

Safer, **culturally-aware** chatbots for addressing gender-based violence

By Chayn

April 2021

A version of this report was commissioned by UNICEF's "Safer chatbots for gender-based violence" programme. Their final report can be found [here](#).

Table of Contents



1. Executive Summary
2. Background
3. Methodology
4. Key learnings
5. High-risk keywords
6. Opportunities for further research
7. Empathetic conversation design
8. Patterns of use from Little Window
9. Trauma-informed design principles for chatbots
10. Future of chatbots in gender-based violence

Executive Summary

Chatbots are an exciting technology and advancements in natural language processing, the applications of it in the human rights space - including gender-based violence are increasing.

Chayn's approach in the chatbots has been cautious which is why when we developed Little Window in 2017, our remit was very clear: little chat, more signposting. It was set up to be a search assistant. There was little natural language processing, limited opportunities for free text which reduced the likelihood of high error rate and clear indication through

language and design that the service was not manned by a human. We used a cat as a mascot for the service.

Despite these precautions, a review of the chat logs clearly indicates that people were using it in ways we wanted to avoid - as a crisis service. This resulted in an alarmingly high error rate and it was clear from the chat log that people, whether that was because people are desperate to talk to a human, they ignored or didn't pay attention to our start up message that indicated it wasn't staffed by humans.

Background

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?', 'why am I always down?' or 'excuses to stop bf forcing me to sleep with him'. We want these chatbots to be smart enough to know that when someone says 'how can I make my husband love me?', it's likely that this person is talking about domestic abuse and is experiencing coercive control. [Chayn's](#) been helping survivors of abuse across the world navigate the online and offline world by providing information, resources and links to the frontline support in multiple languages.

In 2017, Chayn was commissioned to do a research project ([Tech vs Abuse](#)) by Comic Relief, alongside Snook and Safelives, on how domestic abuse survivors in the UK, as well as the organisations supporting them, are using the web to search for information and support. We found that often women had very short windows of opportunity when their partners were not around to find support online, but that the online world can be a 'rabbit hole' of information. This is especially true for women who have multiple factors that affect their circumstances. For instance, if you're a Pakistani immigrant with low English skills stuck in an abusive relationship in the UK, not only do you need information related to laws and organisations in the UK, but also Pakistan where you may have to return. Similarly, there are many specialist services that you might want to search for, such as charities that cater to women with children.

Here are two quotes from the women we interviewed during Tech vs Abuse research that stood out:

Bad websites make me have a very bad experience. You go round in circles finding information and that can become very frustrating. Both user journey and not finding what you're looking for. Realising that this is just a website. This is the worst part for me.

It's only a website with a number that doesn't work. I had nasty experiences and they make you feel desperate at the point.

It took me 15 clicks to find the information on a local refuge. If you only have 5 minutes alone, that's at least 10 clicks too many.

Little Window is a smart search bot that directs women to the information they are looking for as quickly as possible. Think of it like google search on turbo for all of Chayn's resources and those provided by our friends too. We wanted to drastically reduce the time women take to search for information, which in many cases can save lives.

It's imperative that we expand our knowledge of what the common terms or expressions can mean, and what survivors are expecting to see as a response. This will allow chatbot developers to connect relevant phrases to resources related to gender-based violence. Collecting this information is critical but so is doing it with an intersectional lens. This is where our expertise comes in.

Methodology

We used two sources of data to conduct a mixed-method research into the efficacy of chatbots in being safe, helpful and culturally aware.

Data sets

1. **Little Window:** We reviewed data from our chatbot Little Window, which included more than 800 lines of user submitted data across 4833 conversations between March 2018 and April 2020.
2. **Social media:** We looked at popular forums and groups on social media and comment sections of YouTube videos where survivors of abuse discussed their stories.

We formed a multi-cultural and multilingual research team of survivors and allies to look through both data sets as well as combining it with our own experience of supporting survivors from the friends, family and wider networks of volunteers and staff.

Languages

Little Window was available to Chayn's visitors in English and French, but people did input other languages in free text fields, which we have included in our analysis.

For research into social media forums, our team looked at Arabic, Bangla, English, French, Hindi, Spanish and Urdu. This was purely based on the composition of our research team.

Process

This was a short research project. Therefore, a research methodology was developed that allowed our team to complete the project within a few days.

Collection

For research on Little Window, data was extracted from Chayn's database. For social media sites, team members were asked to look at groups where women ask advice and share experiences. We considered the ethics of using social media content, which though public, had not been shared with the knowledge that it may be used for research. On the balance of the benefit of doing a one-of-a-kind research project that will help survivors of abuse, we decided it was justified that we used it. None of the original posts were stored.

Individual analysis

Each member of the research team individually identified the sentences or words that are likely to be used when attempting to share a story or ask a question related to gender-based violence through social media sites and Little Window.

Comparative analysis

We worked together in pairs and in one large group to compare and contrast each others' work, identifying gaps, commonalities and differences.

Key learnings



The following learnings should support the development of more inclusive and accurate chatbots.

Phonetic spellings and abbreviations may cause errors.

Phonetic spellings and abbreviations are **common**, e.g. 'plz', 'coz', 'thru' or 'three', '2 wks', 'wat 2 do', 'jerny', 'wunt', 'dunno', and 'an'. We found these to be the most common in native English speakers who had not gone through advanced schooling or are native Welsh speakers, or spoke rudimentary English as a second language. These could cause the chatbot to fail to react to a disclosure or respond with an **error message**. This could also

be the case for **language barriers**, e.g. **'how to don't kill myself'** or **'my husband is treading me'**.

Text 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 NLP 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'.

Survivors are likely to not use formal terms to describe their experience.

Terms for GBV are not used directly in a lot of disclosures, and are instead more **broad disclosures**, e.g. **'he is making me hand over money'** or **'he stops me seeing friends'**. Though we know these examples would tend to fall under financial abuse or social control, the chatbot may not react to these disclosures, so these should be categorised as common expressions under different forms of abuse.

The specificity of disclosures varies a lot.

Disclosures are in some cases **not specific enough** and may cause errors in categorising, e.g. **'I need help'**, **'I don't know what to do'**, or **'victim resources'**. These disclosures could relate to a number of issues faced by the user and the chatbot may not respond to the current one.

There are however some disclosures that have specific points that need **direct advice**, e.g. **'I am on a spouse dependent Visa and my visa will expire on 17 Jan 2021 can I apply my own Visa as I am going through domestic violence mental abuse'**. With these disclosures there are multiple points of support potentially needed.

Some users share statements rather than questions.

A lot of disclosures are statements rather than questions, or are in simple terms, e.g. **'advice on martial rape'** or **'domestic abuse'**. This may lead to the chatbot responding with a lot of information that may not be relevant to the user and their needs for support not met.

Country-specific disclosures need to be considered.

There are some disclosures that are specific to certain countries, e.g. **'I live in New York and facing dv from my spouse'**, **'how to get divorce in France?'** or **'where is the nearest refuge in London'**. The resources for these will of course vary depending on the country and the laws there, which the chatbot would ideally be able to provide.

There are interlacing issues so our support needs to be that too.

For many of our users and those that we researched, **many were not solely concerned with the abuse**. There was an additional concern about migration, and legal rights. Migration related issues were the strongest when we came across migrants e.g Bangladeshi women living in UAE, Syrian and Pakistani women living in the UK and USA.

Consent featured in most conversations.

Not surprisingly, **themes of consent** were present in most user queries and comments across all languages, especially English. We also noticed a high amount of user expressions related to **emotional abuse**. Both should feature heavily in the resources compiled by chatbots and the narrative within the chatbot should be written with this in mind.

In communities where talking about sex and abuse is taboo, many used the English language to talk about sexual assault and consent.

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.

Spanish speakers are more likely to use direct and precise terms to describe their situations.

Survivors whose mother language is Spanish were more likely to use words like 'divorce', 'domestic abuse', 'child abuse' and 'rape' to describe their situations. This is derived from observing forums and social media where domestic abuse and violence against women was discussed by survivors. We also noted that in many public forums, there was a lot of **public shaming** and **victim-blaming**, sometimes done by the survivors themselves to each other. This form of internalised misogyny was also found in other language and cultural groups.

French speakers may be more likely to use more obtuse terms to describe their situations.

Language used to describe abuse, especially sexual harassment, was often more **obtuse and crude**. Some words that we use in English such as 'slap' or 'yell', have a different significance in French, where different things are more socially and culturally acceptable. We noticed whilst researching that resources for survivors vary a lot, and are quite different from resources available in the UK, for example. The language used often lacked the care and gentleness that is important when addressing survivors, which may be

indicative of lack of knowledge or interest in this provision, as well as socially accepted misogyny in language.

High-risk keywords



From our research, we were able to highlight the high-risk keywords, **triggers** and **'red-flags'** that users may disclose and that chatbots would need to be alert to for escalation, with regards to gender-based violence. We worked with the framework of some words being indicative of high-risk of their own (pills, die) and others needing further confirmation.

Identifying trigger words and red-flags.

From our research we needed to determine which of the high-risk keywords would be single triggers and need immediate escalation and which needed further 'red-flags' for the chatbot to pick up for this escalation.

We began by mapping the **single trigger words**, such as **'suicide'**, **'unsafe'** or **'trafficked'**, which alone should alert the chatbot and trigger a triage sequence. These are words that are very loaded and serious, and ones that the chatbot should identify and escalate in the first instance. In our research we highlighted certain other trigger words that should be escalated immediately in the various languages, which will be noted further on.

We then reviewed trigger words that would need to be checked alongside other 'red-flag' words for escalation. If the conversation contains a red-flag word, there should be an escalation. If not, then there is a chance it is a false positive with the initial trigger word. This is keeping in line with our findings regarding language or the end user themselves - for example, **'forced'** would be a trigger word for the chatbot to identify, but for escalation would need more information, as the word can be interpreted differently. Red-flag words such as **'marriage'**, **'sex'**, **'abortion'** or **'touching'** would need to be monitored through the chat for a potential match and appropriate escalation.

Forced → trigger word

Forced + abortion → escalation

Complex disclosures.

There are many words that the chatbot could or should be alert to through a disclosure, though there were some identified which were much more complex. Feelings are commonly disclosed e.g. **'I feel overwhelmed'**, **'I am afraid'** or **'I am sad'** that the chatbot should be able to recognise and respond to accurately. Words such as **'depressed'**, **'slaps'**

or **'broken'** may be subject to the user's vernacular and how they use that word. For example, 'broken' may not necessarily mean that they have broken bones, or even if it does, it may not mean it is as a result of gender-based violence and that the user needs support. We also needed to consider language, similar to the points raised in our key learnings. Obtuse descriptions or words for abuse in some languages may make the escalation process difficult, and so the red-flags have been identified to aid the chatbot in being alert to a user disclosure.

This was not English specific and we also noted these trigger words and red-flags for French, Spanish, Urdu and Arabic. We also considered the use of English across the Middle East, Latin America and South Asia.

Ensuring a cleaner dataset through words, not phrases.

We included words and stripped function words out of phrases where possible to produce a cleaner dataset. For instance, 'I am feeling depressed' becomes 'depressed' because it is this word that causes us concern, rather than the preceding words.

There are some phrases, however, that we believe should be complete triggers, such as **'I am not safe'**. This is because this as a complete phrase is a clear indication of the user needing support, and also could encapsulate various instances of support needed, which would then be highlighted from further user disclosure and second red-flags. Other examples noted in the research are **'makes me have sex'** and **'honour killing'**, which we noted should be a complete trigger phrase that the chatbot should be alert to.

		Trigger	Red-flags
	Escalate if trigger is present	<i>This word or phrase should trigger a triage sequence. If there is nothing in the "red-flags" column, then it should be escalated immediately. If it has a list, then the conversation should be checked against it. If there is a match, it should be escalated.</i>	<i>If empty, "Trigger" should be escalated. If there is a list in here, please check for a match and then triage. If there is no match, then there is a chance it is a false positive.</i>
English	x	abuse; abusive	
English	x	sexually assaulted; sexual assault	
English	x	suicide	
English	x	kill	
English	x	gun	
English	x	knife	
English	x	die	
English	x	paedophile	
English	x	rape; raped; raping	
English	x	starve	

Given the nature of conversational text input, we treated both the 'trigger' and 'red-flags' as lists.

Sophisticated language modules would be able to interpret different tenses for the same word such as 'hurt, hurting' but for the sake of clarity, we have included these in both lists.

Urgency in disclosures.

We noted that there are some words that could alert the chatbot when entered with other terms, but it would be difficult without advanced knowledge for the chatbot to pick up on this urgency. The word '**leave**' was a clear example of this. A user disclosing 'I need to leave now I am in danger' would need much more urgent support and referral than if they entered 'how do I plan to leave'.

There is always the potential for grey areas in reporting certain words.

Through this work we identified this grey area in a number of disclosures, where the risk of false positives and the risk of not escalating were equally high. Words such as '**choke**', which we deliberated on, would either be viewed as a single trigger word, or would need second red-flag words to match. We decided the latter would provide more clarity, as the user might be using that word for other reasons than to disclose abuse, such as sexual health questions. There are a number of words like this in the research, and we had to discuss each word and determine whether the user would disclose them for purely GBV related issues, or for something else.

Considering the user themselves through the disclosure process.

For words such as '**slaps**' or '**nasty**', these could be identified as less high-risk words, but if we consider that a child may be disclosing this, they take on new meaning. This is where the categorisation of second red-flag words would help with the chatbot through the disclosure process, as it would be able to alert a human for further support when certain words are entered within the same chat.

The 'trigger' terms that should trigger an escalation response without a need for a double confirmation are as follows.

ARABIC: مسدس ، لجوء ، انتحار ، اغتصاب ، مسلح ، اسلحة ، سكين ، شبرية ، اعتداء جنسي ، اعتداء جسدي ، ضرب

ENGLISH: *abuse; abusive; sexually assaulted; sexual assault suicide kill gun knife die paedophile rape; raped; raping starve acid sexual violence trafficking; trafficked danger kidnap child porn i*

am not safe self harm violent unsafe blackmailing; blackmail hack makes me have sex pill;pills honour killing abduct stalk; stalked; stalking.

FRENCH: *automutilation; aggression; suivi; inceste; trafic sexue;; trafic de drogue; trafic d'êtres humains; viol; mutilations génitales.*

HINDI and URDU: *zeher; khudkushi; jaan lay lungi; tashadud; gula daba; gala daba.*

SPANISH: *me mata; mató; pega; azota; amenaza; golpea; tráfico; hackeo; matoneo; me acosa; acosarme; violencia; agresiones sexuales; maltrato; golpeada; golpea; me pegaba; me pega; me violaba; me viola; suicidio; extorsionando; me haga daño; morboso; tocaciones; pornografía; infantil; acosador; manipulador; delito sexual.*

Finally, **a note of caution** from our team. While this analysis is useful to make bots smarter, it's important to state that many of these categorisations of what words are red flags by themselves, and what need a second confirmation, are highly subjective and not precise. We recommend these lists are treated as working drafts and **updated regularly**.

Opportunities for further research

We did not have the capacity to perform further analysis on the following lines of enquiry at this time but recommend that these are investigated as part of the wider research.

Spelling errors

In future, there is scope to recruit 'smarter' language recognition systems that will be more robust to spelling errors to support the error experience, especially when dealing with traumatic situations and where users may be incapable of correcting errors themselves or where phonetic use will be more common.

Structure of phrases

The system should be smart at picking up on the structure of phrases. Most of our dataset showed the following structure: 'my [person] is [verb]'. It was nearly always a description of abuse.

Use of pronouns in disclosures

It would be interesting to see how people can be redirected to resources based on the kind of pronouns they have used: reflexive as in '**Hurting myself/harming myself**', non-reflexive as in '**my husband is hurting/harming me**'. The former could be directed to

resources about mental health including organisations that provide hands-on support. The latter could go towards gender-based violence resources and organisations.

Subtle differences in use of the same words

We also noticed that there are subtle differences which can be picked up from the phrases. When someone uses the word '**leave**', it's what comes after the verb (a person, a place or nothing) that determines what they are looking for, for example '**leave my partner**' (divorce law) is different to '**leave**' in other settings (e.g., '**How do I leave?**', 'My partner is verbally abusive. **I want to know how to leave**').

At the moment, most chatbots (including our own), would triage this to divorce law, but on examination it is clear that is only relevant for the first example. Further work is needed in this area to suitably address disclosures and provide the most accurate resources and support.

Conjunction

We noted in our key learnings that users typically tend to express their feelings in disclosures. Conjunction words such as 'and', 'or', and 'but' all have different nuances and connotations, but they all help to build a meaningful relationship within a sentence that can alert to something new. It would be extremely worthwhile to have a more specific look into different types of conjunction (e.g. '**how to**'/'**how do I**', '**I need**'/'**I am**'/'**I think I**'). This could allow for better reactions and responses from the chatbot to better support these instances, as well as in the majority of disclosures, especially if they are long disclosures or phrases, and could better streamline support and resources available.

Length of utterance and urgency in disclosure

Another interesting area of exploration is looking at the length of utterance and whether that can determine which request is more urgent and thus must be prioritised in the ticketing or escalation system. For instance, 'please **help**', 'Looking for support to **help** a friend in a pandemic' and '**help** with legal rights' all use the word **help** but it would be reasonable to assume that the first one might be more urgently triaged than the second two simply based on the information we have. This is worth testing because it might be that there is no correlation, or that it is inaccurate for people for whom English is their second language.

Improving the error experience based on larger contextual data sets

In terms of improving the **error experience**, there are some options. Though you would need a lot of data to do so, training bots to ask smarter questions when they don't recognise something would help to vastly improve the user experience, including some of the issues we mention in our Little Window interactions. This would be simpler for typos in disclosures, and more complex for vague expressions, but would still be possible.

To do this, two things would need to be done. A bot would need to be trained to recognise certain situations based on input, as we mentioned when discussing structure of phrases. Then, there would need to be a restructured questioning back to the user, followed by a confirmatory stage. An example would be the input being 'my father' and the response being 'it sounds like a member of your family may be abusing you. Is that correct?'

This would be invaluable in disclosures such as the ones we have included in the appendix. Having the nuances of user inputs and disclosures that are based on larger contextual data sets included would allow for better support and resources, and would especially help in clarifying instances of high-risk and escalation.

Empathetic conversation design



The UN "Safer bots" project was also looking into how to create supportive conversation scripts. We reviewed some bots, and came up with a list of considerations for empathetic conversation design - something Chayn focuses on across all products, including Little Window.

The onboarding experience

A warm, open and non judgemental way of asking someone about what they may be going through without using words like 'Rape' and 'domestic abuse' or 'forced marriage'?

The onboarding experience should always include a **detection** and **acknowledgment** of possible disclosure, with an option for the user to decline referral information. Upon a review of prototype chatbot messages, we have noted the following ways for improved onboarding.

Conversations should begin with empathy and thanking survivors for trusting the service.

The opening messages in a disclosure situation should always be **empathetic** and **warm**. 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. With this, there should be a thank you message given to the user in the first instance of sharing with the chatbot. The response and interaction should be warm and encouraging to ensure the user feels safe in disclosing and that they will not only find help, but support through the disclosure process.

Some emojis don't translate across cultures

The use of emojis should be carefully considered. For instance, 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 - we suggested using the heart emoji.

The disclosure experience

How the chatbot reacts and responds if an end user discloses an experience of GBV. By categorising common expressions, we're making it easy for chatbot to know what 'theme' it should be taking them to.

We've noted our key learnings for how the chatbot should ideally react and respond to a disclosure, especially if an end user is disclosing an experience of GBV. Once this has been done, there should be an empathetic, discrete and rapid response to this disclosure, as well as referral information provided, before the user is handed back to the main user report experience.

Embodying empathy through automated responses can be hard but must be done

Users should feel supported throughout this entire experience and should have referrals as soon as they are needed. There may be challenges to this with regards to language errors as we noted in our key learnings, but when the disclosure is identified, the warm tone should still be present.

Suggested responses could be:

- 'I'm sorry to hear what you are going through. Thank you for sharing this with me (the chatbot). It can be hard to reach out.'
- 'I've got some resources that may help that are safe and understanding. We hope you get the help you need - when you are ready. 💕'

Safely deleting chat messages

Where possible, providing advice on deleting messages that should always be included throughout this process. User safety is the most important part of the disclosure process, and should be assured at all times. These steps should be either as clear as possible, or linked out for the ease of the end user to relate back to. This could be for deleting the messages, or replacing them with a stock message related to something completely different, or to the first message in the chat again. With this though, we may need to consider a way for end users to come back to the chat, to ensure they are not being re-traumatised in disclosing multiple times, for example if they are worried their partner will look at their phone and see the messages, but still need the support. This should also be considered.

Warm offboarding for users

There should be a warmer offboarding for the user, whether they choose to replace the messages or not. This message will serve to ensure the experience is as supportive as possible for the user and that they leave the experience with kind affirmations from the chatbot. We suggest the following to illustrate our point:

'When things happen out of our control and do us harm, we can feel like there won't be a better tomorrow but that's not true. You are here and deserve to be here. Your life matters to us. Remember that we're still here and you can come back to this chatbot anytime.'

The error experience

How a chatbot reacts and responds if users type something the chatbot can't recognize. We're trying to decrease this error rate.

It's expected that the chatbot could or would not understand many things users say through their disclosures, or incorrectly match user statements to a theme that is not correct or helpful.

We recommend that as an answer to both, the chatbot offers pre-set options and apologises for not understanding things correctly. Also offering free text options for users to tell us why it was wrong and submitting it to the development team allows a feedback loop, though it is important in the short term to give the users an explanation of why such errors occur.

When reviewing the user input data for Little Window, 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. This is a serious limitation to chatbot technology in general, and the more open the chatbot is with the users about its

limitations throughout and providing an 'off ramp' to other services, the better it would be, and the more support can be given to users.

Patterns of use from Little Window



Little Window was available in French and English. It was put offline for maintenance in April 2020. We reviewed data from our database for Little Window engagement, which included more than 264 lines of user submitted data across 4833 conversations between March 2018 and April 2020. Here are our insights, which were both surprising and heartbreaking at times.

Types of interactions

We analysed hundreds of lines of conversations with user input to categorise the kind of interactions we were having. It's worth pointing out that the following are likely only a snapshot of what users actually want as every platform, due to its design and accessibility, will encourage a particular kind of audience. Below, you will find actual statements taken from our chatbot records.

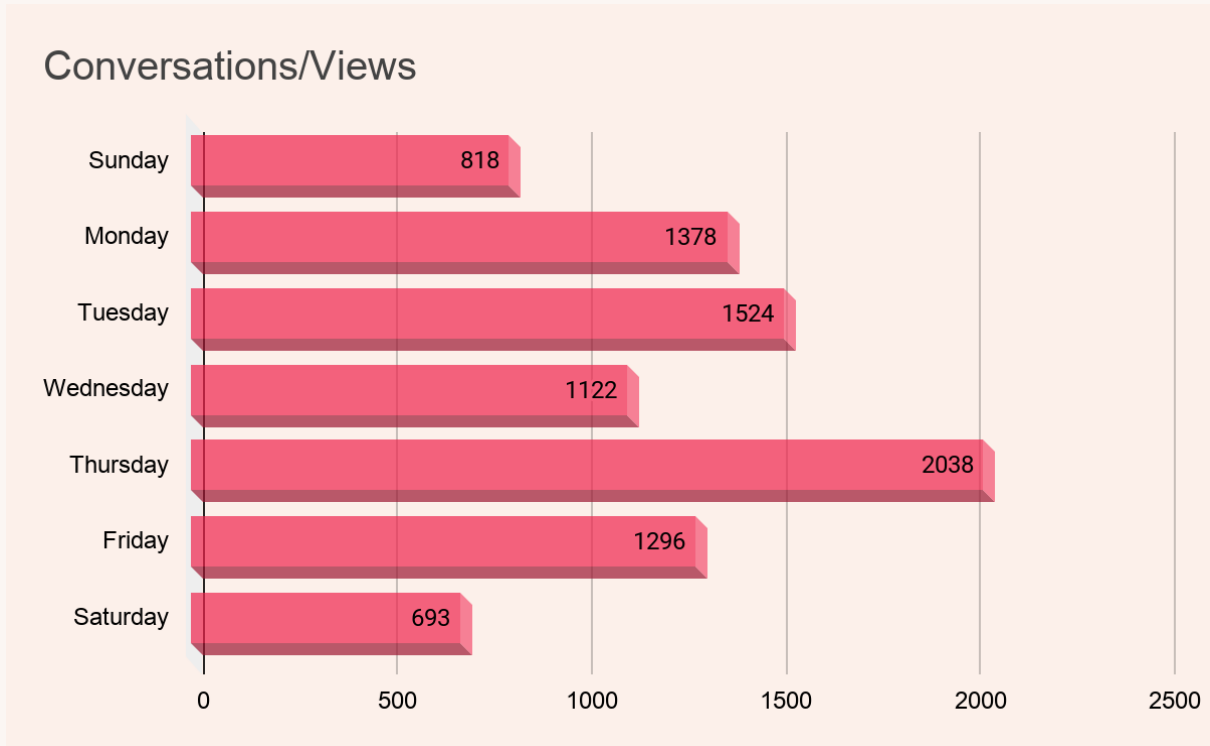
Note: We have not corrected spelling and grammatical errors because we wanted to share the nature of the raw data chatbots are expected to comprehend.

Type	This can sound like...
Validation of experience of abuse (esp. Emotional abuse)	i think i'm depressed but i'm not sure I am unsure about the health of my relationship. I don't know if it is healthy or unhealthy. Can you help me think about this?
Catharsis	I am getting abused by my family every day my husband abandoned me 2 days ago.
Advice on how to get out of abusive situations	My ex boyfriend has been harrassing me since we broke up 10 months ago and I don't know what to do How can I stop my husband swear to me in front of my kids I have problems with my husband at home, I don't know what to do?

Migration	funding for asylum seekers DDV application to apply indefinite leave to remain
Legal advice	I joint own my house but had to flee due to Domestic Violence, I currently stay in supported accommodation, do I have to pay for my housing costs i just want any general advice that can be given I cant afford a lawyer and a final judgent hearing is about the following matter this month on the 22nd Evidence needed for cohersive control
Online safety	maybe you'll know what browser is more safety I am looking for support with tech abuse what do i do when someone is blackmailing me with nudes
Self-harm	How to don't kill myself?
Divorce	Is there any help i can get to survive my divorce?
Looking for direct, one to one support	support group after abuse I need counselling
Asking help for others	My friend is in a difficult situation and I would like to know how can I help I'm concerned about a grieving lady with bipolar whom may be being manipulated my autie in pakistan is suffering domestic abuse, who can she contact for immediate support
Looking for services	what's my nearest domestic abuse service?

Days of the week

We analysed data collected on views and conversations to determine the busiest days with higher rates of conversations. Between March 2020 to April 2021, Thursday was the day with most conversations. The reasons for this are unknown to us, but this information provides us with useful insight into potential user habits and patterns for future insight.



Requested languages

The top countries that use and benefit from Chayn’s other projects are: UK, US, India, Pakistan, Canada, Germany, France, Australia, Italy, Brazil, Belgium, South Africa, Lebanon, UAE, and Mexico.

There is some overlap in languages with these countries but it’s still interesting to see that the languages requested aren’t always the ones we would expect. The most requested languages (sorted by number of requests) were: Arabic, Chinese, Dutch, Spanish, Farsi, Hindi, Italian, Polish, Portuguese, Russian, Swahili, Tamil, Turkish and Urdu.

Languages that were requested once were as follows: Ahoj, Albanian, Amharic, Azerbaijani, Bangla, Bulgarian, French, German, Japanese, Kalabari, Lithuanian, Norwegian, Philippino, Punjabi, Romanian, Shona, Sinhala, Swedish and Zulu.

Trauma-informed design principles for chatbots



At Chayn, we've been creating, testing and experimenting with design principles that create trauma-informed language and services since 2013. The following is a snapshot of our current framework. This might change over time as we like to adapt as we learn more.



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 chat bots, this can look like:

- Not saving information on the user's end as they might be using a shared device
- Providing safety advice before, and during conversation
- Strong digital safeguarding framework
- Giving users options to replace information within the chat, in case they are concerned about others finding this information
- Acknowledging that the abuse isn't their fault and that they deserve to feel happy and respected in their bodies and life
- Conversation design that holds space for users having mixed feelings about their situations e.g whether to report parents for historical abuse or just distancing themselves in the present.
- Soothing offboarding messages and affirmations for users



EQUITY

All of our interventions need to be designed with inclusion and accessibility in mind. Survivors are not a homogenous group and they will not all benefit from the same types of support. We must consider how position, identity, vulnerabilities, experiences, knowledge and skills shape trauma and recovery, and create solutions that leave no one behind.

In chat bots, this can look like:

- Clearly indicating if a human is behind the screen
- Creating a 'soft ramp' of conversation with a user who may have a safeguarding risk rather than abruptly telling them they need to be referred to the appropriate agency right away. This becomes hard to balance with the need to safeguard children but we still need to ensure first, that the users feel they can talk to us and will return to us when they feel vulnerable.
- Signposting to organisations outside the intended geographical area to cater to vulnerable people who may stumble across the bot and migrants in the intended area
- Providing intersectional resources for LGBTQ+ users



PLURALITY

To do justice to the complexity in 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 chat bots, this can look like:

- Not assuming someone's language based on where they are coming from as they could be from a minority or migrant group
- Allowing people to choose multiple languages, and locations they want to access resources for
- If anonymising data to protect personal information, not anglicising (or equivalent majority ethnic/religious group) user names
- Allowing users to communicate their experiences without any assumptions placed before them
- Maintaining and linking to support for multiple areas e.g health, housing, legal



AGENCY

Abuse, inequalities and oppression strip people of agency. We must always make sure we do not use tactics of oppression to ensure we can redistribute power and agency by providing information, community and/or material support.

In chat bots, this can look like:

- Not forcing users down one outcome for their situation e.g only calling the police
- Giving a range of resources for each theme so users can choose what they would like at that moment (as this may change for repeated users)
- Having consideration for language barriers or grammatical differences in disclosures



ACCOUNTABILITY

For Chayn, this also means practising the values of openness and collaboration with our partners and users alike, banishing the spectacle of perfection 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 chat bots, this can be:

- Explaining reasons for the limitations of the chatbot upfront like in the error experience
- Including space for users to suggest new content, features and give feedback on their experience



HOPE

The people coming to these chatbots know something of the pain of trauma. They do not need to be reminded of it by harsh words and sad pictures. It's scary and brave to reach out for help and our virtual spaces need to feel like an oasis for users. Abuse can leave us feeling like no one cares about us, and at times, that we don't even care about ourselves. Empathetic, warm, soothing and minimally-designed interfaces and narrative should feel like a virtual hug, motivating people to both ask and embrace help. It should validate their experience.

In chat bots, this can be:

- An aesthetic that feels more like a cafe, and less like a police station
- Words that feel more like flicking a magazine, and less like reading a legal document
- Empathetic messages and a good balance of emoji usage

PRIVACY

A survivor's personal information including their trauma story - such as data, images, videos, or statements - must be kept secure and not disclosed, unless the survivor decides to do so. At the same time, we should remove unnecessary obstacles from users getting to the information and help they require.

In chat bots, this can be:

- Explaining GDPR upfront in simple and summarised narrative
- Restricting cookies

POWER SHARING

Power must be distributed more widely among communities and individuals who are impacted. Interventions should be co-designed and co-created with survivors.

In chat bots, this can be:

- During error experiences, offering users free text options so they can tell us why it was wrong.





About Chayn

Chayn is a global award-winning network of volunteers that produce simple crowdsourced resources (such as websites, toolkits & interactive web applications) using technology to empower women and non-binary people experiencing violence and abuse. Our methodology is 'design with, not for' as 70% of our volunteers are survivors of abuse who manage the total product cycle from ideation to project management and design.

www.chayn.co

Researched and written by

Ayesha Khalid, Beatriz Diaz, Carolina Moyano, Chloe Cross, Dina Ariss, Hera Hussain, Kim Burnell, Lakshmi Manjoosha Adapa, Naomi Alexander Naidoo, Nooreen Khan, Sarah Jaffri and Zoe Firth.

For further information, contact team@chayn.co or reach us on social media.