

# A 'design justice' approach to developing digital tools for addressing gender-based violence: exploring the possibilities and limits of feminist chatbots

Nicola Henry <sup>a</sup>, Alice Witt <sup>a,b</sup> and Stefani Vasil <sup>c</sup>

<sup>a</sup>Social Equity Research Centre, RMIT University, Melbourne, Australia; <sup>b</sup>Deakin Law School, Deakin University, Melbourne, Australia; <sup>c</sup>Australian Catholic University, Sydney, Australia

## ABSTRACT

Digital tools, such as safety apps, reporting portals, and chatbots, are increasingly being used by victim-survivors of gender-based violence to report unlawful activity and access specialized support and information. Despite their limitations, these interventions offer a range of potential benefits, such as enhancing decisional certainty, promoting safety behaviors, and fostering positive psychological outcomes. In this paper, we introduce an innovative 'design justice' approach to the development of digital tools for addressing gender-based violence. Drawing on our experience of building a feminist chatbot focused on image-based sexual abuse, we argue that the integration of feminist principles throughout the design, content, and evaluation stages is crucial for mitigating the risk of harm and promoting positive outcomes. Our theory-informed and practice-led approach can help to guide the development of other digital tools for addressing gender-based violence. Nonetheless, more scholarly research is needed to investigate the use, efficacy, and impacts of such interventions, at the core of which should be interdisciplinary collaboration between subject matter experts, victim-survivors, technical specialists, and other key stakeholders.

## ARTICLE HISTORY

Received 13 July 2023  
Accepted 14 May 2024

## KEYWORDS

Design justice; digital tools; gender-based violence; feminist chatbots; image-based sexual abuse

## Introduction

Victim-survivors of gender-based violence (GBV) experience multiple barriers to reporting their experiences, as well as seeking help, information, and resources after the abuse (Munro-Kramer et al., 2017).<sup>1</sup> Research shows that low rates of reporting or disclosure are commonly the result of victim-blaming attitudes and rape myths, language barriers, geographical isolation, discrimination, stigma, shame, and guilt (e.g., Campbell et al., 2001; Ullman & Filipas, 2001). These and other barriers can lead victim-survivors to downplay the seriousness of their experiences and avoid seeking help (Munro-Kramer et al., 2017). Research also shows that many victim-survivors do not seek professional

**CONTACT** Nicola Henry  nicola.henry@rmit.edu.au  Social Equity Research Centre, RMIT University, GPO Box 2476, Melbourne 3001, Australia

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

help because of the prohibitive costs of care, privacy concerns, a lack of knowledge about where to go for help, and the increasing wait times for over-stretched and under-funded support services (Fakhry et al., 2017).

Barriers to reporting and getting help may be different or more pronounced for some groups, including indigenous peoples, migrants, refugees, younger or older people, people with disabilities, sex workers, and LGBTIQ+ people (Fiolet et al., 2021; Henry, Vasil, Flynn, Kellard, & Mortreux, 2022). In a systematic review, Bundock et al. (2020) found that adolescents who experienced dating violence encountered multiple barriers to seeking help, including: concerns about confidentiality; fears of isolation; fears about partner retaliation; a lack of trust in professionals; and shame and embarrassment. In a scoping review of Indigenous people's help-seeking behavior for family violence, Fiolet et al. (2021) found that the most common barrier was shame and embarrassment, followed by concerns about community members finding out, and mistrust or fear of service providers. As a result of these and other barriers, most victim-survivors of GBV do not report their experiences to the police or to other authorities (Ceelen et al., 2019; Fisher et al., 2003), and many do not seek professional care through medical or support services (Ansara & Hindin, 2010; Ullman & Filipas, 2001). Instead, they are more likely to disclose their experiences to a friend or family member or not at all (Campbell et al., 2001; Fisher et al., 2003).

Increasingly, victim-survivors of GBV are turning to digital tools, such as applications ('apps') and chatbots, to obtain support or information. Some victim-survivors prefer using these tools, perceiving them as more accessible, private, and responsive than traditional or 'face-to-face' services (Alhusen et al., 2015; Bundock et al., 2020; Glass et al., 2017). Victim-survivors by and large want to 'access care on their own terms and in a confidential manner' and this often 'surpasse[s] their desire to seek out other sources of care or report the assault in the immediate post-assault period' (Munro-Kramer et al., 2017, p. 298). Digital tools are thus a promising alternative or complementary pathway for more traditional reporting, help-seeking, and prevention. Nonetheless, the use of digital tools raises several concerns, especially when developers do not integrate best practice principles, such as trauma-informed or data privacy approaches, or fail to draw on the knowledge of GBV experts. Digital tools may also do little to prevent GBV from occurring in the first place, especially when they are not targeted to potential perpetrators or bystanders. Despite these limitations, given the rapid pace of digital innovation, there is an opportunity to develop ethical and socially responsible digital interventions that can meet the needs of victim-survivors of GBV.

In this article, we propose a theory- and practice-based 'design justice' approach for developing feminist digital tools for addressing GBV. Design justice is a methodological framework that centers the experiences of communities and involves collaboration and co-design in the design process (Constanza-Chock, 2020). We draw on our own experiences of designing and developing 'Umibot', a text-based, feminist chatbot that provides information, support, and general advice to Australian victim-survivors, bystanders, and perpetrators of image-based sexual abuse. Image-based sexual abuse (IBSA) refers to a range of different acts involving the non-consensual taking, creating, or sharing of intimate images (nude or sexual photos or videos), including threats to share intimate images. This type of abuse can also involve pressuring, coercing, or threatening someone into sharing their intimate images, sending unwanted and unsolicited intimate images,

and creating digitally altered images, including ‘deepfakes’ (see Henry2020 et al., ). Umibot provides information on an extensive range of IBSA-related topics, including laws in Australia, reporting options, support services and tools, and online safety strategies.

In the first section of the article, we provide a summary of the scholarly literature on digital tools for responding to GBV, including feminist chatbots. In the second section, we outline our three-stage design justice approach to building Umibot for victim-survivors, bystanders, and perpetrators of IBSA. We argue that our practice-based, design justice approach, which is underpinned by feminist, trauma-informed, and reflexive principles, can usefully guide the design, development, and deployment of other digital interventions. We argue that design justice is crucial for enhancing individual users’ experiences, reducing potential harms, and promoting transparency and accountability, especially when the subject matter is potentially traumatic and distressing. We conclude by outlining future research directions and underlining the important role that experts, including lived experience and academic experts, should play in developing digital tools to address GBV.

### **Digital tools for responding to gender-based violence**

Digital tools for addressing GBV are diverse and rapidly evolving. Examples include: safety planning and decision-making apps (e.g., Alhusen et al., 2015; Glass et al., 2017); bystander decision aids or training through apps, games, or virtual reality (e.g., Potter et al., 2019); anti-groping devices (e.g., Chowdhury, 2022); wearables, such as anti-penetration devices, personal safety alarms, and GPS ankle bracelets (e.g., Wilson-Barnao et al., 2021); spatial crowd mapping technologies (e.g., Grove, 2015); online professional or peer-to-peer counseling options (e.g., Littleton et al., 2016); alternative reporting tools (e.g., Loney-Howes et al., 2022); and chatbots (e.g., Bauer et al., 2020; Maeng & Lee, 2022).

Meta-analysis of the use of digital tools shows promising results in some contexts, including bullying (e.g., Chen, Chan, et al., 2022), alcohol and drug abuse (e.g., Schouten et al., 2022), and smoking cessation in teen pregnancy (e.g., Griffiths et al., 2018). In relation to GBV, research shows that digital tools can provide meaningful support to victim-survivors. For instance, in a meta-analysis conducted on digital interventions for enhancing mental health outcomes for victim-survivors of intimate partner violence, Emezue and Bloom (2020, p. 7) found that digital interventions can enhance victim-survivors’ decisional certainty, safety behaviors, and psychological outcomes. Emezue and Bloom (2020) note that digital interventions ‘prioritize privacy, confidentiality, and user safety while offering reliable and personalized real-time care to meaningfully and clinically improve health and wellbeing’ (p. 2). In a systematic review, El Morr and Loyal (2020) likewise found that information and communications technology (ICT)-based interventions have ‘the potential to be effective in spreading awareness about and screening for IPV [intimate partner violence],’ and that ‘ICT use show promise for reducing decisional conflict, improving knowledge and risk assessments, and motivating women to disclose, discuss, and leave their abusive relationships’ (p. 10). While these findings are promising, chatbots that aim to address different forms of GBV have received very little scholarly attention to date, and more research is needed to investigate the efficacy and impacts of such interventions. In the section below, we briefly

explain how chatbots work, before then exploring different chatbots for addressing GBV and inequality.

### ***Feminist chatbots***

A chatbot, sometimes referred to as a ‘virtual assistant,’ ‘conversational agent,’ or ‘bot,’ is a computer program that mimics human conversation using text, voice, or a combination thereof. Users can access chatbots in a range of ways, such via messaging apps or standalone websites. Advancements in artificial intelligence (AI), particularly those in the AI subfields of machine learning and natural language processing (NLP), have heralded a new era of so-called ‘intelligent’ text-based chatbots over the past decade (Gupta et al., 2020; Rapp et al., 2021). Text-based chatbots can be categorized into three general types (Gupta et al., 2020). The first type is a ‘button-based’ or ‘rules-based’ chatbot that allows users to select from a menu of pre-determined buttons which trigger relevant responses and draw on pre-programmed code. The second type is a ‘hybrid’ chatbot that is both rules- and context-based, enabling users to select from a menu of buttons *and* input text (e.g., ask questions or pose statements). The third and most technically advanced type is a ‘contextual chatbot’ that uses machine learning, neural networks, and other types of AI to interpret and respond to user sentiment and intention in ways that attempt to mimic a free-flowing, interactive human conversation. Newer ‘generative AI’ chatbots, like ChatGPT, learn through vast volumes of human-created datasets, pre-programmed scripts, as well as interactions with users using NLP to create original content and language.

Chatbots are an increasingly common feature in customer service support for travel, banking, online shopping, and other services. Households around the globe have adopted chatbot technology in the form of virtual personal assistants (e.g., Apple Siri, Google Home, and Amazon Alexa), as well as using conversational chatbots as part of their everyday information search activities, such as ChatGPT and Google Gemini. Chatbots are also used to provide alternative pathways for accessing support, advice, and information for people facing interpersonal challenges. For example, counseling-style chatbots rely on psychotherapeutic interventions, such as cognitive-behavioral therapy (CBT), to help people experiencing bereavement, abuse, addiction, and mental health difficulties (e.g., Durden et al., 2023). These chatbots can be (1) ‘conversational’ to mimic a text-messaging conversation a user might have with another human; (2) ‘informational’ to raise awareness and/or provide validation and referrals to information and human support; (3) ‘reporting-oriented’ to facilitate reporting to the police or other agencies; or (4) a combination of all three.

Chatbots have also been designed and developed for victim-survivors of domestic, family, or sexual violence as a potential solution to the many barriers associated with reporting abuse and accessing specialist support. As explained above, these barriers include concerns about confidentiality, shame, and embarrassment, and increasing costs and waiting periods. Chatbots can also reduce the burden on overly stretched support services. Examples of chatbots in the context of domestic, family, or sexual violence include: ‘#MetooMaastricht,’ a chatbot which provided support to victim-survivors of sexual harassment (Bauer et al., 2020) (no longer in operation); Chayn’s ‘Little Window Bot’ for women experiencing abuse (no longer in operation); Good Hood’s ‘Hello Cass,’ a

web and SMS chatbot for Australian victim-survivors of sexual or domestic violence (no longer in operation); 1800 RESPECT's 'Sunny' for Australian women with disabilities who have experienced sexual violence; AI for Good's 'rAIInbow' (or 'Bo') for South African survivors of domestic abuse (no longer in operation); and Kona Connect's 'Sophia' chatbot for victim-survivors of domestic violence around the world. There are also chatbots and online reporting tools for reporting sexual violence, such as 'Talk to Spot,' and chatbots for feminist activism, such as the text-based 'Betânia' chatbot for Brazilian feminists to mobilize against a Constitutional amendment criminalizing abortion (Toupin & Couture, 2020), and Ciston's (2019) 'ladymouth' chatbot that explained feminism to men's rights activists on Reddit. We speculate that many of these chatbots are no longer in operation due to the significant costs associated with their development and maintenance.

Several chatbots for people who have experienced technology-facilitated abuse have also emerged in recent years. Technology-facilitated abuse is an umbrella term that describes the use of digital technologies to perpetrate different forms of interpersonal harassment, abuse, and violence in the context of sexual or family violence, prejudice-based hatred, and othering (e.g., doxxing, online sexual harassment, online hate speech, cyberbullying, IBSA – see e.g., Bailey, Flynn, & Henry, 2021). For example, Plan International and Feminist Internet's 'Maru' chatbot was co-designed with young activists to assist women and girls experiencing or witnessing online harassment. The UK Safer Internet Centre's 'Reiya' chatbot was designed to assist victim-survivors with online abuse, including bullying, impersonation, harassment, and IBSA. In 2023, Netsafe, New Zealand's online safety agency, launched the 'Kora' chatbot for people seeking information and support about online safety and abuse. There are also IBSA-specific chatbots although they are not currently available to the public. These include Falduti and Tessaris's (2022) 'proof of concept' chatbot to help Italian victim-survivors to report IBSA, and Maeng and Lee's (2022) hybrid IBSA chatbot for South-Korean victim-survivors. None of the chatbots mentioned here are based on generative AI models. Instead, they are predominantly rules-based to deliver specific information and support to users.

Not all chatbots for addressing GBV can be classified specifically as 'feminist chatbots,' even though they address feminist issues. At present, there is no definition in the scholarly literature of what a 'feminist chatbot' is, and only one article that uses this term (Toupin & Couture, 2020). We define a feminist chatbot as *a digital tool that provides trauma-informed, survivor-centric information and support to individuals (e.g., victim-survivors, bystanders, perpetrators, advocates, frontline workers) with the broader goal of addressing gender-based inequalities and gendered power relations*. We argue that a feminist chatbot is one that is premised on broader political strategies for addressing gender inequality to promote survivor empowerment, collective awareness, harm-reduction, and cultural change. Essentially, a feminist chatbot should be informed by feminist theoretical frameworks and feminist design justice principles.

One example of a feminist chatbot is the Maru chatbot, designed by Feminist Internet in partnership with Plan International. Maru was developed using a collaborative methodology, involving co-design with young activists from Global South and North countries. The development of Maru was informed by feminist design principles to consider barriers people face accessing the bot, as well as stereotypes and biases that might exist within the team. The chatbot uses language and design elements that are

empathetic, inclusive, and trauma-informed. While there is no ‘quick exit’ button for users to quickly exit the website, the Maru chatbot provides trigger warnings, cybersecurity tips, a privacy statement, terms of use, and other resources. The content is tightly controlled through buttons-only (there are no free-text options).

As limited scholarly attention has been paid to how chatbots have been, or could be, designed and developed to address GBV in practice, we argue there is a need for a blueprint to guide future initiatives (see UNICEF, 2022). In the sections that follows, we propose a feminist, design justice approach to digital interventions that specifically focus on GBV.

## **A ‘design justice’ approach to developing feminist chatbots**

Design justice, according to Constanza-Chock (2020), ‘... rethinks design processes, centers people who are normally marginalized by design, and uses collaborative, creative practices to address the deepest challenges our communities face’ (p. 6). Design justice can provide both a critical lens for analyzing the reproduction of structural inequalities through design, and a set of strategies to ameliorate those inequalities (Constanza-Chock, 2020). Design justice recognizes the importance of social justice, feminism, and intersectionality in guiding the design process at all stages – from conceptualization to operationalization to evaluation. Design justice is thus part of the increasing shift towards safer and more ethical and equitable ‘tech’ in the multi-disciplinary field of human-computer interaction (HCI), which investigates the ways in which people interact with and use computers and other technological systems (MacKenzie, 2024). An allied social justice approach in this field is the ‘trauma-informed approach to computing,’ which similarly adopts principles of safety, trust, peer support, collaboration, enablement, and intersectionality to inform the development of digital interventions (Chen, McDonald, et al., 2022).

Design justice should be based on key feminist principles, including: challenging and redistributing power; care, wellbeing, and non-violence; respect for the earth; deep democracy; and intersectionality (IWDA, 2019). A design justice approach recognizes the complex and structural nature of gendered power relations, including the ways that multiple structural inequalities converge or intersect to shape experiences of discrimination, exclusion, and violence (Cho et al., 2013). It involves reflexivity on behalf of researchers, including consideration of the relationship between power and knowledge, a focus on the lived experiences of victim-survivors, and attention to the ways in which ‘care’ itself is not necessarily neutral, but can be embedded in racial relations and colonialism (Raghuram, 2021). A design justice approach is also ‘strengths-based’ to promote self-awareness, knowledge of rights, self-efficacy, self-care, problem-solving skills, and access to resources. This aligns with existing strengths-based interventions for responding to GBV, focusing on ability, knowledge, and capacity, rather than deficit or passivity.

In 2021, drawing on these principles of design justice, we developed Umibot: a hybrid and text-based chatbot that provides information, support, and general advice to victim-survivors, bystanders, and perpetrators of image-based sexual abuse (IBSA). As outlined above, we define IBSA in terms of the non-consensual taking, creating, or sharing of intimate images, including threats to share images, unwanted intimate images, and pressure

or coercion to share images. This form of abuse can have devastating impacts, including suicide or suicidal ideation, anxiety, depression, social isolation, powerlessness, paranoia, hypervigilance, low self-esteem, and mistrust of others (Bates, 2017; Henry, 2024; McGlynn et al., 2021). While recognition of the harmful nature of this type of abuse has prompted far-reaching changes in policy and law in jurisdictions around the globe, victim-survivors continue to face a wide range of barriers to reporting and help-seeking. Victim-survivors are often unaware if IBSA is a crime in their country, believe that nothing can be done to address the abuse, do not know where to go for help, or worry about being blamed (Henry, 2024). Victim-survivors often report feeling disempowered, lacking in control, and let down by criminal justice responses or support services when they do report their experiences (Henry, 2024; Henry et al., 2021; Rackley et al., 2021).

Building on this existing knowledge, we adopted a design justice approach to designing and developing Umibot. We centered the experiences of affected communities in the design process through a series of consultations with community organizations, lived experience experts, and academic experts whose work focuses on diversity, gender, and sexuality. We drew on ‘feminist ethics of care’ principles, with a particular focus on producing knowledge based on the lived expertise of victim-survivors, prioritizing a positive duty of care to protect people from risk and vulnerability (Leurs, 2017; Gray & Witt, 2021), and promoting empowerment and self-efficacy. Our approach was also ‘trauma-informed,’ recognizing that IBSA can be a traumatic experience for many people, as well as part of a lifetime of experiences marked by trauma, violence, and abuse. We explain below how we drew on design justice, trauma-informed computing, and feminist principles in the development of Umibot across the three key stages of design, content, and evaluation.

### ***Stage 1: design***

There are three key considerations that can guide the design of feminist chatbots in line with a design justice approach: (1) personality, branding, and conversational tone, including the aesthetics of the artwork and color design; (2) ‘privacy-by-design’ and ‘safety-by-design’; and (3) functionality, including how a chatbot ‘learns’ and delivers information in practice. We discuss each in turn below.

#### ***Personality, branding, and tone***

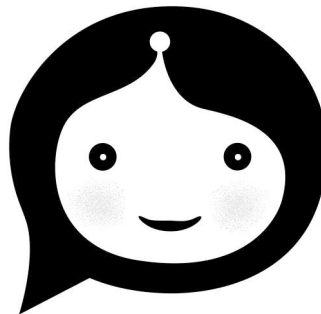
In relation to personality, branding, and tone, there are two interconnected concerns or ‘design dilemmas’ that require careful consideration when designing feminist chatbots. The first dilemma relates to anthropomorphism and whether to create a human-looking or non-human-looking character. Anthropomorphism is the attribution of human characteristics to animals, objects, or abstract ideas, including emotions, motivations, or intentions. Although the research is ‘new’ and ‘fragmented’ on the benefits of anthropomorphized AI-enabled technology (Li & Suh, 2022), some studies suggest that chatbots with human-looking names and design features can build credibility and trust with users, and increase perceptions of similarity between a user and the chatbot, potentially leading to greater user satisfaction (Borau et al., 2021; see also Zogaj et al., 2023).

The second major design dilemma that we encountered was the representation of gender and race. Many commentators have argued that the feminization of digital assistants and chatbots reinforces problematic gender biases and stereotypes, such as women being positioned as more empathetic or more suitable for caring or domestic roles than men (e.g., Strengers & Kennedy, 2021). In relation to race and gender, Vorsino (2021, p. 111) notes that ‘any chatbot or form of communicative AI is, inherently, a project of racialization and gendering.’ This is not only because the visual representations of chatbots can be taken to represent a particular racial and gendered identity, but the conversational style (e.g., colloquialisms, jokes, dialogue patterns) can also reflect certain race and gender biases (Marino, 2014).

For the design of Umibot, which is housed on a standalone website, we undertook a four-stage consultation process to seek advice on the chatbot’s character design and content. The process involved: (1) three online focus groups with feminist researchers, with each focus group comprising up to eight participants and running for 1.5 h: 30 min for participants to test a beta version of the chatbot in advance, and 1 h for the focus group discussion itself; (2) semi-structured online consultations with seven lived experience and intersectionality<sup>2</sup> experts; (3) two online workshops with members of our national and international advisory groups that include academics, victim-survivors, technology companies, government agencies, and support services; and (4) semi-structured interviews with victim-survivors of IBSA, with each interview lasting for 1.5 h, including 30 min to test Umibot. Across the different stages, we collated and prioritized feedback that was raised by the different groups. At the time of writing this article, we were collating and addressing feedback based on our ongoing interviews with victim-survivors, alongside regular (monthly) maintenance of the chatbot.

Based on advice given through these consultations, we chose an anthropomorphized chatbot character design and name that aims to emulate a ‘trusted advisor’ who is empathetic, factual, and reassuring (see Figure 1). We developed a gender-neutral palette that is soft, strong, light, and warm, including the colors of light beige, soft mauve, purple, and orange. We also made Umibot’s character a static and transparent ‘ambiguous image,’ also known as a ‘reversible figure,’ which exploits graphical similarities to represent two or more images simultaneously.

The effect is that some people will interpret Umibot as a robot with an antenna in a black speech bubble, whereas others will see a feminized character with a headscarf or



**Figure 1.** Umibot’s character design.



a bob haircut. As depicted in Figure 1, Umibot's face is transparent in the way that it takes on whatever color is used in the background. These design choices ultimately help to make Umibot an indeterminate race and gender, as requested by our focus group participants and consultants, to ensure the chatbot is more accessible and appealing to diverse groups of people.

### ***Privacy and safety by design***

'Privacy-by-design' and 'safety-by-design' are two fundamental principles of design justice for implementing a reflexive 'ethics of care' approach to data collection, storage, and analysis (see e.g., Leurs, 2017; Luka & Millette, 2018). First, privacy-by-design is of growing importance for the development of digital technologies that place privacy protection at the forefront of design, rather than privacy being an after-thought or add-on (Cavoukian, 2009). Privacy-by-design components can include: creating default settings on web- and app-based platforms that are set to the highest privacy levels; engaging experts to test existing technologies against jurisdictionally-specific privacy laws; and, in a publicly available privacy or other statement, informing users about their rights and how they can enhance their privacy using a tool, app, or platform. This information should be accessible to users through a screen reader or assistive technology.

Privacy protection is crucial for victim-survivors of IBSA and other forms of GBV. At the start of designing Umibot, we undertook a rigorous, multi-stage Privacy Impact Assessment (PIA) in full consultation with RMIT University's Privacy Officer and Legal Counsel, as well as consultations with privacy experts. As part of our PIA, we undertook a detailed analysis of Australian privacy law, mapped all data flows in a conversation with Umibot to identify potential privacy risks (e.g., users inadvertently disclosing personal information in the free-text box), and designed strategies to mitigate potential privacy risks (e.g., Umibot does not collect personal information or keep a record of users' conversations). Once our PIA was approved, we hired Melbourne-based digital agency Tundra, whom we selected through a competitive tender process, to set up Umibot's technical infrastructure in line with our privacy requirements.

We also developed a publicly available Privacy Information Statement (PIS), which aligns with the Office of the Australian Information Commissioner's best practice guidelines. We used plain and understandable language to outline the information that a user might need to make an informed decision, including: the technology behind Umibot (including Amazon Lex, which is specifically for building AI-powered conversational interfaces, and Google Analytics); the types of information that Umibot collects (e.g., only de-identified and aggregate information about how the bot is used, including the total number of users or the most searched for IBSA-related information); why Umibot collects that information for maintenance and research purposes; how information is stored in password-protected data storage; and how users can raise privacy concerns.

RMIT University approved a waiver of consent for chatbot users in line with Australia's National Ethics Statement. A waiver of consent is justified for several reasons, including that the benefits of chatting with Umibot, which provides a confidential, non-commercial, and trauma-informed space for users to chat about their experiences of IBSA, is likely to outweigh any potential risks of harm. It would also would have been impractical and infeasible to obtain users' consent as Umibot collects de-identified

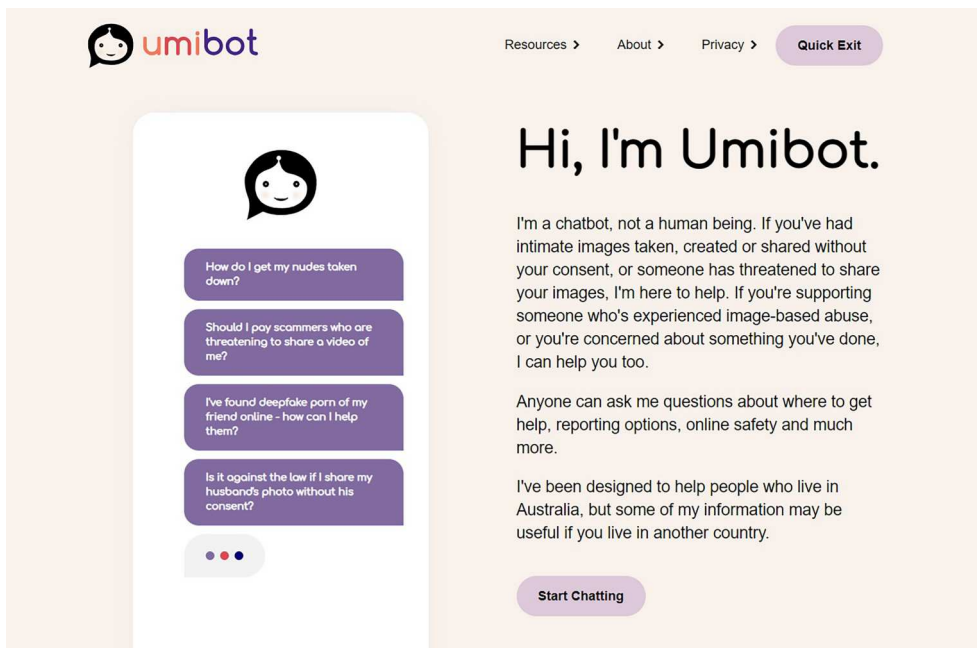
and aggregate data about how the chatbot is used – Umibot does not store any personal information or keep a record of conversations to mitigate privacy issues.

In addition to privacy-by-design, safety-by-design is an interconnected principle that can significantly enhance the ethical design and development of digital tools (Strohmayr et al., 2022). The starting point for our safety-by-design process was to undertake a Cybersecurity Assessment at RMIT University, which ultimately led to us housing Umibot within RMIT’s cloud infrastructure that has additional layers of security, among other benefits. We also implemented a ‘quick exit’ button that users can access at any time (see Figure 2).

Another key component of our safety-by-design approach was to inform users about the possible safety risks of using the chatbot, such as an abusive partner monitoring their device. We did this through a separate ‘Resources’ page on Umibot’s standalone website that functions as a digital library with relevant, publicly available information, including security and safety resources which include information about accessing the chatbot through a safe device, deleting browser history, or using incognito mode. The national ‘support’ resources in the Resources page also provides information on how to seek help through formal and informal support channels, such as victim support services, online safety agencies, the police, workplaces, schools, and universities, as well as more informal support channels, including friends and family members.

### Functionality

One of the most important aspects of how a chatbot functions is the delivery of information to users. Unlike generative AI chatbots that learn through conversations with users and large-scale data from different sources, we adopted hybrid functionality for

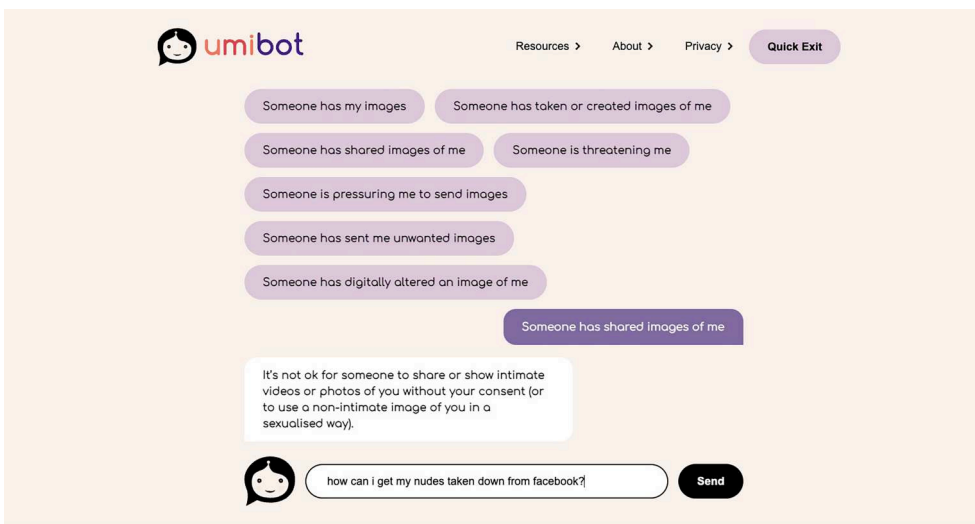


**Figure 2.** Screenshot of Umibot’s landing page, including ‘quick exit’ button.

Umibot. That means that users can select from pre-programmed buttons and type questions or statements in a free-text box, as illustrated in [Figure 3](#). We chose to design a hybrid chatbot as opposed to a less costly and less labor-intensive ‘rules-based’ chatbot. This was because we wanted to give users the opportunity to engage with the bot in an interactive and human-like way. Hybrid models, as previously outlined, attempt to emulate the characteristics of a human conversation by using both rules-based (i.e., basic ‘if this, then that’ logic) and NLP functionality which teaches a chatbot to detect and correctly respond to a range of words and utterances, which can be time-consuming work. We did not go as far as to make Umibot a generative chatbot, or one that uses ChatGPT or other plug-ins, to eliminate the risk of the chatbot going ‘rogue’ by, for example, learning harmful content from its interactions with users and then delivering that content to potentially vulnerable chatbot users (see Vorsino, 2021).

To promote the values of care, safety, and trust, we tightly controlled the ‘learning’ that makes Umibot’s hybrid functionality work. By ‘tightly control,’ we mean that we manually ‘teach’ Umibot a wide range of natural language utterances by inputting text into Amazon Lex (the software behind Umibot’s conversational interface). Utterances range from full questions (i.e., ‘what is image-based sexual abuse?’) to short phrases without punctuation (i.e., ‘what is image-based abuse’) to single words (i.e., ‘hello’), and can include any number of ‘slots’ or synonyms for key words. These utterances and slots ultimately form a tailor-made, expert knowledge-base on IBSA that enables Umibot to respond to freely typed queries (statements or questions) from users. For context, Umibot’s expert knowledge-base comprises thousands of utterances and hundreds of slots, each containing hundreds of synonyms, all of which derive from Umibot’s more than 500 pages of content.

There are several limitations to manually teaching Umibot in this way; principally, that the process of expanding the chatbot’s knowledge is time-consuming and slower than more automated methods. Another downside is that Umibot can miss or not detect



**Figure 3.** Screenshot of Umibot’s hybrid chat window.

queries that are slightly different to the natural language utterances that we have manually inputted, such as utterances that have the same meaning, but different syntax. If Umibot cannot detect an utterance, it returns a ‘fallback response’ (e.g., ‘I’m still learning. Could you please rephrase your question?’). We also taught the bot to respond to certain trigger words or phrases, such as those related to sexual or domestic violence, child sexual abuse, or self-harm. This ensures that users do not get a fallback response when using those key terms and instead are given a sympathetic response coupled with a list of support services. A tightly controlled hybrid approach that is responsive to disclosures of self-harm or violence best aligns with trauma-informed approaches to the development of digital tools for addressing GBV.

Finally, crucial to design justice is *accessibility*. Prior to launching our chatbot, we commissioned Intopia, a digital accessibility and equitable design agency, to undertake an independent accessibility audit of the chatbot to assess its compatibility with the Web Content Accessibility Guidelines (WCAG) 2.1. Accessibility was measured in terms of the extent to which the chatbot is accessible to as many people as possible. Intopia’s expert accessibility auditors reviewed select pages and features of the chatbot, the results of which are publicly available on the chatbot’s website as part of an Accessibility Statement. The Statement lists the issues identified by the independent auditors and the action steps that we took.

## **Stage 2: content**

Chatbots are resource-intensive and costly to both create and maintain. The foremost technical limitation is that chatbots designed to address GBV do not tend to engage conversationally with users in the same way as generative AI models like ChatGPT. This is mainly due to the need to control content delivered to potentially vulnerable users. In this section, we discuss the importance of content for feminist digital interventions, drawing on five key considerations.

### ***A team of experts***

First, content should be written by, or at least heavily informed by, *experts in the field* to minimize the risk of providing inaccurate or otherwise harmful information to users who may be in a state of significant distress at the time of using the tool. To develop Umibot’s content, we drew on decades of our own GBV and IBSA research, as well as the expertise of community stakeholders with whom we consulted, including victim-survivors, community and governmental organizations, and academic experts.

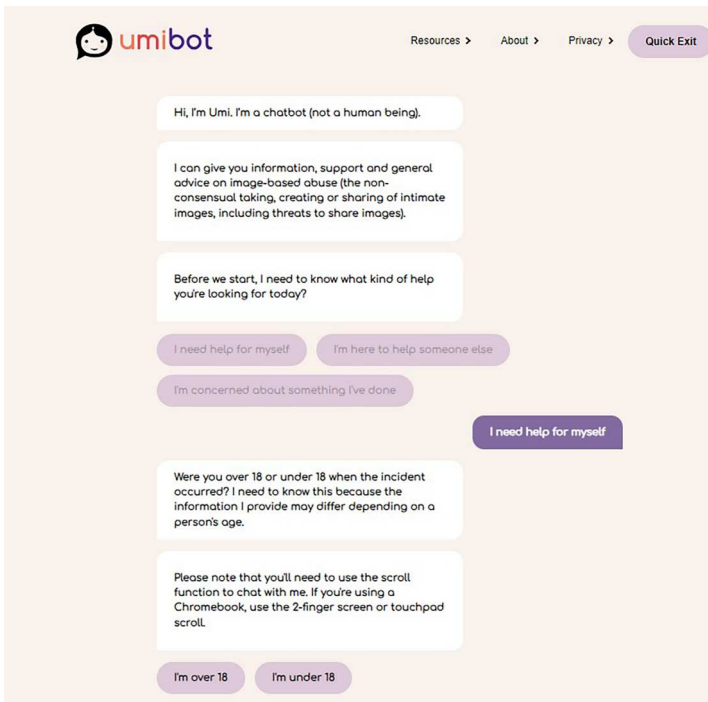
### ***Appealing to diverse audiences***

A second key consideration in the design content for feminist chatbots is appealing to *diverse audiences*. Feminist digital interventions need to be gender-inclusive to provide support to all genders using gender-neutral and non-heteronormative language (e.g., language not relating to a particular gender or sexual orientation) (Debnam & Kumodzi, 2021). For Umibot’s responses, we utilized language that is gender-inclusive, as well as provided support, information, and advice to diverse groups. For instance, at the start of a conversation with Umibot, as shown in [Figure 4](#), users must answer two mandatory questions: first, whether they are seeking help for themselves or someone else, or whether

‘they are concerned about something they have done’; and second, whether they were over 18 or under 18 when the incident occurred. After this point, users are channeled into different pathways according to one of five main user groups: youth victim-survivor; adult victim-survivor; youth bystanders; adult bystander; and youth/adult perpetrator. In total, there are 260 individual, specifically tailored ‘intents’ or main topics of conversation across the entire chatbot.

The victim-survivor content is the most comprehensive in Umibot, providing information, support, and advice to people who have had one or more experiences of IBSA. The bot provides content on the following topics: ‘someone has my images,’ ‘someone has taken or created images of me,’ ‘someone has shared images of me,’ ‘someone is threatening to share my images,’ ‘someone is pressuring me to send images,’ ‘someone has sent me unwanted images,’ or ‘someone has digitally altered an image of me.’ Victim-survivors are presented with pre-programmed buttons, such as ‘What can I do?’, ‘Laws in Australia,’ ‘Reporting options,’ ‘Get support,’ and ‘Collect evidence.’ These buttons lead to other pre-programmed options at various points in the conversation, such as ‘Report to the police,’ ‘Report to a platform,’ ‘What will happen to the person I’m reporting?’, ‘Get legal advice,’ ‘Looking after yourself,’ and ‘Domestic and sexual violence support’ (and many more).

Bystanders are another important target audience for GBV interventions as they can enact significant harm through victim-blaming and judgmental attitudes, yet they can also provide invaluable support, practical assistance, and witness statements (Alhusen et al., 2015). In the chatbot, we provide a range of information and support to bystanders,

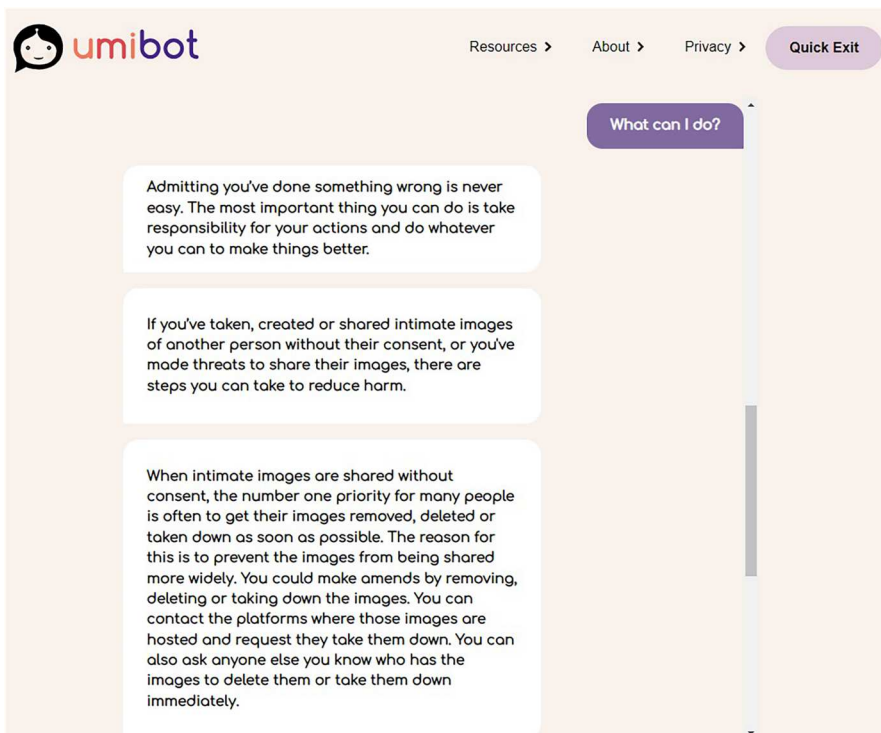


**Figure 4.** Screenshot of Umibot’s two mandatory questions.

including the types of behavior that can constitute IBSA, how to respond to someone who has experienced IBSA (e.g., to avoid victim-blaming statements), where and how to report the abuse, and a comprehensive list of expert support services. We provide an overview of laws in Australia, and more general information about IBSA-related topics, including the meaning and importance of dynamic, affirmative consent. We also emphasize the importance of self-care and outline how bystanders can take care of themselves after witnessing or hearing about abuse, including support services that they can seek help from.

The chatbot additionally provides limited information for perpetrators or people who are concerned about something they have done. There are several challenges in developing perpetrator-specific content. The content needs to be educative, with a view to reducing harm and preventing abuse, while not colluding with wrongdoers or condoning their behavior. For Umibot, we chose and drafted the content for perpetrators carefully, deciding to focus on the nature of IBSA and why it is harmful. We provide advice, useful resources, and support options to this group of users. The aim was to provide advice about actionable steps that the person could take, referrals for mental health support services, as well as information about the laws, which were carefully written to avoid alarm for the user.

As illustrated in [Figure 5](#), we did not make the free-text box available in the perpetrator or bystander pathways due to resource constraints as well as our desire to carefully



**Figure 5.** Example of content for perpetrators.

control the information provided in those pathways, especially for perpetrators as explained above.

Umibot also provides links and tailored content for people from diverse backgrounds and experiences. This is either through pre-programmed buttons (e.g., on domestic or sexual violence), through the free-text functionality, or through the information provided in the Resources page (e.g., information on reporting, platform policies, security and safety, support services, and collecting evidence). For instance, Umibot provides specific information for Indigenous people, LGBTQIA+ people, migrants, older people, and sex workers, including specific laws (e.g., laws in residential care; anti-discrimination, piracy, and copyright laws etc.), support organizations, and suggested resources (e.g., the UK Revenge Porn Helpline's tips for sex workers).

Another key example of how we approached diversity is through the different pathways for youth and adults. For example, the youth victim-survivor pathway includes referral information for people who have experienced child sexual abuse. The language, advice, and information provided are also different to that of the adult content; for instance, the laws on child sexual abuse material, or information on different support services. Furthermore, we were careful to strike a balance between giving information about potential laws in Australia that criminalize young people for consensually taking, storing, and/or sharing their own intimate images, and not causing alarm about the existence of such laws. Relatedly, we were careful about giving advice to young people on how to engage in safe sexting without putting them at risk of being prosecuted for child sexual abuse offenses.

### ***A sex and ethics approach***

Third, a '*sex and ethics*' approach (Carmody, 2008) provides important guidance for the development of content for feminist digital tools. This is a strengths-based approach which strikes a balance between information and education on risk and pleasure. It also prioritizes user confidence and decision-making autonomy through presenting different choices and options for action. We aimed to achieve this through language that is empathetic yet empowering, as well as more practically by presenting users with different options to help them to make an informed decision for their next steps. Research has shown that abstinence-focused educational messaging is disempowering and undermines people's decision-making about their own sexual health, safety, and wellbeing (Döring, 2014). Digital tools, including feminist chatbots, need to raise awareness about the wrongs of the non-consensual sexual acts, with a focus on respectful relationships and ethical sexual practices. For example, Umibot acknowledges that consensual intimate image sharing is a normal practice that can be mutually pleasurable, but also mentions some of the risks associated with intimate image sharing and how to mitigate these risks (e.g., ensuring images do not have an identifiable backdrop or features). Umibot also acknowledges the unfairness of having to take steps to protect against the non-consensual sharing of intimate images and carefully presents mitigation strategies as optional steps that users can consider.

### ***A trauma-informed approach***

Fourth, the content for feminist chatbots needs to be '*trauma-informed*.' Trauma is an emotional response to a deeply distressing or disturbing event that can include shock,

difficulties with processing the experience, reliving the event, or feeling overwhelmed, depressed, or anxious (Herman, 2015). A trauma-informed approach to developing digital interventions recognizes the unique impacts of GBV across a person's lifetime and seeks to reduce the possibility of retraumatization and further harm. Drawing on the authors' decades of research on sexual violence and trauma, we were mindful that trauma is a contested concept that on the one hand can be validating of the harms, yet on the other hand can be used in ways that fail to capture the complexity of experience after abuse. As such, we recognized that IBSA is often experienced as constant, life-altering, and 'devastating' (McGlynn et al., 2021) – in some instances, as 'traumatic' or a 'trauma,' particularly when it co-occurs with other forms of abuse, including sexual violence, sexual harassment, or domestic and family violence (Henry et al., 2021; Henry, Gavey, & Johnson, 2023). However, we do not assume that all victim-survivors necessarily have experiences that fit neatly into this schema.

At a practical level, we adopted several trauma-informed strategies, including establishing a list of trigger words that - when typed - divert the course of the conversation to focus on providing immediate support and resources the user can access. Trigger words includes those relating to danger, self-harm, child sexual abuse, sexual or domestic violence, and police or emergency services (e.g., 'I was raped,' 'I no longer want to live,' 'my husband scares me,' or 'I need the police'). We used sensitive and non-victim blaming responses that are empathetic, supportive, caring, and non-judgmental. The bot also recognizes the diversity of the impacts of IBSA, and the diversity of pathways for recovery.

### **Recognizing structural power and inequality**

Finally, feminist digital tools need to recognize the ways that *structural forms of power, inequality, and positionality* shape experiences. This entails a recognition that people not only experience GBV differently, but the many barriers to reporting or seeking help are shaped by intersecting systems and structural inequalities that affect people differently. Some scholars suggest that such interventions often prioritize criminal justice responses over alternative justice approaches, legitimizing a system that perpetuates injustices against indigenous women and women from other marginalized communities (Shelby et al., 2021). For Umibot, we considered the problems often encountered by victim-survivors with the police and the criminal justice system and were careful to provide users with different options for reporting. Umibot explicitly acknowledges that for some people, reporting to the police or other authorities might not be an option because of past experiences, negative perceptions, and/or safety concerns. For example, when a user seeks information on reporting options, Umibot does not present reporting to the police or criminal justice authorities as the first or only option. Umibot also provides general advice to users who feel like the police or other authorities have not taken them seriously.

### **Stage 3: evaluation**

To assess the efficacy of feminist digital tools, including chatbots, it is crucial to undertake evaluation and review. As Jewkes and Dartnall (2019) note, 'we need more data about the niche that web-based technologies are designed to fill before we invest in them more extensively ...' (p. 270). To enhance our understandings of how these technologies work in practice, including their design and content, stakeholders can use a range of



research methods to collect empirical data including analytics software, randomized controlled trials, questionnaires, exit surveys, semi-structured interviews, and focus groups. These investigations can occur across different stages (Stages 1, 2 and 3 outlined above) and at different times (e.g., pre-, and post- public launch). Prior to release, stakeholders might also run focus groups to elicit feedback on the design and content of a beta version of a digital tool or co-design the tool with target user groups.

After launching the tool, researchers can collect data in two main ways. First, focusing on the impacts and outcomes of digital tools facilitates an assessment of user acceptability, benefits, risks, and limitations, including psychosocial outcomes (e.g., decision-making, self-esteem, and mental health), information gathering (e.g., raising awareness, increasing knowledge, access to referral pathways), and safety promotion (e.g., privacy, security, and safety). Analytics software can collect this data in real-time when users first encounter and use the tool. Second, drawing on intersectionality (Figueroa, Luo, Aguilera, & Lyles, 2021) as well as HCI perspectives and methodologies, can inform understandings of the user experience in terms of whether the tools are acceptable to users, whether needs are being met, the barriers of use, and the ways in which the tool is used. This might include collecting textual, verbal, or biofeedback data through interviews, the use of virtual reality, surveys, or focus groups (Lazar et al., 2017).

Since Umibot's launch in late 2022, we have been using analytics software to collect de-identified aggregate data about how the chatbot is used, such as information about the total number of users and most frequently sought-after information. At the time of writing, Umibot has had over 3,380 conversations with more than 2,600 unique users in over 60 countries. We are also conducting semi-structured interviews with victim-survivors to assess the chatbot's usability, helpfulness, and acceptability. Part of the interview involves us conducting a 'walk-through' of Umibot with participants, who we then ask a series of questions about the tool's design, content and what future versions of the chatbot should look like. Finally, we are using a Qualtrics survey to collect aggregate (de-identified) data about the use and impacts of the chatbot in practice, such as whether users learned new information about IBSA that helped guide their decision-making.

## Conclusion

In this article, we proposed a design justice approach to developing feminist tools for detecting, preventing, and/or responding to GBV. We focused on feminist chatbots as one technological solution to the many known barriers to help-seeking in this context. Chatbots are different to other feminist digital tools for reporting or mapping crime or building safety strategies for victim-survivors of domestic violence. As many victim-survivors and perpetrators of GBV choose not to report or disclose their experiences due to shame and stigma, or fear of chastisement, chatbots can provide much-needed confidential, non-judgmental, personable, interactive, and timely support, information, and general advice. Such interventions can also ease the burden on generalist sexual and domestic violence support services that are often overwhelmed and not always equipped with the specialist knowledge that is needed to assist victim-survivors experiencing technology-facilitated abuse. While chatbots are not a replacement for human support, they can function as an important intermediate step, and can help to connect individuals with human support services.

**Table 1.** A three-stage design justice approach to feminist digital interventions**Stage 1: Design**

Key design considerations:

- Personality, branding, and tone, including how gender and race are represented;
- Privacy and safety-by-design principles and, most critically, how they can be embedded into the technology itself;
- Functionality, including how a digital tool ‘learns’ (e.g., supervised machine learning) and/or delivers information in practice.

Methods of refining and testing a tool’s design include:

- Focus groups and consultations with lived experience and other experts;
- Hiring a technical developer to build a tool;
- Privacy impact assessments;
- Cybersecurity assessments;
- Accessibility audits in line with the latest Web Content Accessibility Guidelines (WCAG).

**Stage 2: Content**

Key content considerations:

- Information should be written by, or at least heavily informed by, experts in the field to minimize the risk of providing inaccurate or otherwise harmful information;
- The tone of content should be empathetic, and strike a balance between detail and brevity, particularly given that some users might be distressed and have varying degrees of literacy when seeking information, support, and/or advice;
- Content should cater to audiences from different backgrounds.

Methods of drafting and refining the content of a digital tool include:

- Document readability testing;
- Expert external review of draft content.

**Stage 3: Evaluation**

Key evaluation considerations:

- Qualitative and quantitative data collection for empirical analysis of a tool’s use, efficacy, and impacts in practice;
- Evaluative data should be publicly available, validated, and understandable to diverse audiences.

Methods of evaluating a digital tool:

- Analytics software (e.g., Google Analytics, Amazon Lex analytics);
- Surveys;
- Qualitative interviews with victim-survivors and other key stakeholder groups.

The consequences of victim-survivors not reaching out for support can be significant, including social isolation, maladaptive coping mechanisms (e.g., alcohol and substance abuse), worsening mental health (e.g., including depression, anxiety, suicidal ideation, and self-harm), poor physical health outcomes, and the breakdown of relationships. Chatbots have the potential to help address this by providing accessible, non-judgmental, and round-the-clock support to individuals in need, many of whom might be in a state of crisis or distress. In this way, chatbots can play a role in the broader ecosystem of responses to GBV across the domains of primary prevention, early intervention, and response. Chatbots can also play an important role in disseminating information to people who may not readily seek out external help and support, as well as to the general public, including bystanders. In the context of the growing recognition of how technology can be weaponized by perpetrators of GBV, feminist chatbots also provide a useful and timely illustration of the different ways in which digital technologies can be harnessed for good.

Despite the recognition of the potential benefits of technological solutions to seeking help for GBV, limited scholarly attention has been paid to the processes of designing and developing feminist chatbots and other digital interventions, and *how* they work in practice. We sought to address this gap by outlining our theory-driven and practice-led design justice approach to developing Umibot, which principally serves as an informational tool for victim-survivors, bystanders, and, to a lesser extent, perpetrators of image-based sexual abuse. As shown in Table 1, we outlined our design justice approach

across three interlinked yet distinct stages: design, content, and evaluation. In the design stage, a key priority was embedding privacy- and safety-by-design principles into the chatbot's technical infrastructure, as well as creating the expert knowledge base behind bot's machine learning functionality. In the second stage, our aim was to ensure the bot's content was informed by research in the field, and was empathetic, trauma-informed, and comprehensive enough to cater to diverse user groups. In the third stage, which we are currently undertaking, we focus on collecting data on the useability, acceptability, efficacy, and impacts of the bot in practice.

Overall, we argue that feminist, trauma-informed, and reflexive principles must underpin the development of digital tools from the outset, as part of a design justice approach more broadly. Such an approach should be evidence-based; safe, secure, and respectful of privacy and data ethics; inclusive, intersectional, and socially-culturally relevant; attune to the risks of unconscious bias and resulting power asymmetries; and strengths-based, trauma-informed, and survivor-centered.

Finally, teams building chatbots and other feminist digital interventions to address GBV need both subject matter expertise and technical 'know-how' to meet the needs of users, including input from frontline support services, victim-survivors, community organizations, technology developers, graphic designers, online safety organizations, accessibility experts, and the media. Without multi-stakeholder input, chatbots and other digital tools risk causing further harm to victim-survivors and other potentially vulnerable users.

Future research should investigate victim-survivors' experiences of using feminist tools. This can be done by collecting textual, verbal, or biofeedback data through interviews, virtual reality, surveys, or focus groups. Disaggregated data collected through analytics software can also help to shed light on how they are being used in practice. More research is also needed to empirically investigate the role and efficacy of feminist digital interventions for bystanders and perpetrators for addressing problematic gender attitudes, beliefs, and norms that are key drivers of GBV. Our design justice approach can help to guide this future research, serving as a valuable resource for other developers of feminist digital tools, as well as scholars, frontline services, and other stakeholders investigating the most effective ways to address the global problem of gender-based violence.

## Notes

1. We use the term 'gender-based violence' (GBV) to refer to a set of physical, psychological, and technology-facilitated behaviors, including sexual assault, sexual harassment, domestic and family violence, and other gender-based harms.
2. Intersectionality refers to the combined effect of different structural inequalities, such as racism, colonialism, sexism, ableism, capitalism, homophobia, and transphobia. These structural inequalities create multiple layers of oppression, not only shaping people's experience of violence, abuse, and discrimination, but also creating barriers for help-seeking and support (see, e.g., Cho et al., 2013).

## Acknowledgements

The authors acknowledge funding from the Australian Research Council (ARC) (FT200100604), and the assistance and guidance of: the Tundra team; the Amazon team; the Australian eSafety

Commissioner and team; Justine Henry; Courtney Vowles; Audrey Fitzgerald; Gemma Beard; Alana Ray; Zahra Stardust; Michael Salter; Michelle Gissara; Jasmin Chen; Sophie Hindes; Noelle Martin; Gala Vanting; Clare McGlynn; Nicola Gavey; Sophie Mortimer; Elena Michael; Mary Anne Franks; Asia Eaton; Michelle Gonzalez; Jane Bailey; Emma Holten; Karuna Nain; Michelle Lewis; Kara Hinesley; Mia Garlick; Samantha Yorke; Bronwyn Carlson; Bridget Harris; Nic Suzor; and Karen Bentley.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by Australian Research Council [grant number FT20010064].

## Notes on contributors

**Nicola Henry** (she/her), PhD, is an Australian Research Council (ARC) Future Fellow in the Social Equity Research Centre at RMIT University (Melbourne, Australia). Her research focuses on the prevalence, nature, and impacts of gendered violence, including legal and non-legal responses to addressing and preventing violence. Her current research is focused on technology-facilitated abuse and image-based sexual abuse [email: nicola.henry@rmit.edu.au].

**Alice Witt** (she/her), PhD, is a Lecturer in the Deakin Law School, Faculty of Business and Law at Deakin University (Melbourne, Australia). She researches the exercise of governing power in the digital age, focusing on the intersections of regulation, technology, and gender. In addition to contributions to edited volumes, Alice's socio-legal work has appeared in journals such as *Social & Legal Studies*, *Feminist Media Studies*, and *Artificial Intelligence and Law* [email: alice.witt@deakin.edu.au].

**Stefani Vasil** (she/her), PhD, is a Lecturer in the Thomas More Law School at Australian Catholic University (Melbourne, Australia). She was previously a Postdoctoral Research Fellow with the Monash Gender and Family Violence Prevention Centre (MGFVPC) at Monash University. Her research is focused on the intersections between migration and gendered violence, including domestic and family violence. [email: stefani.vasil@acu.edu.au].

## ORCID

Nicola Henry  <http://orcid.org/0000-0003-2241-7985>

Alice Witt  <http://orcid.org/0000-0003-0505-422X>

Stefani Vasil  <http://orcid.org/0000-0003-1722-8477>

## References

- Alhusen, J., Bloom, T., Clough, A., & Glass, N. (2015). Development of the MyPlan safety decision app with friends of college women in abusive dating relationships. *Journal of Technology in Human Services*, 33(3), 263–282. <https://doi.org/10.1080/15228835.2015.1037414>
- Ansara, D. L., & Hindin, M. J. (2010). Formal and informal help-seeking associated with women's and men's experiences of intimate partner violence in Canada. *Social Science & Medicine*, 70(7), 1011–1018. <https://doi.org/10.1016/j.socscimed.2009.12.009>
- Bailey, J., Flynn, A., & Henry, N. (Eds.). (2021). *The Emerald international handbook of technology-facilitated violence and abuse*. Emerald Publishing Limited.

- Bates, S. (2017). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn on female survivors. *Feminist Criminology*, 12(1), 22–42. <https://doi.org/10.1177/1557085116654565>
- Bauer, T., Devrim, E., Glazunov, M., Jaramillo, W. L., Mohan, B., & Spanakis, G. (2020). #Metoomaastricht: Building a chatbot to assist survivors of sexual harassment. In *Machine learning and knowledge discovery in databases: International workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, proceedings, Part I* (pp. 503–521). Springer International Publishing.
- Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing*, 38(7), 1052–1068. <https://doi.org/10.1002/mar.21480>
- Bundock, K., Chan, C., & Hewitt, O. (2020). Adolescents' help-seeking behavior and intentions following adolescent dating violence: A systematic review. *Trauma, Violence, & Abuse*, 21(2), 350–366. <https://doi.org/10.1177/1524838018770412>
- Campbell, R., Wasco, S. M., Ahrens, C. E., Sefl, T., & Barnes, H. E. (2001). Preventing the 'second rape': Rape survivors' experience with community service providers. *Journal of Interpersonal Violence*, 16(12), 1239–1259. <https://doi.org/10.1177/088626001016012002>
- Carmody, M. (2008). *Sex and ethics: Young people and ethical sex* (Vol. 1). Palgrave Macmillan.
- Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario, Canada*. <http://jppaulgibson.synology.me/ETHICS4EU-Brick-SmartPills-TeacherWebSite/SecondaryMaterial/pdfs/CavoukianETAL09.pdf>
- Ceelen, M., Dorn, T., van Huis, F. S., & Reijnders, U. J. (2019). Characteristics and post-decision attitudes of non-reporting sexual violence victims. *Journal of Interpersonal Violence*, 34(9), 1961–1977. <https://doi.org/10.1177/0886260516658756>
- Chen, Q., Chan, K., Guo, S., Chen, M., Lo, C. K. M., & Ip, P. (2022). Effectiveness of digital health interventions in reducing bullying and cyberbullying: A meta-analysis. *Trauma, Violence, & Abuse*, 24(3).
- Chen, J. X., McDonald, A., Zou, Y., Tseng, E., Roundy, K. A., Tamersoy, A., Schaub, F., Ristenpart, T., & Dell, N. (2022, April). Trauma-informed computing: Towards safer technology experiences for all. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–20).
- Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society*, 38(4), 785–810. <https://doi.org/10.1086/669608>
- Chowdhury, R. (2022). Sexual assault on public transport: Crowds, nation, and violence in the urban commons. *Social & Cultural Geography*, 24(7), 1–17.
- Ciston, S. (2019). 'Ladymouth': Anti-social-media art as research. *Ada: A Journal of Gender, New Media, and Technology*, 1–14.
- Constanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- Debnam, K. J., & Kumodzi, T. (2021). Adolescent perceptions of an interactive mobile application to respond to teen dating violence. *Journal of Interpersonal Violence*, 36(13-14), 6821–6837. <https://doi.org/10.1177/0886260518821455>
- Döring, N. (2014). Consensual sexting among adolescents: Risk prevention through abstinence education or safer sexting. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 8(1), 9.
- Durden, E., Maddison, C. P., Rapoport, S. J., Williams, A., Robinson, A., & Forman-Hoffman, V. L. (2023). Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent "Woebot." *Internet Interventions*, 33, 1–11.
- El Morr, C., & Loyal, M. (2020). Effectiveness of ICT-based intimate partner violence interventions: A systematic review. *BMC Public Health*, 20(1), 1–25. <https://doi.org/10.1186/s12889-020-09408-8>

- Emezue, C., & Bloom, T. L. (2020). PROTOCOL: Technology-based and digital interventions for intimate partner violence: A meta-analysis and systematic review. *Campbell Systematic Reviews*, 17(1), e1132. <https://doi.org/10.1002/cl2.1132>
- Fakhry, S. M., Ferguson, P. L., Olsen, J. L., Haughney, J. J., Resnick, H. S., & Ruggiero, K. J. (2017). Continuing trauma: The unmet needs of trauma patients in the postacute care setting. *The American Surgeon*, 83(11), 1308–1314. <https://doi.org/10.1177/000313481708301137>
- 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* (pp. 96–102).
- Figueroa, C. A., Luo, T., Aguilera, A., & Lyles, C. R. (2021). The need for feminist intersectionality in digital health. *The Lancet Digital Health*, 3(8), e526–e533.
- Fiolet, R., Tarzia, L., Hameed, M., & Hegarty, K. (2021). Indigenous peoples' help-seeking behaviors for family violence: A scoping review. *Trauma, Violence, & Abuse*, 22(2), 370–380. <https://doi.org/10.1177/1524838019852638>
- Fisher, B. S., Daigle, L. E., Cullen, F. T., & Turner, M. G. (2003). Reporting sexual victimization to the police and others: Results from a national-level study of college women. *Criminal Justice and Behavior*, 30(1), 6–38. <https://doi.org/10.1177/0093854802239161>
- Glass, N. E., Perrin, N. A., Hanson, G. C., Bloom, T. L., Messing, J. T., Clough, A. S., Campbell, J. C., Gielen, A. C., Case, J., & Eden, K. B. (2017). The longitudinal impact of an internet safety decision aid for abused women. *American Journal of Preventive Medicine*, 52(5), 606–615. <https://doi.org/10.1016/j.amepre.2016.12.014>
- Gray, J., & Witt, A. (2021). feminist data ethics of care for machine learning: The what, why, who and how. *First Monday*, 26, 18. <http://doi.org/10.5210/fm.v26i12.11833>.
- Griffiths, S. E., Parsons, J., Naughton, F., Fulton, A. E., Tombor, I., & Brown, K. E. (2018). Are digital interventions for smoking cessation in pregnancy effective? A systematic review and meta-analysis. *Health Psychology Review*, 12(4), 333–356. <https://doi.org/10.1080/17437199.2018.1488602>
- Grove, N. S. (2015). The cartographic ambiguities of HarassMap: Crowdmapping security and sexual violence in Egypt. *Security Dialogue*, 46(4), 345–364. <https://doi.org/10.1177/0967010615583039>
- Gupta, A., Hathwar, D., & Vijayakumar, A. (2020). Introduction to AI chatbots. *International Journal of Engineering Research and Technology*, 9(7), 255–258.
- Henry, N. (2024). 'It wasn't worth the pain to me to pursue it': Justice for Australian victim-survivors of image-based sexual abuse. In G. Marco Caletti & K. Summerer (Eds.), *Criminalizing intimate image abuse: A comparative perspective* (pp. 301–309). Oxford University Press.
- Henry, N., Gavey, N., & Johnson, K. (2023). Image-based sexual abuse as a means of coercive control: Victim-survivor experiences. *Violence Against Women*, 29(6-7), 1206–1226.
- Henry, N., McGlynn, C., Flynn, A., Johnson, K., Powell, A., & Scott, A. J. (2020). *Image-based sexual abuse: A study on the causes and consequences of non-consensual nude or sexual imagery*. Routledge.
- Henry, N., Vasil, S., Flynn, A., Kellard, K., & Mortreux, C. (2022). Technology-facilitated domestic violence against immigrant and refugee women: A qualitative study. *Journal of Interpersonal Violence*, 37(13-14), P12634–P12660.
- Herman, J. L. (2015). *Trauma and recovery: The aftermath of violence—from domestic abuse to political terror*. Hachette UK.
- IWDA. (2019). *Promoting women's right to safety and security: A framework to guide IWDA's work*. <https://iwda.org.au/assets/files/IWDA-Safety-and-Security-Framework.pdf>.
- Jewkes, R., & Dartnall, E. (2019). More research is needed on digital technologies in violence against women. *The Lancet Public Health*, 4(6), e270–e271.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- Leurs, K. (2017). Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*, 115(1), 130–154. <https://doi.org/10.1057/s41305-017-0043-1>

- Li, M., & Suh, A. (2022). Anthropomorphism in AI-enabled technology: A literature review. *Electronic Markets*, 32, 2245–2275. <https://doi.org/10.1007/s12525-022-00591-7>
- Littleton, H., Grills, A. E., Kline, K. D., Schoemann, A. M., & Dodd, J. C. (2016). The From Survivor to Thriver Program: RCT of an online therapist-facilitated program for rape-related PTSD. *Journal of Anxiety Disorders*, 43, 41–51. <https://doi.org/10.1016/j.janxdis.2016.07.010>
- Loney-Howes, R., Heydon, G., & O'Neill, T. (2022). Connecting survivors to therapeutic support and criminal justice through informal reporting options: An analysis of sexual violence reports made to a digital reporting tool in Australia. *Current Issues in Criminal Justice*, 34(1), 20–37. <https://doi.org/10.1080/10345329.2021.2004983>
- Luka, M. E., & Millette, M. (2018). (Re)framing big data: Activating situated knowledges and a feminist ethics of care in social media research. *Social Media + Society*, 4(2).
- Mackenzie, I. S. (2024). *Human-computer interaction: An empirical research perspective*. Morgan Kaufmann: Cambridge.
- Maeng, W., & Lee, J. (2022). Designing and evaluating a chatbot for survivors of image-based sexual abuse. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–21).
- Marino, M. C. (2014). The racial formation of chatbots. *CLCWeb: Comparative Literature and Culture*, 16(5). <https://doi.org/10.7771/1481-4374.2560>
- Mcglynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., & Powell, A. (2021). ‘It’s torture for the soul’: The harms of image-based sexual abuse. *Social and Legal Studies*, 30(4), 541–562.
- Munro-Kramer, M. L., Dulin, A. C., & Gaither, C. (2017). What survivors want: Understanding the needs of sexual assault survivors. *Journal of American College Health*, 65(5), 297–305. <https://doi.org/10.1080/07448481.2017.1312409>
- Potter, S. J., Flanagan, M., Seidman, M., Hodges, H., & Stapleton, J. G. (2019). Developing and piloting videogames to increase college and university students’ awareness and efficacy of the bystander role in incidents of sexual violence. *Games for Health Journal*, 8(1), 24–34. <https://doi.org/10.1089/g4h.2017.0172>
- Rackley, E., Mcglynn, C., Johnson, K., Henry, N., Gavey, N., Flynn, A., & Powell, A. (2021). Seeking justice and redress for victim-survivors of image-based sexual abuse. *Feminist Legal Studies*, 29(3), 293–322.
- Raghuram, P. (2021). Race and feminist care ethics: Intersectionality as method. In H. Mahmoudi, A. Brysk, & K. Seaman (Eds.), *The changing ethos of human rights* (pp. 66–92). Edward Elgar Publishing.
- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630. <https://doi.org/10.1016/j.ijhcs.2021.102630>
- Schouten, M. J., Christ, C., Dekker, J. J., Riper, H., Goudriaan, A. E., & Blankers, M. (2022). Digital interventions for people with co-occurring depression and problematic alcohol use: A systematic review and meta-analysis. *Alcohol and Alcoholism*, 57(1), 113–124. <https://doi.org/10.1093/alcalc/agaa147>
- Shelby, R., Harb, J. I., & Henne, K. E. (2021). Whiteness in and through data protection: An intersectional approach to anti-violence apps and #MeToo bots. *Internet Policy Review*, 10(4). <https://doi.org/10.14763/2021.4.1589>
- Strengers, Y., & Kennedy, J. (2021). *The smart wife: Why Siri, Alexa, and other smart home devices need a feminist reboot*. MIT Press.
- Strohmayr, A., Bellini, R., & Slupska, J. (2022). Safety as a grand challenge in pervasive computing: Using feminist epistemologies to shift the paradigm from security to safety. *IEEE Pervasive Computing*, 21(3), 61–69. <https://doi.org/10.1109/MPRV.2022.3182222>
- Toupin, S., & Couture, S. (2020). Feminist chatbots as part of the feminist toolbox. *Feminist Media Studies*, 20(5), 737–740. <https://doi.org/10.1080/14680777.2020.1783802>
- Ullman, S., & Filipas, H. (2001). Correlates of formal and informal support seeking in sexual assault victims. *Journal of Interpersonal Violence*, 16(10), 1028–1047. <https://doi.org/10.1177/088626001016010004>

- UNICEF. (2022). *Safer chatbots: implementation guide*. <https://www.unicef.org/media/114681/file/Safer-Chatbots-Implementation-Guide-2022.pdf>.
- Vorsino, Z. (2021). Chatbots, gender, and race on Web 2.0 platforms: Tay.AI as monstrous femininity and abject whiteness. *Signs: Journal of Women in Culture and Society*, 47(1), 105–127. <https://doi.org/10.1086/715227>
- Wilson-Barnao, C., Bevan, A., & Lincoln, R. (2021). Women's bodies and the evolution of anti-rape technologies: From the hoop skirt to the smart frock. *Body & Society*, 27(4), 30–54. <https://doi.org/10.1177/1357034X211058782>
- Zogaj, A., Mähner, P. M., Yang, L., & Tscheulin, D. K. (2023). It's a match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research*, 155, 113412. <https://doi.org/10.1016/j.jbusres.2022.113412>