoric behind this choice? We also asked (RQ2) what motivates the use of conversational agents by children and adolescents. Indeed, researchers and users often have different reasons for using chatbots. Then, we asked (RQ3) how conversational agents influence (positively or negatively) the work of practitioners. As the practitioners are currently crucial in the effort to reduce the prevalence of bullying and intervening when bullying occurs, we wanted to determine how the chatbots affect their work: are they designed to help or to replace the practitioner? Globally, in the case of already existing chatbot-based intervention, we tried to assess whether chatbots effectively intervene in bullying among children and adolescents.

To find the relevant papers, we search the following keywords (Chatbot, bot, conversation agent, natural language processor, bully, harass, child, teenager) in Science Direct, Scopus, ACM, IEEE Xplore, ERIC database, EBSCOhost, Wiley library, Google Scholar (first 10 pages). The research was done in November 2022 and did not have a time exclusion criterion. The sources were uploaded into PICOportal, an online platform that helps with the various steps of a systematic literature review which uses the PRISMA method. To be selected articles needed to be about children or teenagers (they were excluded when about the adult population), the intervention needed to be with a chatbot technology (excluded when it was not the main intervention) and they needed to be mainly about bullying (excluded when not primarily about it). After sorting the 660 initial papers, our literature review includes 28 articles, chapters, and theses.

Results

Overview of the primary uses of chatbots and data

This section aims to give a general overview of the research included in our literature review. In 44% of the studies, chatbots were intended for children and teenagers who are victims of bullying. Forty-four percent of chatbots were aimed at children and teenagers in general (meaning not only bullying victims or perpetrators). These studies aimed at the youth's general population in a preventive manner. Only a few studies work on chatbots for adult users who need to intervene in children and teenagers' bullying

cases (7% for adults in general, such as parents, and 4% specifically for teachers and service providers). It then becomes clear that researchers view chatbots as tools for youths and not for adults.

Regarding chatbots' objectives, 52% were conceived to support victims, either emotionally or with the denunciation process. Thirty percent of chatbots in our literature review aimed to inform children and teenagers about the various forms of bullying and their consequences. Fifteen percent were designed to support victims during the legal procedures of bullying cases. Other intended uses include detecting online bullying behaviors (7%), adult training (7%), and direct intervention (7%). Our data also included another literature review which is not included in this because it presents various uses of ICT in the bullying context. This overview highlights how supporting the victims is the researchers' and developers' primary concern.

Seventy-nine percent of the chatbots were not explicitly designed to be used in a school context. Rather their understanding of bullying is more general. Similarly, 54% of them were not specific to bullying: While bullying was part of their intervention scope, it was part of a more extensive approach (e.g. supporting youths with mental health issues, bullying is a frequent cause of mental health issues for them). Finally, 52% of the papers presented design or co-design studies (generally without testing), showing that this study area is still in the early stages. Other studies included design testing (33%) and intervention (15%).

RQ1: why use a chatbot in the context of bullying?: Researchers' rhetoric

Our first research question is about why researchers think that chatbots are a promising avenue for research and intervention in bullying and cyberbullying. Indeed, papers often begin with authors justifying why it is worthwhile to explore the possibilities offered by artificial intelligence and chatbots. Almost all the papers mentioned accessibility, in terms of technical support (often cell phones) and time, as one of the main reasons they chose to work with a chatbot. The latter is the main reason researchers see chatbots as a promising way to address bullying. Indeed, as Santos and colleagues said, 'By providing access to mental health services, these chatbots can offer immediate and accessible help and support that does not require medical treatment' (2020, p. 2).

Depending on countries, regions, and social classes, professional help can be difficult to find, or even completely inaccessible, for youths. In this context, a chatbot accessible via a cellphone or school platform can provide temporary support for minor problems or while waiting for a meeting with a mental health professional. In some cases, speaking with someone (or even with a digital agent) can offer some relatively accessible relief: 'Conversational technology can partially alleviate […] problems by offering private, accessible, convenient, and synchronous communicative solutions and act as a virtual audience for adolescents to engage with' (Lopatovska et al., 2022, p. 2). Despite being seen as 'accessible', it is worth remembering that cell phones and the Internet are not accessible to a large portion of the population, especially in the global South, and that technological solutions contribute to accentuating inequalities.

One study from Thailand specified that chatbots could also help gain accessible legal advice, sometimes needed in the context of bullying and sexual exploitation. Indeed, as chatbots are seen as private and safe, youths can access information without feeling judged and having to pay (Socatiyanurak et al., 2021).

Another important reason why researchers see chatbots as a promising avenue is children's and teenagers' comfort. First, they are at ease with technology (Ahn et al., 2020; Elgibreen et al., 2020; Grové, 2020; Grundmann, 2022; Iivari et al., 2021; Piccolo et al., 2021). Having grown up with several types of technological devices (cell phones, tablets, computers, gaming consoles, etc.) makes it easy for them to understand the technology and to relate with digital agents. Moreover, the interactivity of chatbots helps reach out to and engage children and teenagers (Rita & Shava, 2021) as they are used to fast-paced digital interactions.

Despite being born with technology, including AI and other virtual agents, youths are comfortable with them for various reasons including privacy and the feeling that chatbots do not judge them. Indeed, several researchers affirmed in their papers that children and teenagers have an impression of privacy and anonymity when interacting with digital agents (Agarwal et al., 2021; Høiland, 2019;

Lopatovska et al., 2022). When questioned about their feeling toward texting with a machine, children displayed comfort with the privacy of such an exchange (Agarwal et al., 2021, p. 29). This impression of privacy can help youths express personal, and even shameful thoughts, as chatbots are perceived as non-judgmental (Ahn et al., 2020; Høiland, 2019; Sanoubari et al., 2021). It then gives them an outlet to express themselves without fearing a negative backlash. 'Additionally, as youths may perceive it as shameful to seek help for mental difficulties (Solvang & Kilsti, 2000; Tveit, 1998), being anonymous on digital health services may make it easier to bring up difficult and taboo issues (Jensen, 2014)' (Høiland, 2019, p. 10). To summarize, as youths perceive chatbots as private and non-judgmental, they encourage self-disclosure and self-expression (Ahn et al., 2020; Iivari et al., 2021; Piccolo et al., 2021; Rajwal, 2023): '[According to Bethel et al. (2017), children] are able to […] share more sensitive information with the robot-interviewer' (Iivari et al., 2021, p. 5).

Researchers also note that chatbots can give learning opportunities to youths and adults alike. Indeed, it provides a place to practice repeated tasks for adults (such as teachers and service providers) to learn how to perform interventions in bullying cases (Grundmann, 2022; Schussler et al., 2017). Chatbots can be used for valuable knowledge input about emotions (Rajwal, 2023; Young & Dautenhahn, 2022) and legal procedures (Morgan et al., 2018). They can be used as a safe place for experimentation. 'We believe robots can provide a safe medium for children to explore emotional and sensitive issues because, unlike humans, they are non-judgmental. Also, they quite literally enable children to see the world from another being's eyes, which facilitates perspective-taking' (Sanoubari et al., 2021, p. 179). In the case of children on the autism spectrum, chatbots can provide useful tools to nurture and reinforce social skills (Ireland et al., 2018).

Some other reasons were told, although by fewer papers. First, chatbots can be a way to automate some tasks like parts of therapy (Bae Brandtzæg et al., 2021), emotional support (Piccolo et al., 2021; van der Zwaan et al., 2012; Wiederhold & Riva, 2012) emotion detection (Laorden et al., 2013), intervention (Young & Dautenhahn, 2022), or even data collection (Gaci et al., 2020; Kamar et al., 2022). Indeed, 'Since one-on-one online counselling is very labour intensive, automating this kind of support could help to reach more victims' (Wiederhold & Riva, 2012, p. 243) by delegating and automating some parts of the job. Second, chatbots are relatively new: in the context of hard-to-solve problems like bullying, sometimes just trying something new is worth it (Grové, 2020; Young Oh et al., 2020), even without a specific idea of how it could help. Third, in the specific case of robots, researchers affirm that having a body might increase the efficacy of the intervention, especially with young children.

In brief, the data we collected shows that researchers see chatbots as a potential solution to bullying for many reasons, but mainly because it is a technology always accessible via cell phone and internet connection and youths seem comfortable interacting with them. Privacy and lack of judgment are the main reasons why they think it is promising to explore this intervention avenue. This could reduce the stress on school counselling services and mental health professionals.

RQ2: why do children and teenagers want a chatbot?

The following research question concerns youths' motivation for engaging with a chatbot in the context of bullying and other similarly sensitive subjects. The papers in our literature review agree that comfort is the main reason. That being said, children and teenagers are comfortable with chatbots for various reasons. Interacting with a chatbot reduces social difficulties usually encountered with peers and adults (Bae Brandtzæg et al., 2021; Gaci et al., 2020; Ireland et al., 2018; Piccolo et al., 2021; Santos et al., 2020; Socatiyanurak et al., 2021). Gaci and colleagues present in their articles the current knowledge about the comfort of youth with chatbots and digital agents, most of them being social in nature. Some researchers (Lucas et al., 2014; Pickard et al., 2016) showed that children prefer to use virtual agents because they feel less judged, and criticized and without verbal or non-verbal reactions. Also, the digital interface makes it easier for children and teenagers to text with virtual agents as they usually do with their friends. Finally, they feel less anxiety, embarrassment, and guilt when discussing with chatbots adding to a feeling of privacy (Gaci et al., 2020, p. 459).

Indeed, it is hard to speak about bullying and sexual grooming for children and teenagers (Rajwal, 2023; Young Oh et al., 2020). Chatbots then provide an always accessible (Gaci et al., 2020; Høiland,

2019) 'safe place' for them to express themselves and seek out advice (Bae Brandtzæg et al., 2021; Piccolo et al., 2021; Santos et al., 2020). This feeling of safety is partly due to the anonymity of online discussions (Bae Brandtzæg et al., 2021; Gaci et al., 2020; Piccolo et al., 2021; Santos et al., 2020). Moreover, 'Teenagers are referred to as "mobile natives," who grew up using smartphones from an early age. They display symptoms of "call phobia", which means they prefer to communicate by text, mobile messenger, or email rather than by voice' (Ahn et al., 2020, p. 2). Texting with a digital agent then seems to be perfectly adapted for youths.

Despite the comfort of youths, other reasons were present in the papers. Children and teenagers usually regard technologies and artificial intelligence positively (Rita & Shava, 2021; Sanoubari et al., 2021; Wiederhold & Riva, 2012), which makes them want to interact with it. Similarly, youths are often curious about new technology (Agarwal et al., 2021). These last two reasons motivate children and teenagers and can be used by researchers to help them intervene in bullying. Interestingly, two papers also noted that some youths want to build personal relationships with chatbots (Agarwal et al., 2021; Gaci et al., 2020).

> Children kept coming back to chat even after the day's content was over and they wanted to establish a personal relationship with a chatbot called Wulu. However, Wulu is yet to evolve to be able to respond to friendly messages and build a trusting relationship. (Agarwal et al., 2021, p. 31)

Having a chatbot become a permanent social presence in their life could be a promising way to monitor their mental health and provide accessible emotional relief, especially in the case of minor issues (which do not necessarily require human intervention).

To summarize and answer this research question, children and teenagers want to interact with a chatbot in the context of bullying mainly because they are more comfortable discussing sensitive subjects with digital agents. Indeed, since the low rate of denunciation is a problem regarding bullying, chatbot-based platforms could be a way to denunciate and bypass the shame around bullying and the need for help. In addition, chatbots are a relatively accessible way to provide youths with quick emotional support.

RQ3: how do chatbots affect the work of service providers?

Within the social context of school, most of the work regarding bullying cases is done by service providers like social workers. Our third research question aims to understand how the use of chatbots affects (or could affect) their work. A lot of papers agree that it can reduce cases and workload by raising awareness and educating youths (Agarwal et al., 2021; Ahn et al., 2020; Høiland, 2019; Laorden et al., 2013; Louchart et al., 2004; Rita & Shava, 2021; van der Zwaan et al., 2012; Young Oh et al., 2020; Young & Dautenhahn, 2022). In the specific case of autism spectrum children, chatbots are valuable tools for practicing social skills. Those skills are crucial to later deal with bullying, especially for social resilience and reducing victimization (Ireland et al., 2018).

As a potential education and intervention tool, chatbots can help prevent bullying cases and have a profound and lasting effect on the number of cases. Similarly, chatbots can provide tools for intervening (Elgibreen et al., 2020; Ireland et al., 2018; Sanoubari et al., 2021; Wiederhold & Riva, 2012). Nonetheless, efforts to prevent bullying always had mitigated results, it is probably naïve to think that technologies and chatbots can have a significant effect where all the other methods failed. That being said, bullying is an essential problem for youth's mental health and well-being (which can have lasting effects on adult life). In that context, any reduction of cases (even minor) is still worth a try.

Some chatbots also try to make bullying cases easier to report (Gaci et al., 2020; Piccolo et al., 2021; Rita & Shava, 2021; Socatiyanurak et al., 2021). As children and teenagers find it difficult to talk about bullying to adults, a digital agent can act as an interface between the service provider and the victim. By initiating contact between them (often after asking the child or teenager if they want to do an official report), youths might feel less intimidated or ashamed to ask for help (Gaci et al., 2020; Grové, 2020;; Morgan et al., 2018; Piccolo et al., 2021; Rita & Shava, 2021; van der Zwaan et al., 2012). During the initial conversation with the chatbot, it can gather relevant information and data about the case (Morgan et al., 2018), thus the service provider does not need to ask factual questions. While gathering data

about the case, chatbots can also act as a filter, reporting serious cases to professionals: 'At the very least the anti-cyberbullying agent should be able to detect and deal with cases it cannot handle, either by referring the user to a specialized helpline, or call in a human counselor that takes over the conversation' (van der Zwaan et al., 2012, p. 7).

In the context of scientific research, chatbots can also be used to perfect knowledge, which can, in turn, be used to adapt service providers' work. For example, one study used chatbots to act as children on websites known as hunting grounds for sexual groomers. The chatbots were used to attract and observe the groomer's behaviors. The study aimed to evaluate how different familial settings discourage sexual groomers (Kamar et al., 2022). The different types of parenting (passive, active, or absent) displayed by those chatbots help understand how parents should act to undermine the predator. This type of knowledge has a scientific as well as intervening value. On the negative side, one study noted that, even though they tried to design a chatbot to help service providers practice their intervention skills, it undermined their confidence (Grundmann, 2022).

Globally, regarding chatbots' influence on service providers, like counselors, data shows that they could help reduce their workload, by dealing with minor cases, acting as a facilitator, and collecting data about cases. In addition, chatbots could also offer tools to raise awareness and prevent bullying cases, thus reducing the number of youths in urgent need of care. Chatbot-based interventions thus seem to be a promising avenue to help service providers.

Discussion: is chatbot-based intervention working?

Globally, our study aimed to assess if chatbot-based interventions currently in place or being tested seem to work. Since most of the research in our literature review was still in the ideation/design phase, there were not a lot of results for this analysis. Nonetheless, we can present a brief overview of the main results. Several papers state that it is not yet working, but could be effective mainly because youth seems to want it (Bae Brandtzæg et al., 2021; Grové, 2020; Høiland, 2019). However, one study shows that not all children and teenagers like to interact with chatbots (Bae Brandtzæg et al., 2021, p. 10), so it could be an effective solution only for a certain portion of them.

For other researchers, it is too early to know if it will be an effective way to intervene in bullying cases (Elgibreen et al., 2020):

> While chatbots are arguably seen as the sexy new solution to autism intervention, continuous improvements must be made to existing interventions and all emerging interventions need to be validated for their efficacy to drive real-world outcomes for individuals on the spectrum (Ireland et al., 2018, p. 116).

One of the reasons it seems to be too early is that there are still too many errors at the moment. For example, 'The chatbot cannot correctly detect the child's emotion all the time. This is caused by a limitation in EREN's [the chatbot] emotion detection ability which currently relies on assessing the nouns, verbs, and adjectives found in the child's input' (Santos et al., 2020, p. 9). Another study states that the chatbot's language has yet to be adapted to how children and teenagers speak to be effective (Laorden et al., 2013). A researcher notes that the chatbot is too predictable in the long run (Grové, 2020) and needs to be less repetitive to build a relationship with youths.

For one study, the results were negative, meaning that it is not only not working, but it harmed the people using the chatbot (Grundmann, 2022). This study developed a chatbot to train future helpline counselors with various intervention scenarios, but the results show that it undermined their confidence. Nonetheless, as the author says, it can perhaps be explained by the fact that the trainees were more conscious of their failings, which can still be positive in a learning context.

Chatbots seem to be effective in one area in particular: altering the attitude of children and teenagers concerning bullying and raising awareness about its consequences (Agarwal et al., 2021; Bae Brandtzæg et al., 2021; Ueda et al., 2021; Young Oh et al., 2020). For example, in one study aiming to raise awareness about various sensitive subjects in India, 'Most participants indicated that they liked the content on bullying the most. This is also supported by the greater positive shift in attitude under questions regarding bullying as observed in the quantitative data' (Agarwal et al., 2021, p. 25).

In summary, altering youth's attitudes seems to be a promising avenue of chatbot-based intervention, even though it is still early and lots of work is yet to perfect chatbots' behavior.

Conclusion: are chatbots a promising avenue for reducing the prevalence of bullying among children and teenagers?

Considering these results, we ask ourselves if chatbots can help reduce the prevalence of bullying among children and teenagers. The most promising way to do that seems to be by altering the attitude of youths toward bullying and acts of violence. Indeed, several studies agree that chatbot is an efficient way to raise awareness among children and teenager about the behavior associated with bullying and its effects (Agarwal et al., 2021; Bae Brandtzæg et al., 2021; Elgibreen et al., 2020; Piccolo et al., 2021; Ueda et al., 2021; Young Oh et al., 2020; Young & Dautenhahn, 2022). Also, as an education tool, chatbots could empower youths to react when they see bullying acts, which is a promising way to reduce the duration of bullying relationships since it often stops when a witness intervenes.

Our results also show other ways chatbots could help with bullying. First, they could help by providing emotional support to children and teenagers (Gaci et al., 2020; Grové, 2020;; Høiland, 2019; Morgan et al., 2018; Rajwal, 2023; Santos et al., 2020). While they cannot replace social workers and other trained humans, they could help with minor cases where youths mainly need to express themselves. As such, chatbots could also act as a denunciation interface, which gathers facts about the case and transfers them to a service provider (Gaci et al., 2020; Ireland et al., 2018; Rita & Shava, 2021). Indeed, since it is easier for children and teenagers to talk about bullying to chatbots than to humans, it could be an effective way to increase the number of cases reported. In the case of cyberbullying on social media and other public forums, chatbots could help detect bullying patterns and then intervene in the discussion to prevent it from degenerating (Laorden et al., 2013). Finally, chatbots could help with service providers' formation, although it would need several major adjustments (Grundmann, 2022).

Even though several researchers seem to believe that chatbots are an interesting way to reduce the prevalence of bullying among children and teenagers, our results show that it is still too early to evaluate its effectiveness, we were not able to answer our primary research question. Indeed, most current studies show technological limitations that undermine the intervention's assessment or are still in the conceptual phase. Nonetheless, we were able to identity some futures avenues for research. Researchers agree that there is still a lot of work to be done before chatbots can effectively answer youth's needs in the context of bullying mainly because chatbot technology is still limited in emotion recognition, adaptation to youth's language and conversation are often repetitive. Nonetheless, AI technology is constantly progressing, it is important to test new AI development to see how it could address those current technological limitations in bullying intervention. In this context, future studies should focus on adapting AI to the specific emotional and social needs of children and teenagers. Also, empirical studies are needed to assess the effectiveness of the new developments regarding prevention and intervention.

Disclosure statement

Funding

About the authors

Laura Iseut Lafrance St-Martin is a professor at the Université du Québec à Chicoutimi, specializing in video game studies and narrative design. Her research focuses mainly on the experience of players and video games as a communication tool, particularly from the perspective of ethical design.

Stéphane Villeneuve is a professor at the Université du Québec à Montréal, specializing in digital integration in education. His expertise lies in the professional competence to integrate digital tools for future and practicing teachers, as well as addressing cyberbullying in education.

ORCID

Laura Iseut Lafrance St-Martin 🆔 http://orcid.org/0000-0002-9607-5167
Stéphane Villeneuve 🆔 http://orcid.org/0000-0002-7538-9562

References

Agarwal, D., Agastya, A., Chaudhury, M., Dube, T., Jha, B., Khare, P., & Raghu, N. (2021). *Measuring effectiveness of chatbot to improve attitudes towards gender issues in underserved adolescent children in India*. Harvard Kennedy School.

Ahn, Y., Zhang, Y., Park, Y., & Lee, J. (2020, April 25). *A chatbot solution to chat app problems: Envisioning a chatbot counseling system for teenage victims of online sexual exploitation* [Paper presentation]. CHI EA '20.

Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., & Følstad, A. (2021, May 7). *When the social becomes non-human: Young people's perception of social support in chatbots* [Paper presentation]. CHI '21.

Bethel, C. L., Henkel, Z., Eakin, D. K., May, D. C., & Pilkinton, M. (2017). Moving toward an intelligent interactive social engagement framework for information gathering. In *Proceedings of the 15th IEEE International Symposium on Applied Machine Intelligence and Informatics (SAMI 2017)*. IEEE.

Biswas, T., Scott, J. G., Munir, K., Thomas, H. J., Huda, M. M., Hasan, M. M., David de Vries, T., Baxter, J., & Mamun, A. A. (2020). Global variation in the prevalence of bullying victimisation amongst adolescents: Role of peer and parental supports. *EClinicalMedicine*, *20*, 100276. https://doi.org/10.1016/j.eclinm.2020.100276

Elgibreen, H., Almazyad, S., Shuail, S. B., Qahtani, M. A., & Alhwiseen, L. (2020). Robot framework for anti-bullying in Saudi schools. *Fourth IEEE International Conference on Robotic Computing (IRC)*, Taichung, Taiwan, (pp. 166–171).

Gaci, M., Vonèche Cardia, I., & Gillet, D. (2020). Privacy-aware digital mediation tools for improving adolescent mental well-being: application to school bullying. In *Proceedings of the 13th International Conference on Advances in Computer-Human Interactions (ACHI 2020)*. DigitalWorld.

Grové, C. (2020). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, *11*, 606041. https://doi.org/10.3389/fpsyt.2020.606041

Grundmann, S. (2022). *A BDI-based virtual agent for training child helpline counsellors,* Delft University of Technology [Master's thesis].

Høiland, C. G. (2019). *"Hi, can I help?" An exploratory study of designing a chatbot to complement school nurses in supporting youths' mental health*, University of Oslo [Master's Thesis].

Iivari, N., Ventä-Olkkonen, L., Sharma, S., Molin-Juustila, T., & Kinnunen, E. (2021, May 7). *CHI against bullying: taking stock of the past and envisioning the future*. CHI '21.

Ireland, D., Bradford, D., & Farr-Wharton, G. (2018). Social Fringe Dwellers: Can chat-bots combat bullies to improve participation for children with autism? *The Journal of Community Informatics*, *14*(1), 105–119. https://doi.org/10.15353/joci.v14i1.3405

Jensen, K. L. (2014). *Fra offline til online: internetbaseret rådgivning og terapi*. Dansk Psykologisk Forlag.

Kamar, E., Maimon, D., Weisburd, D., & Shabat, D. (2022). Parental guardianship and online sexual grooming of teenagers: A honeypot experiment. *Computers in Human Behavior*, *137*, 107386. https://doi.org/10.1016/j.chb.2022.107386

Laorden, C., Galán-García, P., Santos, I., Sanz, B., Hidalgo, J. M. G., & Bringas, P. G. (2013). Negobot: A conversational agent based on game theory for the detection of paedophile behaviour. In *International Joint Conference CISIS'12-ICEUTE 12-SOCO 12 Special Sessions*. Springer Berlin Heidelberg.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine*, *151*(4), W65–94. https://doi.org/10.7326/0003-4819-151-4-200908180-00136

Lopatovska, I., Turpin, O., Davis, J., Connell, E., Denney, C., Fournier, H., Ravi, A., Yoon, J. H., & Parasnis, E. (2022). *Capturing teens' voice in designing supportive agents*. CUI '22.

Louchart, S., Aylett, R., Pickering, J., & Romano, D. (2004). "Don't hit me!"–Combining language and action in FearNot! AISB 2004 Convention: Motion, Emotion and Cognition, University of Leeds, United Kingdom.

Lucas, G. M., et al. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100.

Man, X., Liu, J., & Xue, Z. (2022). Effects of bullying forms on adolescent mental health and protective factors: A global cross-regional research based on 65 countries. *International Journal of Environmental Research and Public Health*, *19*(4), 2374. https://doi.org/10.3390/ijerph19042374

Morgan, J., Paiement, A., Seisenberger, M., Williams, J., & Wyner, A. (2018). A chatbot framework for the children's Legal centre. In *Legal knowledge and information systems* (pp. 205–209). IOS Press.

Piccolo, L. S. G., Troullinou, P., & Alani, H. (2021). *Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements* [Paper presentation]. DIS '21: Designing Interactive Systems Conference 2021, 20210628. https://doi.org/10.1145/3461778.3462114

Pickard, M. D., et al. (2016). Revealing sensitive information in personal interviews: Is selfdisclosureeasier with humans or avatars and under what conditions? *Computers in Human Behavior*, *65*, 23–30.

Rajwal, S. (2023). Design of a chatbot for four-to ten-year-old children based on emotional intelligence. In: Gupta, D., Khanna, A., Bhattacharyya, S., Hassanien, A. E., Anand, S., Jaiswal, A. (Eds.), *International Conference on Innovative Computing and Communications*. Lecture Notes in Networks and Systems, vol 473. Springer. https://doi.org/10.1007/978-981-19-2821-5_57

Rita, M. N., & Shava, F. B. (2021, August). *Chatbot driven web-based platform for online safety and sexual exploitation awareness and reporting in Namibia* [Paper presentation]. 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa. https://doi.org/10.1109/icABCD51485.2021.9519375

Sanoubari, E., Muñoz Cardona, J. E., Mahdi, H., Young, J. E., Houston, A., & Dautenhahn, K. (2021, June 24). *Robots, bullies and stories: A remote co-design study with children* [Paper presentation]. In Proceedings of the 20th Annual ACM Interaction Design and Children Conference (IDC '21). Association for Computing Machinery, 171–182. https://doi.org/10.1145/3459990.3460725

Santos, K.-A., Ong, E., & Resurreccion, R. (2020). *Therapist vibe: children's expressions of their emotions through storytelling with a chatbot* [Paper presentation]. Proceedings of the 20th Annual ACM Interaction Design and Children Conference. Association for Computing Machinery, New York, USA.

Schussler, D., Frank, J., Lee, T.-K., & Mahfouz, J. (2017). Using virtual role-play to enhance teacher candidates' skills in responding to bullying. *Journal of Technology and Teacher Education*, *25*(1), 91–120.

Socatiyanurak, V., Klangpornkun, N., Munthuli, A., Phienphanich, P., Kovudhikulrungsri, L., Saksakulkunakorn, N., Chairaungsri, P., & Tantibundhit, C. (2021). LAW-U: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. *IEEE Access*. *9*, 131440–131461. https://doi.org/10.1109/ACCESS.2021.3113172

Solvang, L., & Kilsti, S. (2000). *"Ungdom og mental helse.".* Du skal ikke ha psykiske problemer. Slik er det bare [Paper presentation] En kvalitativ undersøkelse for helse-og sosialdepartementet og organisasjonen Mental helse. Norsk Gallup Institutt A/S.

Tveit, A. (1998). *Ungdoms tanker om mental helse. Hovedoppgave i helsefag*. Norges teknisk-naturvitenskapelige universitet.

Ueda, T., Nakanishi, J., Kuramoto, I., Baba, J., Yoshikawa, Y., & Ishiguro, H. (2021, November 9). Cyberbullying mitigation by a proxy persuasion of a chat member hijacked by a chatbot [Paper presentation]. *HAI '21*, online event.

van der Zwaan, J. M., Dignum, V., & Jonker, C. M. (2012). A BDI dialogue agent for social support: Specification and evaluation method. In *Proceedings of the 3rd Workshop on Emotional and Empathic Agents@ AAMAS*. p. 1–8.

Wiederhold, B. K., & Riva, G. (2012). User validation of an empathic virtual buddy against cyberbullying. *Studies in Health Technology and Informatics*, *181*, 243–247.

Young, J., & Dautenhahn, K. (2022). *Designing an anti-bullying serious game: insights from interviews with teachers*. In *Joint International Conference on Serious Games* (pp. 102–121). Cham: Springer International Publishing.

Young Oh, E., Song, D., & Hong, H. (2020). Interactive computing technology in anti-bullying education: The effects of conversation-bot's role on K-12 students' attitude change toward bullying problems. *Journal of Educational Computing Research*, *58*(1), 200–219. https://doi.org/10.1177/0735633119839177

Zhu, X., Griffiths, H., Eisner, M., Hepp, U., Ribeaud, D., & Murray, A. L. (2022). Developmental associations between bullying victimization and suicidal ideation and direct self-injurious behavior in adolescence and emerging adulthood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *63*(7), 820–828. https://doi.org/10.1111/jcpp.13529

Appendix 1: Overview of the papers included in the literature review

Title	First Author	Pub Year	Users	Chatbot's intention	To be used in schools	Used only in bullying context	Study type
Parental guardianship and online sexual grooming of teenagers: A honeypot experiment	Kamar, Eden	2022	Adults—General	Detection	No	No	Intervention
Robots, Bullies and Stories: A Remote Co-design Study with Children	Sanoubari, Elaheh	2021	Youths—General	Raising awareness	Yes	Yes	(Co-)design
LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors	Socatiyanurak, Vorada	2021	Youths—victims	Legal declaration	No	No	(Co-)design
Social Fringe Dwellers: Can chat-bots combat bullies to improve	Ireland, David	2018	Youths—victims	Intervention; Victims' support	No	Yes	(Co-)design

(continued)

Appendix 1. Continued.

Title	First Author	Pub Year	Users	Chatbot's intention	To be used in schools	Used only in bullying context	Study type
participation for children with autism?							
Using virtual role-play to enhance teacher candidates' skills in responding to bullying	Schussler, Deborah	2017	Adults—teachers	Education (adults)	Yes	Yes	Intervention
Therapist vibe: children's expressions of their emotions through storytelling with a chatbot	Santos, Kyle-Althea	2020	Youths—General	Victims' support	No	No	Design test
Interactive computing technology in anti-bullying education: The effects of conversation-bot's role on K-12 students' attitude change toward bullying problems	Young Oh, Eun	2020	Youths—General	Raising awareness	Yes	Yes	Intervention
Measuring effectiveness of chatbot to improve attitudes towards gender issues in underserved adolescent children in India	Agarwal, D.	2021	Youths—General	Raising awareness	No	No	Intervention
Design of a Chatbot for Four-to Ten-Year-Old Children Based on Emotional Intelligence	Rajwal	2023	Youths—General	Victims' support	No	No	(Co-)design
Privacy-Aware Digital Mediation Tools for Improving Adolescent Mental Well-being: Application to School Bullying	Gaci, Maria	2020	Youths—victims	Legal declaration; Victims' support	No	Yes	(Co-)design
Designing an Anti-Bullying Serious Game: Insights from Interviews with Teachers	Young, James	2022	Youths—victims	Raising awareness	Yes	Yes	(co-)design
Negobot: A conversational agent based on game theory for the detection of paedophile behaviour	Laorden, Carlos	2013	Adults—General	Detection	No	No	Design test
A chatbot framework for the children's Legal centre	Morgan, Jay	2018	Youths—victims	Legal declaration	No	No	(Co-)design
Co-developing a mental health and wellbeing chatbot with and for young people	Grové	2021	Youths—General	Victims' support	No	No	(Co-)design
"Don't hit me!"–combining language and action in FearNot!	Louchart, Sandy		Youths—General	Raising awareness	No	Yes	(Co-)design
" Hi, can I help?" An exploratory study of designing a chatbot to complement school nurses in supporting youths' mental health	Høiland	2019	Youths—General	Victims' support	Yes	No	Design test
A BDI-based Virtual Agent for Training Child Helpline Counsellors	Grundmann	2022	Adults—service providers	Education (adults)	No	No	Design test
Robot Framework for Anti-Bullying in Saudi Schools	Elgibreen, H.	2020	Youths—General	Raising awareness	Yes	Yes	Design test
A BDI dialogue agent for social support: Specification and evaluation method	van der Zwaan, J.M.	2012	Youths—victims	Victims' support	No	Yes	(Co-)design
CHI Against Bullying: Taking Stock of the Past and Envisioning the Future	Iivari, Netta	2021	Review	Review	No	Yes	Review
When the Social Becomes No-Human: Young People's Perception of Social Support in Chatbots	Bae Brandtzæg, Petter Bae	2021	Youths—General	Victims' support	No	No	Design test
Capturing Teens' Voice in Designing Supportive Agents	Lopatovska, Irene	2022	Youths—victims	Victims' support	No	No	(Co-)design
Cyberbullying Mitigation by a Proxy Persuasion of a Chat Member Hijacked by a Chatbot	Ueda, Tomoyuki	2021	Youths—victims	Intervention	No	Yes	Design test
A {Chatbot} {Solution} to {Chat} {App} {Problems}: {Envisioning} a {Chatbot} {Counseling} {System} for {Teenage} {Victims} of {Online} {Sexual} {Exploitation	Ahn, YunaZhang	2020	Youths—General; Youths—victims	Legal declaration; Raising awareness; Victims' support	No	No	(Co-)design
Simulating peer support for victims of cyberbullying	Van der Zwaan, J. M.Dignum	2010	Youths—victims	Victims' support	No	Yes	(Co-)design

Appendix 1. Continued.

Title	First Author	Pub Year	Users	Chatbot's intention	To be used in schools	Used only in bullying context	Study type
User {Validation} of an {Empathic} {Virtual} {Buddy} against {Cyberbullying	Wiederhold, B KRiva		Youths—victims	Victims' support	No	Yes	Design test
Chatbots to {Support} {Children} in {Coping} with {Online} {Threats}: {Socio}-technical {Requirements	Piccolo, Lara Schibelsky						
GodoyTroullinou		2021	Youths—victims	Victims' support	No	No	(Co-)design
Chatbot {Driven} {Web}-based {Platform} for {Online} {Safety} and {Sexual} {Exploitation} {Awareness} and {Reporting} in {Namibia	Rita, Marsela NurShava	2021	Youths—General	Raising awareness; Victims' support	No	No	Test de Design