

## Civil Society Chatbots: A Plurality of Conceptual Approaches

SOPHIE TOUPIN\*  
Université Laval, Canada

This article examines chatbots that have been developed by civil society for social, political, and/or cultural purposes. I ask what conceptual approaches have influenced the design of civil society chatbots and how these frameworks are actualized in practice. To do this, I first propose the concept of “civil society chatbots” to understand how and why civil society has chosen to use chatbots to support their campaigning and advocacy work. I then examine four conceptual approaches that have influenced them in the design of their chatbots and illuminate each with short case studies. The conceptual approaches under study are entertainment education, community media and technology, feminist design, and human and digital rights. Using these approaches helps situate a practice that is increasing and plural within civil society but relatively eclipsed by the hype around generative artificial intelligence chatbots.

*Keywords: chatbot, activism, advocacy, civil society, Global South*

In 2017, the Brazilian women-led laboratory for civic engagement and activism *Nossas Cidades* (Our Cities) designed a feminist chatbot for Messenger called BetaBot as part of its pan-Brazilian advocacy campaign (Toupin & Couture, 2020). Its aim was to support a countrywide feminist movement to challenge a bill restricting reproductive rights that would have made abortion illegal. BetaBot is a good example of how a nongovernmental organization (NGO) developed a feminist chatbot. BetaBot is not unique in its use of a conversational agent for activism, advocacy, or campaigning. It is part of a growing trend by NGOs, big and small, to experiment with human-computer interaction.

Past and current research on chatbots has divided them according to categories such as social chatbots (Ferrara, Varol, Davis, Menczer, & Flammini, 2016; Shum, He, & Li., 2018), political chatbots (Howard, Woolley, & Calo, 2018; Kim & Lee, 2023), or AI4good chatbots (Bhunu Shava, Rita, & Chitauro, 2022) among others. In this article, I contribute to the scholarship on chatbots by first proposing the category of civil society chatbots and second by focusing on the plurality of approaches that animate them. I ask what are civil society chatbots? Which conceptual frameworks have influenced their design? And how are these frameworks actualized in practice? To do this, I first explain what civil society chatbots are to understand how and why many NGOs have chosen to use chatbots to support their work. I then examine four conceptual frameworks that have influenced civil society in the design of chatbots and illuminate each of them with short case studies. The conceptual approaches under study are entertainment education (EE),

---

Sophie Toupin: [sophie.toupin@com.ulaval.ca](mailto:sophie.toupin@com.ulaval.ca)

Date submitted: 2023-06-27

Copyright © 2024 (Sophie Toupin). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

community media and technology, feminist design, and human and digital rights. Using these approaches helps situate an emerging technical practice and culture that is increasing and plural among NGOs but relatively eclipsed by the hype around generative artificial intelligence (AI) chatbots.

Methodologically, this article is based on various qualitative approaches to gain a better understanding of the design and use of civil society chatbots. First, it relies on a literature review of academic journals and civil society writings to understand how they frame and present their chatbots. Second, it is grounded in a non-exhaustive list I created of 48 civil society chatbots, using targeted Internet searches with keywords (such as #Activist Chatbot, #GenderBasedViolence Chatbot, #Feminist Chatbot, #HumanRights Chatbot, #Community Chatbot, # Antiracist Chatbot, #AIforGood Chatbot, and #Indigenous Chatbot), and supplemented with conversations with the actors that built them. In my semi-structured interviews, I asked interviewees to tell me about the NGO-led chatbots they were inspired by or the ones they knew in their country or region. The criteria used to add a civil society chatbot to the list were simple. It needed to have been developed by a not-for-profit group/organization or an individual such as an artist or activist; in the rare cases (a total of three) that it was developed by a social enterprise, it needed to have been designed for a social purpose, be free to use, and its host partnered with a local NGO.

I conducted semi-structured interviews with people who had designed or who were involved in developing chatbots for a social justice purpose. I conducted six interviews between 2020 and 2023 that lasted between 30 minutes and one hour. The themes addressed in the interviews included the intentions and reasons for choosing to design a chatbot, the design process, how the data collected were treated, and the funding received among others. All of the interviewees finalized were involved in the design process, all of them identified as women, transwomen, or nonbinary individuals (six of six), and the majority were in the Global South (five of six). The literature review of scholarly and NGO publications, the chatbot list, and the interviews together allowed me to identify growing trends in the development of civil society chatbots, the geography they emerge from, the purpose for which they are developed, and the conceptual approaches or influences they stem from. Third, this research is grounded in my own experience of developing two rule-based chatbots to better understand if—and if so, the extent to which—they can amplify social movement activism.

### **Setting the Terms and Understanding the Context**

Chatbots, also called virtual or conversational agents, automate text-based or multimedia-based conversations or interactions (Velkovska & Relieu, 2020). In this article, I focus specifically on civil society chatbots to understand the growing practice of the use and development of chatbots by NGOs. Building on Velkovska and Relieu's (2020) definition, I understand civil society chatbots as multimedia conversational agents of varying technical complexity developed by civil society actors to support social, political, or cultural causes. Civil society chatbot can be designed using the publicity feature that is offered by Facebook Messenger or WhatsApp. As an alternative to corporate ecosystems, they take the form of an open-source Web-based platform or a downloadable app. They are driven by algorithms of varying complexity and respond to users' messages by selecting the appropriate expression from preprogrammed schemas (rule-based chatbot) or adaptive machine learning algorithms (AI-powered chatbot), or a combination of both.

I use the term *civil society* to distinguish it from the state, corporate, and United Nations (UN) sectors. Andrew Calabrese (2004) has influenced my understanding of the term civil society as applied to chatbots, explaining that before the Velvet Revolution and the crumbling of the Union of Soviet Socialist Republics, Western intellectuals came to understand the term civil society as carrying connotations of resistance. Much of the optimism regarding civil society as an enabler of democracy and empowering the public sphere “was grounded in the possibilities of the means of communication to serve as lightning rods for the rapid flow of new and empowering ideas” (Calabrese, 2004, p. 321). Now, however, actors such as activists and researchers find the term civil society to have a less optimistic stance. In a context of economic stagnation and austerity measures, civil society must do more with less, has been pushed to align with development agency agendas, and has come to replace the state in some service delivery. In the British context, Nathalie Fenton (2023) has observed that the political agency of civil society has in fact diminished, reducing its capacity to act meaningfully to transform the world. Fenton (2023) has spoken of the depoliticization of civil society to explain the hollowing out of NGOs’ political agency in the context of proliferating digital tools. My use of the term *civil society chatbot* reflects this changing context.

Civil society chatbots have emerged due to several factors. First, the rapid uptake of conversational agents stems from the widespread use of smartphones and messaging apps in both the Global South and North. Corporate chatbots started taking center stage in the digital ecosystem when vocal assistants based on speech recognition technology were launched. Successive voice-powered conversational agents were launched with Siri in 2011 (Apple), Alexa in 2014 (Amazon), and Cortana in 2014 (Microsoft). To keep businesses, clients, and the general public within a commercial platform ecosystem, in 2016 Facebook launched its new chatbot design application for Messenger, thereby offering new types of customer relations services and publicity functions (Dale, 2016; Marcus, 2016).

The excitement about chatbots led corporations, governments, and UN agencies to experiment with them and, as the rhetoric goes, increase availability of access (a chatbot is available 24/7 as opposed to a human), improve efficiency of service by automating answers regularly asked of a human, and lower the cost of employing humans (in a neoliberal framework the human is costly). A chatbot-heavy digital ecosystem means that a growing number of people have come to interact with them, making them a viable and palatable option for NGOs to test and experiment.

The interest of civil society in developing chatbots needs to be understood as part of a larger historical, economic, social, and technological context where structural adjustment programs and privatization of services among others have all contributed to a paucity of access to universal basic services. This lack has led to the funding of chatbots as the “better than nothing option” to act as a surrogate for a social worker, psychiatrist, or health professional. This trend has been accelerated by the hype to invest in AI and the belief that machine learning can be leveraged to bring economic prosperity and other social, health, and educational benefits using chatbots.

The uptake of civil society chatbots is also linked to an increased critical technical practice and culture within NGOs. In the AI scholarship, Philip Agre (1998) refers to a critical technical practice as making sense of the close articulation between practices of technological development or tinkering with critical reflection. Agre’s (1998) articulation is helpful in understanding how NGOs’ emerging critical

chatbot-making practice represents a socio-technical terrain of political negotiation that is implicated in broader social and political struggles. The observed practice that we see emerging in civil society is part of a set of socially, economically, technologically, and historically situated relations that make visible the co-construction of technology and society (Vicente & Dias-Trindade, 2021). Civil society is an active agent in shaping how chatbots can support their work while at the same time, chatbots are shaping the type of campaigning and advocacy work civil society undertakes. This interactive process of critical chatbot-making practice embodies what Jasanoff and Kim (2015) have called socio-technical imaginaries, which means they are

collectively held and performed visions of desirable futures (or of resistance against the undesirable) and they are also animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology. (p. 19)

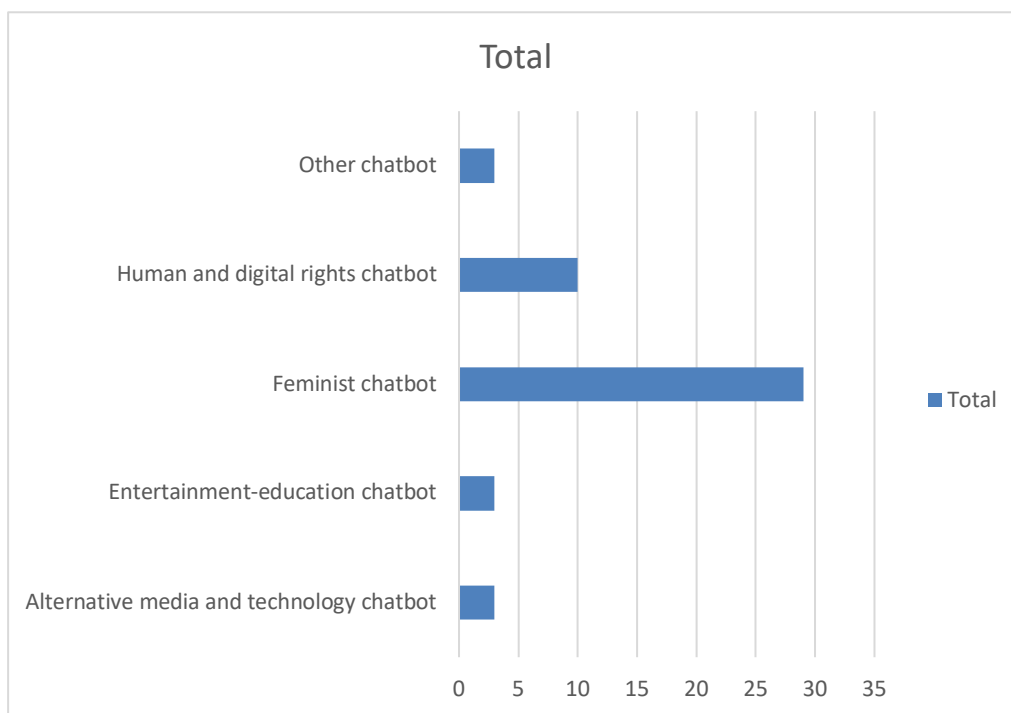
The authors go on to stress that they are “collective, durable, capable of being performed; yet they are temporally situated and culturally particular” (Jasanoff & Kim, 2015, p. 19).

#### **Four Conceptual Approaches**

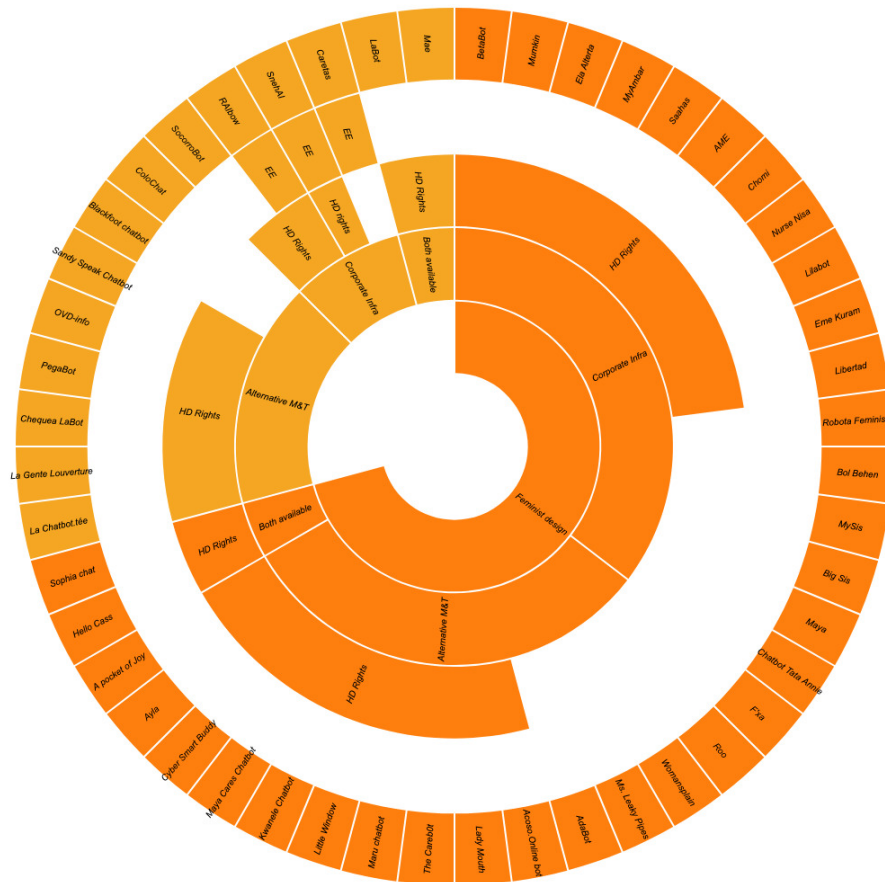
The development and use of chatbots by civil society for advocacy, campaigning, and activism is fairly new, but the critical traditions they are influenced by are not. Below, I present four conceptual approaches that have influenced the development of civil society chatbots: Entertainment education, community media and technology, feminist design, and human and digital rights. I categorized these four approaches based on the chatbot list I constituted, the online description of the surveyed chatbots, the interviews I conducted, and my own experience with developing chatbots (see Table 1 and Figure 1). This means that these approaches emerge from the ways in which the actors themselves present and understand what they did. While each chatbot has been categorized according to one dominant approach, it is important to stress that many of the chatbots under study often borrow from more than one approach. This means that the approaches are not necessarily mutually exclusive. Furthermore, a handful of the chatbots surveyed did not fit well in any of the identified approaches. Unfortunately, due to the lack of space and choices made for this article, I have not been able to further research the other approaches. Researching those limits and gaps will be important to add valuable conceptual and critical approaches to influencing the development of civil society chatbots.

**Table 1. Numbers of Civil Society Chatbots Associated With Their Dominant Conceptual Approach.**

Approaches	Numbers
Alternative media and technology	3
EE	3
Feminist design	29
Human and digital rights	10
Other	3
<b>Total</b>	<b>48</b>



**Figure 1. Numbers of civil society chatbots associated with their dominant conceptual approach.**



**Figure 2. A sunburst diagram representing the 48 chatbots under study and their associated approaches. The root of the sunburst is based on the feminist design approach since more than half of the chatbots surveyed borrow from that approach. Done with RawGraph 2 Beta.**

**EE Chatbots**

Entertainment education is the first approach I explore as a conceptual framework that has influenced the development of civil society chatbots. Drawing from classic theories of media for development, organizations have employed an EE approach to educate communities about social problems and motivate them to change their behavior. Arvind Singhal and Everett Rogers (2001), two proponents of the diffusion of innovation theory, have defined EE as “the process of purposely designing and implementing a media message to both entertain and educate, in order to increase audience members’ knowledge about an educational issue, create favorable attitudes, shift social norms, and change overt behavior” (p. 5). Thomas Tufte (2005), who studies communication and social change, has emphasized the changing nature of EE by defining it as

the use of entertainment as a communicative practice crafted to strategically communicate about development issues in a manner and with a purpose that can range from the more narrowly defined social marketing of individual behaviors to the liberating and citizen-driven articulation of social change agendas. (p. 162)

In the above definition, Tufte (2005) shows that EE has changed over time, drawing first from social marketing techniques that positioned the key challenge of social problems as a lack of information to a second generation that emphasized participation and a third that focuses on structural inequalities and power relations.

The concept and practice of EE has existed since at least the 1950s. One common story about the emergence of EE features Mexican broadcaster Miguel Sabido, who, in the 1970s, created soap opera TV shows (telenovelas in Spanish) to change the behavior of Mexican society with regard to specific health interventions and other social issues (Singhal, 2004). Although it is contested, the story goes that the innovation lens was then used and adapted in India, Kenya, and South Africa in the 1980s and 1990s, and it is still used today (Tufte, 2005).

At least two civil society chatbots stand out in this review that explicitly frame their conversational agent through the lens of EE. The first EE chatbot I want to highlight is called rAIInbow (2020). It was launched in November 2018 by the now folded social enterprise AI for Good UK in partnership with the intersectional feminist South African township-based not-for-profit Soul City Institute for Social Justice. After hearing stories of gender-based violence (GBV) from friends and acquaintances in South Africa, the founder of AI for Good UK, Kriti Sharma, decided to leverage smart AI technologies to address the endemic problem of GBV (Zisengwe, 2019). Soul City was a perfect partner as, in the mid 1990s, it had pioneered a participatory approach to EE (Singhal, 2004; Singhal & Rogers, 1999; Tufte, 2005). Started by two doctors at the dawn of South African democracy, Soul City recognized the need for training on basic health issues after 50 years of apartheid. Soul City's approach to EE was rooted in participatory research, including a standard literature review and focus group interviews with affected communities. This process allowed the organization to craft key messages in their EE programs. The TV and radio soap opera-style shows ran from 1994 until 2014 and addressed subjects such as maternal health, HIV/AIDS, and GBV, among others.

To design rAIInbow (2020), South City and AI for Good UK employed a participatory research methodology to develop credible fictional stories about survivors of GBV as content for the chatbot as reported in a Law Justice Development Zoom session (Law Justice Development, 2021). Soul City first ran discussion groups with hundreds of GBV survivors to understand what type of chatbot could best support their needs (Zisengwe, 2019). The discussions led to an understanding that the initial assumption about the chatbot's priority needed to change: From reporting GBV to helping identify that one was in an abusive relationship (Law Justice Development, 2021). The chatbot was therefore designed in such a way that it provided users with immersive and personalized EE stories with characters who had experienced GBV and encouraged behavior change (Zisengwe, 2019).

rAIInbow (2020) combined a rule-based module to provide button-based choices, with a retrieval-based module (or "AI") to respond to users' queries and personalize their experience. Facebook Messenger

was chosen at the time (circa 2017–2018) as it was a media that was widely available, accessible, and appropriate to the targeted audience. rAIInbow (2020) developers planned to go beyond the downloadable chatbot on Messenger, connect users to live chat with human beings, and expand the program's reach to North America, Africa, and Asia. rAIInbow (2020) is, however, no longer available on Messenger, and its description page has been hacked, showing text in Vietnamese.

The scholarship presents two understandings of EE civil society chatbots for fighting GBV. The first understanding is that, based on a comparison with online searches, the GBV chatbots seemingly perform better than a search engine when it comes to emotional support, information credibility, and representation of survivors (Falduti & Tessaris, 2022; Maeng & Lee, 2022). Second, scholars have suggested that raising awareness about unhealthy relationships through a gamified edutainment environment might have promising results (De Filippo et al., 2022). However, like Singhal and Rogers' (1999) study of EE telenovelas, De Filippo and colleagues (2022) concluded that while EE chatbots may raise awareness of GBV, they do not lead to behavior change. The potential of EE chatbots to fight against GBV needs to be tempered with what they can and cannot promise. In fact, they represent a low-cost digital technological intervention, while proper GBV interventions are time-intensive and costly to deliver (De Filippo et al., 2022).

The second EE-centered civil society chatbot I will discuss is SnehAI, developed when AI for Good UK was approached in 2019 by the charity Population Foundation of India (PFI) to support the chatbot's development. Again, SnehAI was designed, first, to change the behavior of young people in India regarding issues of sexual and reproductive health and, second, to focus on cyber harassment (Wang et al., 2022). As shown by the history of the use of EE by Soul City in South Africa, the PFI mobilized an EE approach to change behavior in its population through TV and radio soap opera programs for more than 50 years.

With SnehAI, the PFI attempted to extend their EE communication strategy through the interactive medium of the chatbot to reach their young, targeted audience. The chatbot is based on the persona of Dr. Sneha Mathur, a fictitious young woman doctor who is the lead protagonist of a digital EE TV show called *Main Kuch Bhi Kar Sakti Hoon* (I, A Woman, Can Achieve Anything; Population Foundation of India, 2014). In an interview, Poonam Muttreja, the executive director of the PFI said, "The SnehAI chatbot was built as an extension of our transmedia edutainment initiative" (Muttreja, 2020, para. 11). It was added in 2019 as a rule-based Messenger chatbot designed using the inputs of 84 teenagers and 19 adults (Wang et al., 2022). The PFI received more than \$700,000 to produce an updated version using natural language processing (NLP) from the UN Secretary-General fund End Violence Against Children and reach the Sustainable Development Goal 16.2 to end all forms of violence against children by 2030 (End Violence Against Children, n.d.).

The EE approach comes to life in two ways with SnehAI. First, an interaction with the chatbot offers exposure to different sets of conversations. Second, through short video animation, it provides young people in India with sexual and reproductive health conversations and messaging to stay safe from harmful online behavior such as cyber harassment, bullying, trolling, and online grooming, among others. The belief behind SnehAI is that young people interacting with the chatbot will learn socially desirable behaviors from the characters presented. This approach is in sync with Singhal and Rogers (1999), who explain that viewers of EE programs learn from how characters are rewarded for good behaviors and punished for undesirable ones.



However, while an EE communication strategy may increase awareness about a message, it is less successful in influencing attitudes and behavioral change (Singhal & Rogers, 1999).

One digital health-care study of SnehAI has shown that the chatbot has mostly been used by young men (93%; Wang et al., 2022). Their study reveals shocking gender-based violent questions that young men have asked the chatbot. Wang et al.'s (2022) research not only confirms previous studies that chatbots that have a female presenting avatar are more abused than those with male-appearing avatars (Feine, Gnewuch, Morana, & Maedche, 2019), but also raises pressing questions. Specifically, Wang et al. (2022) opens a set of urgent considerations regarding how not to transform NLP-powered civil society chatbots into toxic and misogynist machines. How can its developers guarantee that the AI will not model the misogyny of its previous interactions? More studies are necessary to understand how and to which extent EE approaches will be changed by AI.

### ***Civil Society Chatbots as Community Media and Technology***

Understanding civil society chatbots through this lens reignites a tradition of alternative communication made by and for social movements. Alternative media and technology are well exemplified by the creation of Independent Media Centers (IMCs) at the end of the 20th century. Influenced by the international solidarity with the Zapatista movement, community media and tech activists joined forces in 1999 to create an IMC to counter corporate journalism in content, form, and purpose. Indymedia was launched in Seattle during the protests against the World Trade Organization Ministerial Conference of 1999 to present an inside view of street protests. It then spread to many locations igniting an Indymedia movement. The motto of the communication media and technology approach was "be the media," which consisted of the belief that activists needed their own means of communication outside of corporate media, which distorted their views and perspectives (Kidd, 2003).

Indymedia was unique as it set up a blog based on an innovative open publishing software, which allowed activists to participate in news production, thereby bypassing traditional gatekeepers such as journalists and webmasters (Milan, 2010). The hundreds of IMCs that were created around the world included as a feature this open publishing process, which was grounded in free and open-source software (FOSS) and guidelines on publishing that largely refused to censor most information (Lievrouw, 2011). Tech activists coming out of the FOSS tradition developed the open publishing software along with other tools. The collaboration between media and tech activists stressed the importance of alternative content as much as it did the alternative digital infrastructure.

Two chatbots that accurately represent this framework were designed by Lab Delta. The founders of Lab Delta, one of whom is the author of this article, have been involved in setting up or collaborating with IMCs in Canada and beyond. The first rule-based chatbot aims to increase the awareness of activists, community-based organizations, and students on alternative technologies during the pandemic. On its Web-based platform it says, "The *Chatbot.tée* uses text to (re)activate social combat and digital resistance by proposing an alternative technology" (Chatbot.tée, 2021, para. 2). As lockdowns and other COVID-19 related restrictions moved many of our social, economic, cultural, and political lives online, people including activists started massively adopting corporate digital technologies. It felt imperative to inform the public

and, particularly, activists in Quebec about alternatives to corporate digital infrastructures. This was the basis for launching the chatbot.

From its inception, the idea of *Chatbot.tée* was to offer a free open-source plug-in template for civil society and activists to use to experiment with chatbots. It is licensed according to Creative Commons to allow for such a legal hack. Koumbit, a tech collective based in Montreal, developed the WordPress plug-in. The connections with the Indymedia movement extend to Koumbit, which emerged from *Centre des médias alternatifs du Québec* (Independent Media Quebec). Thanks to funding from the Canadian Internet Registration Agency, Koumbit was paid \$5,000 to develop the chatbot template.

The chatbot's political commitments are stated clearly: It refuses the data and algorithmic capitalism that favor the oligopolistic power of tech giants, and it counters the surveillance regime and all technologies that reproduce unequal structures and practices that undermine the rights and dignity of women, Black, Indigenous, and racialized people, disadvantaged population groups, and migrants and refugees. *Chatbot.tée* brings together a tradition of community media and tech activism that supports an understanding of why developing alternatives to corporate platforms pushes against current state and corporate data extraction and surveillance ecosystems.

The second rule-based chatbot developed by Lab-Delta is called *La Gent L'ouverture*. It emerged from the author's own activism with undocumented people, experimenting with *Chatbot.tée*, and as a way to show how the chatbot template could be used by civil society. *La Gent L'ouverture* was designed as part of the 2022 pan-Canadian campaign to regularize the status of undocumented people. "Its name echoes Toussaint Louverture, the most prominent leader of the Haitian revolution, and gestures towards the need to open borders" (*La Gent L'ouverture*, 2022, para. 1). More broadly, the use of the term *ouverture* (openness) symbolizes the lifting of borders and the freedom for everyone to be mobile, not just those who have the "right" passport. It stresses the importance of using technology for a social justice purpose: "Instead of using artificial intelligence and other emerging technologies to reinforce border violence under the guise of efficiency and cost rationality, *La Gent L'ouverture* uses technology to campaign for status for all" (*La Gent L'ouverture*, 2022, para. 2).

Using audios, images, texts, and videos, *La Gent L'ouverture* gives an account of four migration stories describing the forms of violence, systemic racism, and sexism migrants have experienced while making visible the agency and sense of empowerment that has come from being engaged in a collective struggle for the regularization of these migrants' status. The project had a total funding of \$1,000, the entirety of which went into providing payment for undocumented migrants' time to tell their stories and for visual design.

Activist chatbots, a subset of civil society chatbots, are designed with the belief in a collective future of open culture oriented toward social justice in content, form, and purpose. As Figure 1 shows, several civil chatbots borrow from the media and technology approach in their use of open-source technology. However, not all civil society chatbots using open-source technology constitute small-scale political interventions that reenact the tradition of media and tech activism to disrupt the status quo, confront conservative politics, and create new imaginaries. Activist chatbots are also embedded in a framework of refusal that says no to

data collection as a political orientation (Cifor et al., 2019; Liboiron, 2021). Forms of data refusal imply a desire for autonomy, self-determination, and sovereignty—all principles that are constitutive of a tradition of alternative media and technology. From a refusal-and-repair standpoint, it should be a right by design to decide not to calculate metrics and to keep the autonomy of people who want to interact with a chatbot free from surveillance, data collection, calculation, and metrics.

### ***Feminist Chatbots as a Feminist Design Practice***

To fight and organize against GBV, including technology-facilitated GBV, feminist organizations have started developing feminist chatbots as part of their feminist toolbox (Toupin & Couture, 2020). They range from simple button-based WhatsApp chatbots to finding key resources, to creating a safe and resourceful space on a feminist platform, to chatbots that allow for organizing around feminist issues. As Table 1, Figure 1, and Figure 2 show, most of the civil society chatbots surveyed for this article can be categorized as feminist chatbots.

To understand this category, I draw from Shaowen Bardzell (2010). In her article on feminist human-computer interaction, Bardzell (2010) argues that feminism is a natural ally to interaction design. Her argument stems from the six qualities she identifies that make up feminist design: Pluralism (there is no one design model to follow), participation (participatory process has value), advocacy (its goal is emancipation), ecology (how design impacts society), embodiment (focusing on the agency of the interaction), and self-disclosure (establishes user awareness about what the software is doing). I draw from Bardzell's (2010) work because a few feminist chatbots under study specifically mention her work as an inspiration. The examples I showcase below show how feminist chatbots as a feminist design practice create new, personalized, and interactive feminist digital platforms that can be built on top of a corporate digital infrastructure (Messenger or WhatsApp) or embrace the practice of open culture with their choice of design.

The first feminist chatbots surveyed for this section and mentioned earlier in this article was BetaBot. Launched in 2017, BetaBot used Facebook Messenger's new automated personalized advertisement feature, framing itself as the following: "Facebook's first feminist bot was designed to track women's rights bills and warn when there was a whiff of backlash in the air" (BetaBot, 2017, para. 5). It was designed to provide feminists and LGBTQI+ individuals with a tool for organizing around reproductive justice. BetaBot was a creative way to fight back against a law that threatened to criminalize abortion, by providing information about the bill and how to fight back. It automated sending e-mails to Brazilian members of parliament (MPs) regarding users' disagreement with the proposed bill (Calado, 2018). Its creators chose a rule-based algorithm as opposed to using machine learning because of the shoestring budget they had for the project, to have more control over the human-machine conversation, and to filter the hate that Beta received (Toupin & Couture, 2020). Using machine learning was deemed unnecessary because of the very small size of the data set, and BetaBot was launched a year after Microsoft Twitter Tay's debacle. A rule-based algorithm enabled the designers to have control over the conversation through a preprogrammed script. "Beta is feminist; she needed to say she is not in agreement with racism, sexism or any kind of oppression, she is pro-abortion, and she was not deceptive with regards to being a robot" (Anonymous, personal communication, January 3, 2020). By using a preprogrammed script, Beta did not

make visible the hate that it received nor train its algorithms to spit out violent rhetoric. Only the programmers saw the violent messages Beta received while keeping track of the chatbot's use.

Beta is an example of a feminist design practice that was created to amplify the activism raging on the streets of Brazil in 2017. In fact, it contributed to the thinking about feminist design as a practice by clarifying that the automation of certain time-consuming tasks (such as sending e-mails to MPs) could support the feminist movement. While a Web-based platform still exists to retain the memory of Beta, its Messenger app no longer replies to messages. BetaBot retired in 2020 after three years of activities and five campaigns to protect women and LGBTQI+ rights.

The second feminist chatbot I will discuss, called Maru, was developed by the London-based NGO Feminist Internet together with the international development organization Plan International (PI). It was based on an experiment with designing a feminist version of Alexa, F'xa, in 2019 to serve as a feminist guide to AI bias. It was rooted in Bardzell's (2010) feminist human-interaction principles and in Josie Swords' (2017) feminist chatbot design process, which stress the importance of being self-reflective about the composition of the developers' team, the purpose for which the chatbot is developed, how the data are treated, how the chatbot is depicted, and the disclosure as to how the chatbot functions.

Maru was the result of research on online harassment by PI that revealed that a majority of girls (58%) felt unsafe online, particularly on Facebook (39%; Goulds, Gauer, Corr, & Gallinetti, 2020). Maru was designed as part of the #FreeToBeOnline campaign to support girls who face online harassment on Instagram, Facebook, TikTok, and Twitter. Responding to feminist criticisms of chatbots (Feine et al., 2019; Neff & Nagy, 2016), Maru was designed using a set of feminist design principles including codesigning with young people mostly from countries in the Global South; it avoided reinforcing visual stereotypes and used inclusive and empathetic language (Feminist Internet, 2020). It was designed outside of a social media corporate ecosystem where harassment of girls happens, and it uses an open-source Web-based platform format to respect the privacy of its users by not collecting any data from the chatbot interactions. Its commitment to making its references visible is yet another small feminist design intervention built into the conversation interface. This feminist practice of citing sources (Liboiron, 2021) along with trigger warning symbols is made easy through basic hyperlinks embedded in the chatbot conversation. In comparison with the lack of references—or worse, the made-up references—in a generative AI tool such as ChatGPT, this is a simple and trustworthy feminist practice. Maru provides information about online harassment, how to support others who are experiencing such violence, how to respond to or report online violence, and how to protect oneself digitally. The interactive aspect of Maru presents information in a less static and more dynamic fashion than a regular website, mimicking conversations girls can have on social media. As a feminist design intervention, Maru innovates in its feminist referencing practice to avoid disinformation and develop trust about using the tool.

The third example I use is *Mumkin* (meaning "possible" in Urdu), a feminist chatbot initiated by its chief executive officer Priya Goswami. *Mumkin* was designed to support those in the Borah community in both India and the diaspora to prepare how to have a difficult conversation with family and friends on issues of female genital mutilation (FGM). She described the chatbot idea as a "feminist tech intervention" (Priya Goswami, personal communication, May 3, 2023) that emerged from a conversation between two women survivors of FGM who were tired of arguing with their mothers as well as the need to prepare their arguments. Goswami

described *Mumkin* as “simulated in-app mock conversation” that provides mental wellness for survivors of female genital cutting (personal communication, May 3, 2023). Two major questions she was trying to answer in thinking about designing *Mumkin* were (a) how it was different than using Instagram or Facebook and (b) how users could not be monitored by harassers. As she told me in an interview, she had developed cynicism about corporate social media platforms, because FGM activists have long been the target of harassment for speaking out, and she wanted an alternative place to discuss in peace and safety. *Mumkin* was inspired by corporate digital wellness chatbots such as Headspace Health and Wysa and adapted to its feminist purpose. In speaking about her feminist project and its utility Goswami was very candid:

Why would someone who has been frustrated with their mother’s response for years want to have a conversation with a phone? Why would someone want to sit through a 10- to 15-minute conversation which is very visceral and triggering? Despite these questions, we still believed the project had values. (Priya Goswami, personal communication, May 3, 2023)

*Mumkin* as a chatbot was developed following Goswami’s successful funding application to the Grand Challenge Canada. With the CAD 100,000 prize she paid herself and her team members to work on the design and the visuals, took time to craft well-thought-out conversations through text and audio, and hired a team of developers in Bengaluru. On its website, *Mumkin* (2019) presents itself as an AI-powered chatbot, but the retrieval-based module is not yet available. Since the project is currently on hold, its conversation services are offered only through a rule-based function.

Finally, the last feminist chatbot I highlight is CAreB0t. It was developed by Caroline Sindors and the UK-based design studio Comuzi. It was designed to care for survivors of online harassment while offering best practices about open-source chatbots. CAreB0t is a feminist artistic intervention and political commentary on the violence many equity-deserving groups are subjected to within the social media ecosystem and in a context where insurance for mental issues in the United States and many other countries is unaffordable. In fact, paid subscriptions to wellness chatbots have become a booming digital health industry. In this context, CAreB0t is a commentary on where automating care will take us. As already alluded to, the World Bank and International Monetary Fund schemes for decades have forced countries in the Global South to reduce their spending on health, creating a large market for American-based companies to sell to customers all over the world digital affordable mental health for everyone.

CAreB0t is a small-scale artistic project that does not pretend to have the answers to or solutions for complex historical, social, cultural, and economic issues. Its intent is not to replace the need for a therapist. CAreB0t is politically significant as a feminist political commentary on the state of corporate social media, the lack of access to universal mental health-care systems, the need for civil society to be involved in technological and design choices, and the ability to say no to corporate chatbots that discursively pretend to provide care. The fight is for good universal health care, not for good chatbot health care.

### ***Human and Digital Rights–Based Approach***

The last approach I focus on highlights how civil society chatbots use a rights-based approach including both human and digital rights. I encountered this framing in the description of chatbots that focus on countering disinformation online (such as LaBot, PegaBot) and online violence or harassment (such as

Maru, Little Window, Kuram) and when conducting semi-structured interviews with a handful of designers of civil society chatbots, namely *Socorrobot* (WhatsApp) and OVD-info (Telegram). The UN Sustainable Development Group (UNSDG, n.d.) describes a human rights-based framework as a “conceptual framework for the process of human development that is normatively based on international human rights standards and operationally directed to promoting and protecting human rights” (para. 2). In the last decade, scholars and NGOs have added a digital rights dimension to broaden the human rights-based approach. In their review of what digital rights are Pangrazio and Sefton-Green (2021) highlight that these rights may refer to actions in an exclusively digital field such as freedom of expression online and access to the Internet or how the digital modifies, expands, or extends preexisting rights.

Two chatbots exemplify this framework well. OVD-info is a chatbot based on the open-source software Telegram that focuses on political persecution in Russia. It allows Russians to report an arrest or any other interaction with law enforcement.

The chatbot enables users to receive recommendations on what to do if they are arrested, and what legal texts to quote in front of the police. They can send a message directly to our team in case of need as we are operating outside of Russia. (Anonymous, personal communication, June 15, 2023)

OVD-Info is an independent human rights defense and media group that focuses on rights to freedom of assembly and expression. The group developed the Telegram chatbot to help them circulate vital rights information to protesters. In the following quote they explain how the project came about.

In 2017 there were thousands of people arrested not only in Moscow but in the whole of Russia. They took to the streets to demonstrate after they watched Navalny’s documentary Don’t Call him Dimon. If you have so many detainees, you have a lot of calls. Our hotline was overloaded. We understood that we needed to automate part of the process. The decision to develop a chatbot was a way to solve many problems. (Anonymous, personal communication, June 15, 2023)

Another poignant example is the case of *Socorrobot* (helpbot in Spanish). *Socorrobot* was launched in October 2022 as a rule-based chatbot on WhatsApp to support the families of the disappeared in Mexico. It was designed on a voluntary basis by three Mexican digital rights NGOs: Codeando Mexico, Quinto Elemento Lab, and Technicas Rudas. Before working on the chatbot together, the three organizations were all working on issues related to disappearances in Mexico.

*Socorrobot* was developed in an inherently political context. In Mexico, disappearances have been a major human rights problem since at least the 1960s (Mandolessi & Olalde Rico, 2022). One of my interviewees said,

Disappearance is a massive human rights problem in Mexico, and there are not enough people to support all the collectives composed of the families of the disappeared. We knew there was a need to automatize the process of how a family can report a disappeared person. (Anonymous, personal communication, January 31, 2023)

The idea to develop a chatbot came from the desire to use automation for a human rights issue.

Reporting a disappeared person is an institutional process you need to follow. There are not enough people who know which process to follow to file a missing person and those who do, don't have the time and resources to support all these cases. (Anonymous, personal communication, January 31, 2023)

Together, the three organizations that developed *Socorrobot* decided to meet people where they were rather than design an open-source Web-interface chatbot (Anonymous, personal communication, January 31, 2023). They decided to test the WhatsApp application programming interface chatbot as it is widely used by Mexicans, young and old, and many telecommunication companies give people unlimited access as part of their service. Another important factor that convinced the three organizations to develop a WhatsApp chatbot was the fact that it was not new in people's practice. The designers of *Socorrobot* speculate that because the Mexican government had launched a COVID-19 WhatsApp chatbot, Mexicans had already been exposed to WhatsApp chatbots during the pandemic and were more inclined to use and trust them.

*Socorrobot* was developed over a year and a half. It started with collective discussions among journalists, civic hackers, and social justice activists who were interested in the project. As part of their research and design process, they first interviewed many actors on how to file a missing person's report. These interviews included discussions with the *Comisión Nacional de Búsqueda de Personas* (National Search Commission), lawyers, and family members of the disappeared. Additionally, they submitted access-to-information requests. Once the research process was done, the three organizations met over a weekend to map the flow of the conversation. Then, they conducted several test runs to validate the conversation flow and the tone of the conversation with the families of the disappeared, from a range of genders and ages, and integrated their comments. To prioritize user safety and prevent the identification of victims or their family members, the team decided to collect as little data as possible.

When initially presented with the tool, the families of the disappeared said that they would never use a robot (Anonymous, personal communication, January 31, 2023). But when they saw what the chatbot could do, they took to it immediately and circulated it quickly among the families of the disappeared. In fact, when asked how they publicized the chatbot, my interviewee said,

Not that much. We wanted it to be organic rather than put together a big social media campaign. What we did is sent one press release and organized a small workshop with about 15 family members in Oaxaca, and the next day a second workshop in Mexico City on August 30, the international day of the disappeared. On the second day, when we arrived and presented the chatbot in Mexico City, somebody came to see me and said, "Hey yesterday I got this message about Socorrobot." (Anonymous, personal communication, January 31, 2023)

The one press release they sent led to countrywide media coverage. The Canadian Embassy in Mexico gave the three organizations CAD 35,000 to come up with an updated version of the chatbot. With

this funding, they plan on updating the accuracy of the chatbot—especially regarding migrants from South America who disappear during their journey in Mexico. They are also updating the chatbot to better answer the needs of the families that want to report people who disappeared in the 1960s or 1980s.

*Socorrobot* has led many groups of the families of the disappeared to use the chatbot beyond its initial purpose. The chatbot has become much more than a tool for reporting—it has become a tool for prevention, raising awareness, and breaking stigma. In fact, *Socorrobot* has been included as part of the strategy of families of the disappeared to talk about this very sensitive issue together with the wider community (Anonymous, personal communication, January 31, 2023). The families of the disappeared have appropriated the tool to best support their work and in turn expanded its function in unexpected ways.

### Conclusion

This article has focused on the growing trend among NGOs to experiment with chatbots and the plurality of conceptual approaches that have inspired them. What we come to understand from this growing practice is that the tinkering of civil society with chatbots demonstrates the possibility of thinking about and doing chatbots differently. This emerging practice is not only made possible thanks to a growing technical culture within NGOs but also because they are animated by normative ideas that they are able to perform technically. NGOs have not blindly adopted chatbot templates developed by Big Tech. Rather, they have been active agents in rethinking how and what chatbots can do for their advocacy campaigns. They have tried to align their design to further a world that is discursively and materially oriented toward social justice. Experimenting with chatbots has allowed them to use creatively new human-computer tools and in turn broaden their repertoire of activism and campaigning. The plurality of conceptual approaches presented shows varying and often intertwined socio-technical imaginaries that coalesce around social justice issues.

The practice of developing civil society chatbots constitutes a counterpoint to the development of corporate-led generative AI and a way to (re)imagine digital alternative futures centered around community building, inclusivity, and social justice. What we come to understand with this trend is that civil society has developed a critical chatbot-making practice and culture that draw from multiple sources of inspiration. While the new publicity feature available on social media has influenced them to experiment with chatbots in the first place, they have also come to influence this medium of communication and create a vibrant critical technical practice. The critical chatbot-making practice that has emerged as a result represents innovative ways of digital storytelling to collectively organize around social justice issues, give services, automate tasks, and bear witness to how to better protect and defend human dignity, among others. The appropriation of chatbots by civil society has also allowed for the creation of narratives and practices of digital resistance in the face of widespread data extraction, digital surveillance, and other forms of digital violence on social media.

Civil society chatbots are a distinct and politically inclined practice that deserves our attention. More experiments are needed by activists and NGOs to find ways to further this critical making practice including how to organize securely through chatbots. Moreover, further empirical research is needed to make visible the wealth of approaches that inspire civil society chatbots.



### References

- Agre, P. E. (1998). Toward a critical technical practice: Lessons learned in trying to reform AI. In G. Bowker, S. L. Star, L. Gasser, & W. Turner (Eds.), *Social science, technical systems, and cooperative work* (1st ed., pp. 131–157). London, UK: Routledge.
- Bardzell, S. (2010). Feminist HCI: Taking stock and outlining an agenda for design. In *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1301–1310). New York, NY: ACM. doi:10.1145/1753326.1753521
- BetaBot. (2017). *Our technologies beta*. Retrieved from <https://www.nossas.org/en/technologies>
- Bhunu Shava, F., Rita, M. N., & Chitauro, M. (2022). *Tech4Good: Artificial intelligence powered chatbots with child online protection in mind* (SSRN Scholarly Paper 4331466). doi:10.2139/ssrn.4331466
- Calabrese, A. (2004). The promise of civil society: A global movement for communication rights. *Continuum, 18*(3), 317–329. doi:10.1080/1030431042000256081
- Calado, C. (2018, February 25). *Beta Feminista—Uma entrevista com o time responsável pela criação do bot feminista* [Beta Feminista—An interview with the team responsible for creating the feminist bot]. Medium. Retrieved from <https://medium.com/botsbrasil/beta-feminista-uma-entrevista-com-o-time-respons%C3%A1vel-pela-cria%C3%A7%C3%A3o-do-bot-feminista-bba17-6c3fa285e41>
- Chatbot.tée. (2021). *À propos de la Chatbot-tée* [About the Chatbot-tée]. Retrieved from <https://www.labdelt.ca/cest-qui-la-chatbot-tee/>
- Cifor, M., Garcia, P., Cowan, T., Rault, J., Sutherland, T., Say Chan, A., . . . Nakamura, L. (2019). *Feminist data manifest-No*. Retrieved from <https://www.manifestno.com/>
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering, 22*(5), 811–817. doi:10.1017/S1351324916000243
- De Filippo, A., Bellatin, P., Tietz, N., Grant, E., Whitefield, A., Nkopane, P., . . . Hatcher, A. (2022). *Effects of digital chatbot on gender beliefs and exposure to intimate partner violence among young women in South Africa*. Retrieved from <https://osf.io/preprints/psyarxiv/n6rdj/>
- End Violence Against Children. (n.d.). *Population foundation of India*. Retrieved from <https://www.end-violence.org/grants/population-foundation-india>

- Falduti, M., & Tessaris, S. (2022). On the use of chatbots to report non-consensual intimate images abuses: The legal expert perspective. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good (GoodIT '22)* (pp. 96–102). New York, NY: ACM. doi:10.1145/3524458.3547247
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). Gender bias in chatbot design. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, NL: ACM, Revised Selected Papers* (pp. 79–93). doi:10.1007/978-3-030-39540-7\_6
- Feminist Internet. (2020, November 25). *How we created global anti-harassment chatbot, Maru*. Medium. Retrieved from <https://feministinternet.medium.com/how-we-created-global-anti-harassment-chatbot-maru-c74a36039537>
- Fenton, N. (2023). Understanding civic participation and realizing data justice. *International Journal of Communication*, 17, 3660–3678.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. doi:10.1145/2818717
- Goulds, S., Gauer, M., Corr, A., & Gallinetti, J. (2020). *Free to be online? Girls and young women's experiences of online harassment (Research report No. 3, pp. 1–56)*. Plan International. Retrieved from <https://plan-international.org/uploads/2022/02/sotwgr2020-commsreport-en-2.pdf>
- Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the U.S. 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93. doi:10.1080/19331681.2018.1448735
- Jasanoff, S., & Kim, S.-H. (2015). *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. Chicago, IL: University of Chicago Press.
- Kidd, D. (2003). Indymedia.org: A new communications commons. In M. McCaughey & M. Ayers (Eds.), *Cyberactivism: Online activism in theory and practice* (1st ed., pp. 46–69). London, UK: Routledge.
- Kim, Y., & Lee, H. (2023). Humanizing chatbots for political campaigns: How do voters respond to feasibility and desirability appeals from political chatbots? *Interacting With Computers*, 35(1), 40–48. doi:10.1093/iwc/iwad024
- La Gent L'ouverture. (2022). Qui est La Gent L'ouverture [Who is La Gent L'ouverture]. Retrieved from <https://www.labdelta.ca/qui-est-lagent-louverture/>

- Law Justice Development. (2021). *The rise of chatbots against gender-based violence event report*. Retrieved from <https://jitij.org/2022/10/05/the-rise-of-chatbots-against-gender-based-violence-event-report/>
- Liboiron, M. (2021). *Pollution is colonialism*. Durham, NC: Duke University Press.  
doi:10.1215/9781478021445
- Lievrouw, L. A. (2011). *Alternative and activist new media*. Malden, MA: Polity Press.
- Maeng, W., & Lee, J. (2022). Designing and evaluating a chatbot for survivors of image-based sexual abuse. In *CHI '22: Proceedings of the 2022 Chi Conference on Human Factors in Computing Systems* (pp. 1–21). New York, NY: ACM. doi:10.1145/3491102.3517629
- Mandolessi, S., & Olalde Rico, K. (2022). *Disappearances in Mexico: From the "dirty war" to the "war on drugs."* London, UK: Routledge.
- Marcus, D. (2016, April 12). *Messenger platform at F8*. Meta. Retrieved from <https://about.fb.com/news/2016/04/messenger-platform-at-f8/>
- Milan, S. (2010). The way is the goal. Interview with Maqui, Indymedia London/IMC-UK Network Activist. *International Journal of E-Politics*, 1(1), 88–91. doi:10.4018/jep.2010102206
- Mumkin. (2019). *Mumkin*. Retrieved from <https://www.mumkinapp.com/>
- Muttreja, P. (2020, December 3). *Meet a digital groundbreaker: Population Foundation India*. End Violence Against Children. Retrieved from <https://www.end-violence.org/articles/meet-digital-groundbreaker-population-foundation-india>
- Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 10, 4915–4931.
- Pangrazio, L., & Sefton-Green, J. (2021). Digital rights, digital citizenship and digital literacy: What's the difference? *Journal of New Approaches in Educational Research*, 10(1), 15–27.  
doi:10.7821/naer.2021.1.616
- Population Foundation India. (Producer). (2014). *Main Kuch Bhi Kar Sakti Hoon* [I, a woman, can achieve anything] [Web TV series]. New Delhi, India: YouTube. Retrieved from <https://www.youtube.com/@MainKuchBhiKarSaktiHoon>
- rAInbow. (2020). *Hi rAInbow*. Retrieved from <https://www.hirainbow.org/>
- Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10–26.  
doi:10.1631/FITEE.1700826

- Singhal, A. (2004). *Entertainment-education and social change: History, research, and practice*. New York, NY: Lawrence Erlbaum Associates.
- Singhal, A., & Rogers, E. M. (2001). The entertainment-education strategy in campaigns. In R. E. Rice & C. Atkins (Eds.), *Public communication campaigns* (3rd ed., pp. 343–356). Thousand Oaks, CA: SAGE Publications.
- Singhal, A., & Rogers, E. M. (1999). *Entertainment-education: A communication strategy for social change*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Swords, J. (2017, September). *Feminist chatbot design process (FCDP)*. Retrieved from <https://drive.google.com/file/d/0B036SIUSi-z4UkkzYUVGTGdocXc/view?resourcekey=0-DS-Lj4uCh2VHf1cuogGNfg>
- Toupin, S., & Couture, S. (2020). Feminist chatbots as part of the feminist toolbox. *Feminist Media Studies*, 20(5), 737–740. doi:10.1080/14680777.2020.1783802
- Tufte, T. (2005). Entertainment-education in development communication: Between marketing behaviours and empowering people. In T. Tufte & O. Hemer (Eds.), *Media and glocal change—Rethinking communication for development* (pp. 159–176). Buenos Aires, Argentina: Publicaciones Cooperativas.
- United Nations Sustainable Development Group. (n.d.). *Human rights-based approach*. Retrieved from <https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach>
- Velkovska, J., & Relieu, M. (2020). Pourquoi ethnographier les interactions avec les agents conversationnels? [Why should we ethnograph interactions with conversational agents?]. *Réseaux*, 220–221(2–3), 9–20. doi:10.3917/res.220.0009
- Vicente, P. N., & Dias-Trindade, S. (2021). Reframing sociotechnical imaginaries: The case of the Fourth Industrial Revolution. *Public Understanding of Science*, 30(6), 708–723. doi:10.1177/09636625211013513
- Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (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), 1–18. doi:10.2196/29969
- Zisengwe, M. T. (2019, July 10). *Innovation Q&A: Kriti Sharma, founder of rAInbow*. Civic Tech Innovation Network. Retrieved from <https://medium.com/civictech/innovation-q-a-kriti-sharma-founder-of-rainbow-bcf84e5d321e>