

## **Chatbots and the Complexities of Delivering AI-enabled Support to Survivors of Gender-Based Violence**

**By:** Hera Hussain<sup>1</sup>, Nadine Krish Spencer<sup>2</sup>

<sup>1</sup>Founder and CEO, Chayn

<sup>2</sup>Head of Product and Experience, Chayn

**Keywords:** Chatbots, AI, Online Safety, Gender-Based Violence

### **Abstract**

Chayn, a nonprofit tech organisation dedicated to supporting survivors of gender-based violence and providing them with accessible resources online, decided to take down its chatbot in 2020 after a 3-year pilot. Chayn's experience with culturally aware chatbots highlights valuable insights into the challenges and complexities of delivering AI-enabled support to survivors of gender-based violence.



### **Introduction**

Chayn is a global non-profit that creates digital, multilingual resources to support the healing of survivors of gender-based violence. Our focus is empowering women and other marginalised genders who have experienced domestic, sexual or tech-based abuse. Every decision we make – and every resource we create – has lived experience at its core. Chayn remains one of the only organisations in the world that supports survivors of gender-based abuse across borders - for free. Our resources and services are open to every survivor, irrespective of who they are and where they are based. We put support at the fingertips of all.

So far, we've helped over 530,000 survivors around the world. And – with a mission to make technology a tool for health, not harm – we reach more every day. Our goal is to reach 1,000,000 by 2025.

As a digital native organisation, we are always looking to harness new technologies to support survivors. We are often caught in the challenging space of thinking about how to best use technology to support healing from abuse that is facilitated by technology itself. From the mid-2010s onward, chatbots saw a notable rise in their utilisation and popularity within the humanitarian sector. Despite this increased adoption, comprehensive data regarding the overall achievements, potential pitfalls, and trade-offs associated with this automation remains limited. In 2017, we decided to develop our own chatbot - Little Window. Our remit was very clear: It was set up to be a search assistant. A way to drastically reduce the time women take to search for information, which, in many cases, can save lives. But by 2020, we decided to take it down; this decision was rooted in our commitment to cultural sensitivity and deciding we did not have the right resourcing to keep improving the service.

Here are some lessons learned:

### **1. Human language is messy**

Survivors are likely to not use formal terms to describe their experience. The layers of emotions and social conditioning around how to talk about abuse and neglect make the way people ask for help or explain their situation variable and convoluted. Most people do not type 'I am facing domestic abuse' on chatbots or Google. They are likely to type things like, 'Why does my husband hurt me?', 'How can I make my mother-in-law approve of me?', or 'Excuses to stop bf forcing me to sleep with him.'

For our chatbots to be smart enough to know that when people talk about 'feeling down,' they may mean depressed or anxious, and 'make my husband love me' may also mean they are experiencing abuse or neglect is a very demanding task. Even if, with the advent of ChatGPT, we may look like we can understand and mimic human speech better than ever before, being able to hold conversations on mental health and potentially crisis-related scenarios is a very risky business. Take the recent example of the US National Eating Disorder Association (NEDA), which took their Eating Disorders Support line offline due to the chatbot advising users to take the kind of advice "that led to the development of my eating disorder," according to an influencer who had experience of disordered eating.

Even putting a multilingual capability aside, textual English differs across cultural contexts. Though most words used in English were found to be consistent across regions, colloquial and 'text speak' English presented differently and would be missed by chatbot Natural Language Processing programmes that were only set up to recognise textbook English. For instance, 'nikkahofied' in Pakistan is a way of referring to being married ('Nikkah' being the Islamic marriage and 'fied' being the English verb suffix) or 'pressuring into marriage' is the same as 'forced marriage'.

In addition, irrespective of native language, we found survivors switched to English to describe sexual assault, rape, marital rape, and topics of consent. This was especially prominent among Arabic, Hindi, Urdu, and Bangla speakers. In our discussions, we attribute this to patriarchal cultures suppressing conversation on taboo subjects such as sex. Therefore, survivors often do not know the equivalent of these terminologies in their native languages and revert to English, where these terms are popularised through a globalised media. We also noted how, in Bangla and Urdu, terms about happy and consensual interactions were easier for survivors to say rather than trauma-laden terms.

## **2. Use trauma-informed design**

At every stage of designing Little Window, we employed our trauma-informed design principles of Safety, Equity, Plurality, Agency, Accountability, Privacy, Power-sharing, and Hope. You can read more about these in our white paper from July 2023. A few examples are:

### **Safety**

We must make brave and bold choices that prioritise the physical and emotional safety of users. This becomes critical when designing for an audience that has been denied this at many other points in their lives. Whether it is the interface of your platform or the service blueprint, safety by design should always be the starting point.

In chatbots, this can be reflected in technology decisions such as not saving information on the user's end as they might be using a shared device and giving users options to replace information within the chat in case they are concerned about others finding it. In the user experience, it looks like the opening message is empathetic and warm, thanking survivors for trusting the service. This is a potentially traumatic experience for a user deciding to disclose and find more information, so this non-judgemental and supportive tone should be consistent from the first instance of contact.

### **Plurality**

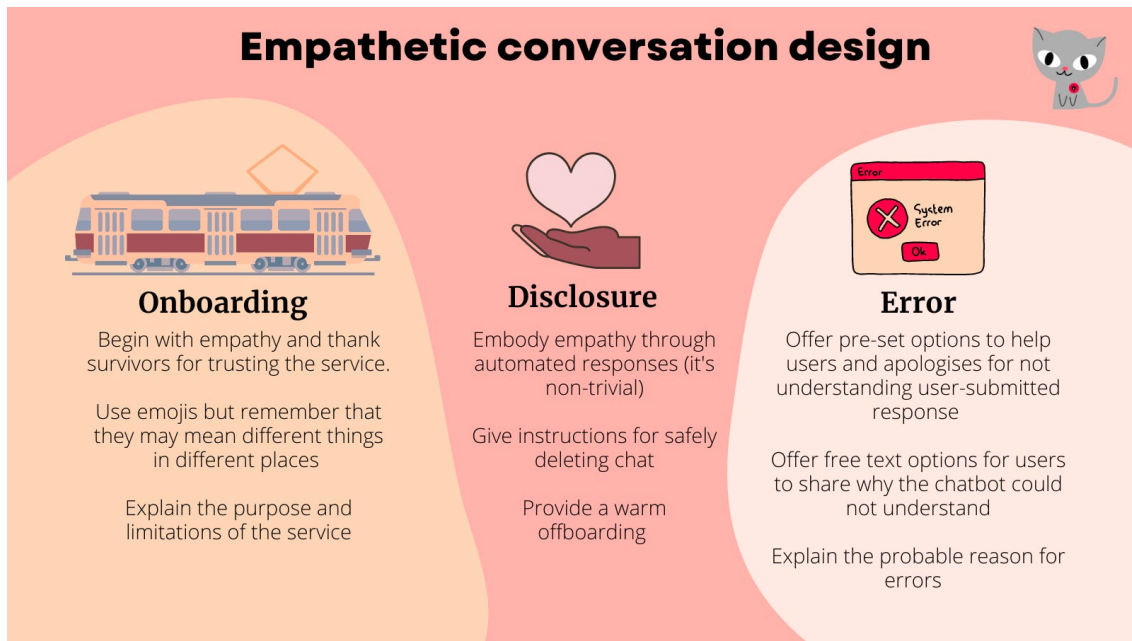
To do justice to the complexity of human experiences, we need to suspend assumptions about what users want or need and thus account for selection and confirmation bias. There is no single-issue human, and therefore, all of our interventions need to be designed with that in mind. Even if our services focus on one aspect, we need to signpost to other needs to provide the best relief.

In chatbots, this means not assuming someone's language based on where they are coming from, as they could be from a minority or migrant group. It can mean allowing people to choose multiple languages and locations for the resources they wish to access. And it can also mean being sensitive in our communication. For example, the use of emojis should be carefully considered. A handshake emoji might mean a friendly hello in some places but may appear more business-like when the aim should be for this to be a supportive, safe experience. If emojis are going to be used, they should be more widely relatable and supportive - at Chayn, we are big fans of the heart emoji.

### **Accountability**

For Chayn, this also means practicing the values of openness and collaboration with our partners and users alike, banishing the spectacle of perfect performance, and embracing the risk of failure that comes with holding uncertainty as dear as knowledge. It's a commitment to be transparent with users about the limitations of our work.

In chatbots, this means being transparent about the very fact that it's a chatbot and including space for users to suggest new content and features and give feedback on their experience.



### 3. Transparency is essential

In keeping with our trauma-informed design principles, Little Window was upfront about being a chatbot (its mascot was a cat). Unlike customer service chatbots, which may try to pass themselves off as real people, Little Window was there only to provide signposting and resources and was transparent about this purpose. There was little natural language processing, limited opportunities for free text, which reduced the likelihood of a high error rate, and clear indications through language and design that the service was not managed by a human.

We spent a great deal of time planning the conversation flow to ensure it was designed sensitively and appropriately. Advising people facing abuse on how to seek safety is a big responsibility, and we wanted the accuracy of our scripts to reflect this. For example, at the beginning of the conversation, if the chatbot did not understand the question that the user asked, it made it clear that it could only answer questions related to a limited set of topics (i.e., only domestic abuse, divorce, and asylum). Giving the user options to choose from immediately was more supportive than risking further frustration due to continued misunderstandings.

### 4. Be brave in constantly analysing your balance between technology and humanity

Despite all of these precautions and the best of intentions—including a careful design process that involved survivors at every stage—a review of the chat logs clearly showed that people were using it in the exact ways we wanted to avoid: as a crisis service. This resulted in an alarmingly high error rate. It was clear from the chat logs that people ignored or didn't pay attention to our opening message that indicated Little Window wasn't staffed by humans. Something that perhaps indicates just how desperate they were to talk to someone. It was painfully obvious how frustrating the error experience was for users. Users may be experiencing panic or desolation, and this interaction is likely to make them feel worse.

Our decision to take down its chatbot underscores the delicate balance between technology and human interaction. We simply didn't have the resources to continually improve the

chatbot with what we were seeing, and decided it wouldn't have been fair to maintain it as it was, given everything we'd learned.

To those who want to launch their own domestic violence support chatbot, our general advice remains: Don't do it. And if you're going to do it, do it for the right reasons. Chatbots offer great ways of helping people navigate structured content. They remove pressure off of staff (and volunteers) and can help direct traffic to the right organisations. What it's not good for is providing conversational support to humans in distress. You cannot replace human care. And even if we could, with generative AI, we just — should — not. The risks are too high if we get it wrong. Chayn now offers a 1-to-1 chat with a trained team of humans via our platform, Bloom, and we're confident that this free, borderless service in 14 languages is precisely the direction we want to continue in.

## References

Hussain, Hera, and Winfield, Jenny. 2023. *Chayn's trauma-informed design principles*.

[https://assets.website-files.com/60fdc9111506063bb9fe8e49/64b081438e3221d7ffc92b12\\_Trauma-informed%20design\\_%20the%20whitepaper%20by%20Chayn.pdf](https://assets.website-files.com/60fdc9111506063bb9fe8e49/64b081438e3221d7ffc92b12_Trauma-informed%20design_%20the%20whitepaper%20by%20Chayn.pdf)

Hussain, Hera. June 2023. *Why Chayn took down its chatbot in 2020 and what we've learned about culturally aware chatbots*. <https://blog.chayn.co/why-chayn-took-down-its-chatbot-in-2020-and-what-weve-learned-about-culturally-aware-chatbots-a9587cf80df8>

Kazansky, Becky, Olivia Johnson, Bárbara Paes, Helen Kilbey, and The Engine Room. 2023. "Chatbots in humanitarian contexts: Learning from practitioner experiences."

[https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623\\_CEA\\_Chatbots.pdf](https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623_CEA_Chatbots.pdf)

Nguyen, Britney. 2023. "Helpline workers for the National Eating Disorder Association say they are being replaced by AI." *Business Insider*, May 26, 2023.

<https://www.businessinsider.com/eating-disorders-nonprofit-reportedly-fired-humans-offer-ai-chatbot-2023-5?r=US&IR=T>