
ORBITS
Technology chapters

How technology enables abuse

There are several features of tech platforms that enable or facilitate tech abuse. While these features are not designed for abusers – they are usually designed for other, valid reasons such as user experience or efficiency – these vulnerabilities can be easily exploited to cause harm.

Vulnerabilities common to many tech platforms include:

★ **Limited user choice in what information is made public:** Most social media platforms make some personal information publicly available, which can be used by perpetrators to identify, harass, and stalk survivors.

★ **Applications connect contacts from phone, email, or social media and alert them:** Many platforms auto-upload contacts from users' phones or other social media accounts to allow people to quickly find friends and acquaintances already using that platform. This can enable abuse by automatically reconnecting survivors with their abuser and/or giving perpetrators frictionless access to many contacts. Some platforms also send alerts to users whenever a contact joins a platform, furthering this problem and the risk of triggering survivors.

★ **Frictionless sharing of photo and video content:** Most platforms allow easy downloading of photo and video content, making it easy for perpetrators to save, share, and use content. In addition, most platforms allow users to take screenshots of others' content and/or conversations without notifying the user.

★ **Sharing enabled for external applications:** Many platforms also include features for quick and easy sharing from one app to another. This feature can be used to easily spread abuse.

★ **Rigid and hard-to-find privacy settings:** While most platforms do offer a variety of options for privacy, these are often inflexible and do not allow people to personalise their privacy preferences. This means survivors are torn between risking their safety or completely privatising their account, which might have other professional or social consequences.

★ **Anonymous accounts:** Anonymous accounts are important for survivors and other marginalised folk. However, they can also be used by perpetrators to carry out abuse without accountability or consequences.

★ **Slow and not fit-for-purpose moderating and reporting mechanisms:** Across many platforms, the tools and processes to report abuse are not easy to find or use, are often slow, and may not be available for some languages at all. Furthermore, algorithms frequently fail to flag abuse, even when it's reported, and when human teams are working on abuse reports, they can fail to recognise and appropriately deal with abuse due

to a lack of training and context-specific knowledge. This issue is particularly pertinent in the Global South, as without sufficient cultural knowledge and training, moderators often do not recognise abusive content as abuse.

> *"People who complain using the reporting mechanisms find that they don't get a reply. It just sort of vanishes. There is no information on what is going to happen etc. There is a complete lack of transparency and that is one of the issues. A complete lack of response."*
>
> Bishakha Datta, Point of View

★ **Lack of timely, appropriate, and culturally adaptive moderation:** Inadequate policies and training of content moderators can create lags and lead towards incorrect decisions that harm survivors.

★ **Harm through content moderation:** Content moderation is often outsourced to poorly paid and supported 'ghost workers', usually based in the Global South. Reviewing abusive content can be traumatising, yet these workers are not given sufficient training or psychological support. This extends, rather than mitigates, harm.

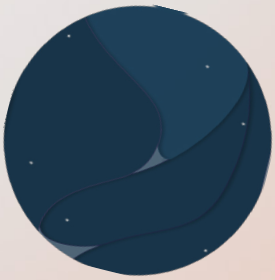
★ **Being able to contact people without pre-approval:** Platforms that allow users to call, message, and nudge people they do not know, without any options to set or change this preference, makes targeted harassment easy.

★ **Ability to create large distribution groups:** This makes room for rapid dissemination of abusive material, such as intimate images.

★ **Keeping users logged in even though they may be on a shared device:** For ease of access, many platforms offer default settings which keep users logged on to their platforms unless they proactively log off. This creates several security risks, including tech abuse.

★ **Limited recognition of the safety needs of people living in countries with oppressive regimes:** Political dissidence or protesting restrictive reproductive rights can be a lot more dangerous for women in countries with oppressive regimes, leading to imprisonment and sanctioning of activists. Women and queer activists are often targetted with dangreous gendered misinformation, dealth and rape threats, and doxxing which can pose a risk to their lives. These platforms are vital places for activists to mobilise their communities and share their work, and therefore their safety has to be ensured.

★ **Lack of blocking and muting options:** Different options for blocking and muting have evolved in recent years. For a long time, this was not possible on Twitter, Slack, and Skype.



Certain tech products also have specific vulnerabilities. For instance, iCloud makes it easy for perpetrators to take over multiple devices and access content, contacts, and more. Snapchat maps enable and encourage the sharing of location data. Facebook groups are used extensively to coordinate abuse. YouTube hosts channels for perpetrators seeking advice, guidance, and techniques to help them abuse. Reddit houses threads which illegally share content from OnlyFans. Clubhouse's onboarding process meant survivors were notified when their abusers joined the app, and both Clubhouse rooms and Twitter Spaces have created platforms for defending abusers and misogynistic speech. Up until late 2021, Google Drive did not allow you to block users, which meant abusive people could keep sharing files on Google Drive and it would still show up on 'Shared with me'. Features such as 'story views' on Instagram and 'viewed your profile' on LinkedIn can be used by stalkers to communicate that they are watching, while LinkedIn may be used for workplace harassment, as it normalises sending private messages to work contacts or colleagues.

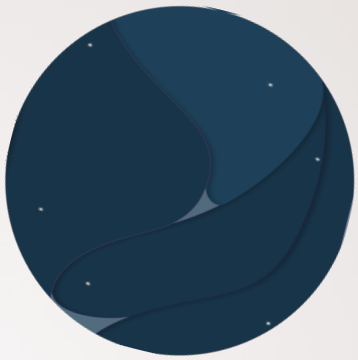
In addition to direct abuse, several platforms have censorship policies and practices that disproportionately harm marginalised groups and those campaigning for social justice, which can serve to reinforce systems of oppression and stall progress on issues such as GBV. For example, 'shadow banning' on Instagram and Tiktok is when a person's content is not shared with their follows, but they are not informed or given reasons for it.

As Safiya Noble has argued in Algorithms of Oppression, even search engines can facilitate harm by embedding biases against women of colour into their algorithm and search results.

Messaging apps also facilitate abuse. The default setting of messaging apps like WhatsApp and Telegram is to show when someone was last online, which can be used to track survivors. The accessibility and anonymity of these apps make them prime platforms for perpetrators. Group chats and the forward function are used for rapid dissemination of abusive material, and it's easy for users to make new groups when old ones are deleted or if they are removed from them.

Privacy features of Telegram in particular, such as heavy encryption and auto-deleting messages, are widely abused to perpetuate TGBV. On Skype, users can message, call, or video call others to harass them without even being added as a contact.

These are just some of the many vulnerabilities in tech platforms that can be exploited by abusers to carry out TGBV. These features have not been designed to facilitate abuse, but they do. The vast number of tech vulnerabilities shows the failure to consider and mitigate tech abuse in regards to tech design.



Case Study:

Electronic Frontier Foundation - Stalkerware and Apple AirTags

Electronic Frontier Foundation (EFF) is a USA based non-profit that works on ensuring civil liberties in the digital world. They champion user **privacy** and freedom of speech and expression, alongside technology development that supports global justice and innovation.

Apple launched the Apple AirTags on April 30 2021. These were marketed as small, inexpensive trackers that can be attached to or slipped into your belongings, so that you can keep track of items like keys or wallets. An iPhone is paired with the owner's AirTag so that they can play a sound on the AirTag or use its geolocation to locate any items they've attached it to. But AirTags can be used nefariously - they can easily be slipped into someone's bag and used to stalk them.

EFF was quick to recognise and draw attention to this risk. By mid-May, Eva Galperin, Director of CyberSecurity, wrote in Wired about these concerns. Apple AirTags are especially of concern in situations of intimate partner violence, where the domestic abuser could easily slip an AirTag into the survivor's bag to track them. This issue is not unique to AirTags, and is equally applicable to other tracking devices, such as Tile. However, Apple has a huge network, which means AirTag is able to show accurate locations by connecting with the Bluetooth of every active device in the Apple network. All Apple devices are added to the tracking network without first asking for the consent of Apple users. While it is possible to opt-out, users must do this for each device they own.

There are two safety features for iPhone users: a notification pops up when an unidentifiable AirTag is nearby, and nearby AirTags can be viewed through phone settings. However, initially, Android users had no way of finding out if there was an AirTag on them. Though AirTags have a serial number printed on them, which can help with finding out who owns it, it's difficult to locate the device on you in the first place as they are deliberately inconspicuous. The only safety feature built within the AirTag was that after 72 hours of being separated from its owner, it would ping at 60 decibels to alert those nearby. Since the sound isn't very loud, this could easily be muffled by placing it between things. According to Galperin, it's also unclear how long the beeping goes on for, and as she pointed out in Wired, 72 hours is a long time. This causes a huge safety concern for the person being stalked, especially if they live with their abuser, who can easily reset the alert every 72 hours. If they don't live with them, it means a person is still being stalked for 3 days without being alerted.

> *"When Apple fails to protect survivors, the consequences can be fatal. Apple leadership needs to give abuse survivors and experts a central place in its development process, incorporating their feedback from the start."* - Eva Galperin

With Galperin's help, journalists at The Washington Post also wrote about the issue, testing the device out in June. EFF proposed that Apple should design an Android app to alert users about Apple's AirTags. In June, Apple decided to change their policy and reduce the time it would take the AirTag to beep, from 3 days to 8-24 hours. In December 2021, Apple launched Tracker Detect, an Android app to help users identify if an AirTag or any other Find My Device is near them. The app shows nearby AirTags as an unknown item and can play a sound within 10 minutes of finding the AirTag. This is a major improvement from Apple, and is a direct result of EFF's advocacy. However, unlike the iOS app, the app won't run in the background and automatically alert the user. Tracker Detect requires that the user opens the app and runs a scan for the devices. The app will then provide instructions on how to disable the AirTag.

While there has been progress, safety concerns remain: the sound of the AirTag alert is still low and innocuous, the Android app isn't issuing alerts, and there's the issue of the alert being reset by an abuser who lives with the survivor. While Apple safety features are generally stronger, Apple users have to rely on the company's automatic scanning and have no way to actively scan, which can be an issue if you're tracked over a short trip. There are also loopholes such as family sharing, where family members can turn off the alerts on the device, or an abusive partner can simply tether the AirTag to the survivor's own iPhone so that they don't get any alerts. In 2022, Vice, the Guardian, the BBC, and others reported on rising cases of AirTags being used for stalking across the USA. Apple is continuing to introduce and investigate new safety features.

Our principles in practice

Though Apple has to be given credit for recognising the need to change their decisions, the case study provides us with a chance to reflect on what went wrong in the design process. When The Washington Post asked Apple if they'd considered domestic abusers and stalkers in their research, they were evasive. In Galperin's assessment, had they consulted an intimate partner violence specialist or survivors, the device design would have been very different from the start. Thus, Apple did not properly consider **safety** concerns when launching the product. Very overtly so, by enabling stalking, an AirTag completely infringes upon survivors' right to **privacy**, though it may very well maintain the **privacy** of the stalker who owns the device. EFF proposed that Apple users should not be automatically added to the tracking network, but should be able to give their consent, because it also makes all Apple users enablers for the stalker or abuser.

EFF also suggested that by giving space to experts and survivors of abuse, and involving them in the design process from the beginning, Apple could come up with better **safety** features for their devices. This would begin the process of **power redistribution**. Furthermore, the initial discrepancy in how Apple users were notified of an AirTag while Android users were not, showed a lack of **plurality** in the design of the device. The cost of having a mobile phone and the price difference between Android and Apple meant there was a class disparity in who this issue would affect, as it would particularly impact lower-income women and those in the Global South.

This posed major **equity** concerns. By addressing this through an Android app, Apple has demonstrated **accountability** for the harm their product decisions can cause. However, concerns remain, given that the **safety** measures for Apple and Android devices are still unequal, and very limited for those without a smartphone.

Galperin and EFF continue to advocate for survivor-centred approaches to eradicate stalkerware.

A systemic problem

Other than the features of tech platforms that are exploited to perpetrate abuse, there are systemic causes and structures that create favourable conditions for abuse to flourish and lead to inaction from tech companies. While these foundational issues are not the focus of this guide, they must be acknowledged as they underpin how and why technology facilitates abuse.

Prioritisation of issues and regions: Addressing tech abuse is not a priority for many tech companies. As tech abuse has gained more attention in recent years, more resources and efforts have been dedicated to tackling it, but this effort remains negligible in comparison to the gigantic turnover of these platforms. The problem is exacerbated by market prioritisation: there are unequal responses to tech abuse and thus different experiences for survivors between different markets, depending on economic priority. In particular, there is a huge discrepancy between the Global North and South, which manifests, for example, in the lack of proper reporting mechanisms in languages other than English.

> *'I think they are actually deciding not to invest in this. I mean, it's not that they are not capable of it, it's that they are deciding not to. They have all the resources. They have financial resources, artificial intelligence resources, they have offices all over the world. They could really be making the difference, and I think it's just that the priorities are not there.'* - Lulú V. Barrera, Luchadoras

> *'All tech companies have priority markets, where they know they have a presence, it can influence other behaviour in a specific sub-region. So that also means that the priority of issues or the priority of solutions go to those specific markets, they just don't trickle down to everybody. I remember once attending Facebook launching a missing child alert in South Africa. And I was wondering, when is it going to roll out to the other countries?''* - Chenai Chair, Mozilla Foundation

Business model: The business models of most social media platforms are built on engagement, whether that is driven by civic or hateful speech. The more people engage, the more profit tech platforms make. Arguably, this business model is incompatible with effectively tackling tech abuse, because it is not in the interest of tech companies to curb abuse as long as it is driving engagement.

'I would say the major problem with social media platforms when it comes to this kind of abuse is that, for most of these companies, their entire business model is in engagement. It doesn't matter what kind of engagement. It doesn't matter if that is good or bad, or destroys someone's life, it's just the more you get people to engage, the better it is for the company. When that is your entire business model, you don't prioritise things like harm, and you don't prioritise things like keeping people safe, you just prioritise having more people engaged.' - Mary Anne Franks

Power asymmetries: As tech giants grow and increasingly monopolise sectors, the power asymmetry between them and citizens, as well as civil society and even governments, increases. The use of technology has become a point of access to more and more vital services, leaving users with nowhere else to go, and no power to reject or question their terms of use. Tech companies wield power over governments by offering relevant tech infrastructure, as was demonstrated with the development of the COVID-19 contracting tracing apps, and Google and Apple's decision to integrate the technology into their operating systems. Tech giants have become too big to fail.

Diversity within teams and leadership: The inequalities of the wider world are often mirrored within tech companies, and discrimination is a major issue. While diversity and inclusion of marginalised groups is an issue in tech at all levels, it is particularly so at decision-making and leadership levels, meaning the concerns of marginalised groups are easy to ignore. The lack of gender diversity in tech - only 20% of the USA tech workforce is made up of women - is especially detrimental when it comes to tackling gendered tech abuse. Worryingly, AI may make this situation even worse, as women are at higher risk of displacement by automation than men.

Transforming technology: designing for healing

We've seen how technology can facilitate abuse. But this is by design, not necessity. We propose a model of design which enables technology to be used as a tool to mitigate harm and support healing for survivors of TGBV.

When designing online tools, we need to approach it as though we are designing a physical space - say, a cafe. What do we want people to think about when they stand on the street, looking at our cafe window? What would it feel like if they stepped inside? Would they want to take a seat and linger, or would they want to quickly grab something they need and leave? Do they feel like they can do both depending on their mood and routine?

Applying an intersectional, trauma-informed, and survivor-centred lens presents us with new questions to consider. To ensure the cafe is inviting and comfortable for a wide variety of people with different needs and life experiences, how might we alter the design? If we know that the cafe will welcome survivors who have experienced trauma, what might we change or add to its design?

Likewise, we can think of large social media platforms like towns or cities made up of different communities, infrastructure, and trends. What does it say about our curation of these spaces that so many people feel comfortable shouting, abusing, and threatening to harm others? This behaviour would be addressed by bystanders, community leaders, and authorities in real life, so why isn't this happening online? How can we reimagine online spaces so they reward community and connection rather than conflict and hate?

These are big questions that scholars, activists and platform designers are grappling with. Ethics of technology is an expansive field and there are an ever-growing number of ethics toolkits such as the Ethics for Designers tools, Ethical Design Guide, Consentful Tech Project(and their Consentful Tech Curriculum), Design Ethically Toolkit and Tarot Cards of Tech, that can ground and guide discussions.

But what does transformative, ethical technology design look like when we focus specifically on gender-based violence?

Systemic problems; systemic solutions

We've seen how the systemic problems of market prioritisation, business models, lack of diversity, and power asymmetries influence the way technology platforms enable tech abuse, as well as fail to respond to it. Orbits is focused on providing practical tools that every researcher, policymaker, and designer can use, and the recommendations in this guide can go a long way in better mitigating and responding to tech abuse. However, we also know that harm will continue unless the root causes are tackled. In parallel to immediate interventions, we advocate for the following systemic solutions to transform the tech ecosystem:

Alternative business and governance models: If technology companies are failing to effectively tackle tech abuse because of how their business models operate, alternative business and governance should be part of the solution. Non-profit models, mutual ownership, stakeholder (rather than solely shareholder) engagement, and democratic governance should all be explored as part of the systemic response to tech abuse. For example, the platform co-op movement advocates for tech platforms which are cooperatively owned and governed.

Open source technology: Open source technology refers to software where the source code is open and available to be viewed, re-used, and adapted by everyone. Open source technology promotes collaboration and shared learning between technology companies, rather than competition. It's also resource efficient, easing the high development costs of technology and duplicating efforts, and enabling those resources to be directed elsewhere. All of Chayn's products and services are open source.

Diverse, inclusive teams and management structures: The lack of diversity within tech companies, especially at the senior level, presents major barriers to addressing tech abuse, and implementing the intersectional, survivor-centred, and trauma-informed approach that is required. To remedy this, we must not only diversify these organisations and decision-making teams, but also transform the organisational cultures, management structures, and HR practices that have dominated until now. It is not enough to give 'a seat at the table' to people from more diverse backgrounds, communities, and identities - we must rebuild the tables and the rooms where decisions are made so they can genuinely hold multiple perspectives and facilitate decisions that reflect them.

Check out Mozilla Foundation, Tactical Tech, Algorithmic Justice League, New Public, and Amnesty Tech to learn more about transforming technology.

Learn more about technology design which centres survivors and other marginalised folk in our favourite technology design books: Design Justice by Sasha Costanza-Chock and Design for Safety by Eva PenzeyMoog. For more on developing tech policy, see Superrr Lab's Feminist Tech Principles.

The Santa Clara Principles provide a framework for transparency and accountability in content moderation. Find out more about best practices for gender-inclusive content moderation, compiled by Trust and Safety professionals from the tech industry, here.

IBM have produced five design principles for technology design which are resistant to coercive control. Catalyst's safeguarding resources are designed to help build safe digital services.

Design principles and applications

1. Safety

Safety by design should be a prerequisite for any product but it becomes critical when designing for an audience that has been denied safety, such as survivivors of TGBV. Often, safety risks are minimised or deprioritised in technology design. Instead, we must embrace risk analysis as a way of ensuring more people can use our products, which will improve future outcomes for all.

Application examples:

★ Testing all technology for abusability by conducting threat modelling at multiple stages of the design lifecycle.

★ N2 Factor Authentication.

★ Safety exit button on websites that take users to a non-conspicuous website in case someone is watching them.To support emotional safety, consider redirecting to something comforting instead.

- ★ Allowing users to opt for disguised emails with fake subject lines, like Chayn's mini-course platform Soul Medicine.
- ★ Designing reporting mechanisms that don't involve resharing or further distributions of harmful content.
- ★ Blocking and filtering content and users.
- ★ Offering options to restrict how people can get in touch with users.
- ★ Not showing people someone they may know, as it can make someone's secret profile discoverable.
- ★ Not saving information on the user's end as they might be using a shared device.
- ★ In chat bots, providing safety advice before and during conversation.
- ★ The ability to use alternative names, which can help stop stalkers and abusers from finding and tracking survivors.
- ★ Sharing last known logins, so survivors can spot if an abuser or stalker has managed to get control of their devices or accounts.
- ★ Creating user controls on how images can be downloaded and shared.
- ★ Digital fingerprinting, to assist with removing offending materials from all platforms and flagging accounts that shared the offending materials.
- ★ Offering to provide safe contact details as these may differ from the ones that they use to access platforms.
- ★ Providing clear terms of use that highlight zero tolerance for abuse and clearly identify examples of harmful behaviours prominently.
- ★ Permitting third party reporting.
- ★ Reporting to platforms for offline behaviour of users.
- ★ Adding perpetrator information to a digital offender database maintained by the company or law enforcement (if applicable).
- ★ Providing adequate support and trauma counselling for moderation staff.

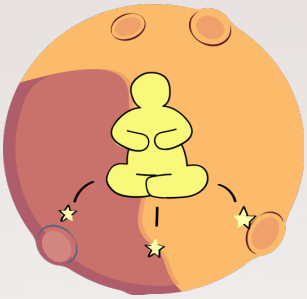
Case Study:

Exit Buttons

Exit buttons are a **safety** feature for websites on sensitive subjects, such as gender-based violence and other forms of abuse. They provide a quick one-click solution to navigate away from the webpage you are viewing, should you need to conceal it from those who are physically nearby. This would be useful in situations where you are in an abusive home, using a public computer, or at work.

As exit buttons have become common practice in recent years, there are some interesting innovations in how to design them. AVA's Breathing Space application lets users choose their own exit page as they are creating an account, and the app remembers their choice. Other websites disguise pages by creating a pop up that covers the website with something innocuous.

For instance, Chayn's exit button 'Leave this site' takes users to Wikipedia's homepage. It used to be Google, but was redirected to Wikipedia to support their mission and because, as the world's number one place to find information, it felt like a good fit. To provide some relief in the moment of panic when someone might need to press the button, not only does the button open a new tab with Wikipedia.com, but also searches 'cute baby animal memes' in the tab where the Chayn website was open. If you click back on the tab, it takes you to a blank screen. In this way, Chayn's button simultaneously deals with physical and emotional **safety**.



2. Agency

Lengthy legal forms that are set out to get consent for data protection are flawed because most users don't want to read through them. Sometimes, it's questioned whether it is safe to expose survivors to co-design processes due to fear of retraumatisation. These attitudes are paternalistic and patronising. We must always centre the user's agency alongside safety, as it is demonstrated that creating environments that value agency can build trust.

Application examples:

★ Offering tools that people can customise and use at their own pace.

★ Refraining from assumptions that survivors of abuse do not want to take an active role in design or feedback.

★ Creating flexible mechanisms that enable people to describe their own experience and share the remedial measures they wish for, rather than forcing reports into rigid, predetermined categories.

★ Allowing people to access essential information without having to create an account.

★ Giving an option of what information is kept public and private, such as full names and location.

★ Building room for consent at various stages, especially in reporting processes. This means actively asking survivors for their consent in sharing information with other agencies and individuals within the organisation, and being clear with survivors about how and why their information is being shared.

★ Providing comprehensive reporting mechanisms that let survivors report even if the perpetrator deactivates/disconnects their account.

3. Equity

Inclusion by design should be the norm, so that products and services can be used by everyone. When designing products that affect diverse groups, it is crucial to actively be aware of and avoid racial, gender, and class stereotyping, as well as geopolitical differences. For instance, accessibility considerations should support access to people with disabilities, prevent exclusion, and produce a superior, more usable design which promotes a sense of belonging for all.

Application examples:

- ★ Designing products that cater to a range of accessibility requirements such as speech and hearing impairments.
- ★ Providing resources and information in multiple formats - for example, captioned videos as well as written resources.
- ★ Ensuring strong referral pathways to specialist services for survivors from marginalised communities.
- ★ Introducing voice-activated reporting mechanisms to account for different literacy levels and the diverse technology needs of different communities.
- ★ Rolling out new safety features simultaneously in all low and high-income countries.
- ★ Making policies and reporting mechanisms available in different languages and dialects.
- ★ Offering reporting processes with accessibility considerations embedded, including an option for low-bandwidth or offline reporting.
- ★ Providing staff training and learning opportunities on anti-oppression and decolonisation.

4. Privacy

In an economy where data is considered the currency of interactions, we must consider the harm we may introduce from intrusive data collection, storing, and selling. This involves understanding that some vulnerable groups will not be able to foresee the risks that may arise when they share their data. Data justice acknowledges that information can often be used as a form of oppression by rendering certain communities invisible or misrepresenting them, and thus we need to actively think about how people are counted, represented, and treated through the lens of data science.

Application examples:

★ Securing all databases.

★ Clearly indicating what data is publicly accessible and what isn't.

★ Automatic disabling of cookies and tracking when survivors report abuse on platforms.

★ Only collecting information that is absolutely necessary and creating clear options for more data storage.

★ Using end-to-end encrypted technology.

★ Exploring the use of privacy-enhancing technologies (PET) such as encryption and data masking.

★ Holding entities liable for misuse of sensitive data.

★ Avoiding misleading language and design that can lead to usage of data in ways people have not agreed to (often for profit).

★ Plainly articulating policies in an easily understandable format. If they are long, there should be a summary available so users understand what they are agreeing to.

★ Seeking explicit consent for selling user data where relevant, especially when it is related to marginalised group.

★ Maintaining strict confidentiality for reporting processes.

★ Withholding survivors' details from the perpetrator during any punitive actions taken.

★ Providing survivors with a digital file of evidence that can support civil and criminal cases, if they want to pursue those routes.

Learn more about data justice: Data 4 Black Lives, Te Mana Raraunga (indigenous data sovereignty in New Zealand) and Data Feminism. To benchmark your organisation's data ethics, see the Open Data Institute's Data Ethics Maturity Model.

5. Accountability

When opaque reporting mechanisms, features, and algorithms are commonplace, survivors learn that they should not place their trust in technology. Therefore, technology companies must deliver timely responses and clearly articulate rationales for decisions which impact the safety and lives of survivors.

Application examples:

★ Providing clear ways to help survivors identify in-platform reporting mechanisms. This means quick access bars for reporting abuse, supported by clear wording about what follows.

★ Communicating to survivors which department deals with the report work and informing them that there is a dedicated and specialist resource to handle reports

★ Actioning user research and feedback in design.

★ Sharing openly when something is not working or is a trial feature.

★ Acknowledging gaps in knowledge or foresight which can contribute to harmful features.

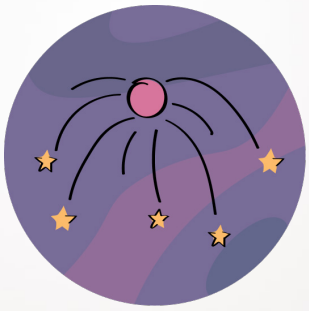
★ Being clear about the hours of your service or the boundaries of your support.

★ Being consistent and predictable in product design - by providing structure and routine, you signal to users that not only have you thought about the service, but are a stable source of support for them. It's not one interaction you're seeking, but the start of a long-term relationship.

★ Committing to long-term change, rather than reacting to scandals and infrequent public outrage.

★ Creating effective and responsive grievance redressal mechanisms on platforms for reporting tech abuse.

★ If applicable, removing the offending user's accounts from other platforms owned by the parent company.



6. Plurality

We need to design for cross-cutting needs, power, and experiences that can change how an individual experiences the digital world and seeks remediation from it. A decolonising design practice will understand the many ways in which harmful stereotypes can turn into assumptions for users.

Application examples:

★ Training moderators to understand cultural context.

★ Refraining from assuming which language is spoken based on location.

★ Offering ways for people to customise their journey on your product or platform.

★ Training staff on the impact of additional vulnerabilities, such as caste, race, religion, sexual orientation, and disabilities.

★ Recognising that people in digital spaces might experience multiple forms of discrimination/hate (for example, gender and race discrimination). Therefore, in complaint processes, it should be possible for survivors to identify multiple offences, including offline ones.

7. Power redistribution

Survivors are often consulted after preventative and restorative measures have been designed. We must ensure that the power to decide those measures lies with the survivor, and that this input is valued through a form of compensation.

Application examples:

★ Giving survivors decision-making power in tech companies through compensated board or committee positions.

★ Consulting communities through different stages of research, design, and implementation.

★ For global firms, using local teams and networks to gather ideas for ways to improve services.

★ Creating community-owned models and practices for governance and evaluation.

★ Translating and localising content and policies.

★ Citing and sharing the work of all feminists and scholars who have influenced or shaped decisions, especially from the Global South.

★ Giving content moderators opportunities to feed into global policies.

8. Hope

In an effort to build rapport with users, some organisations mistakenly use traumatising pictures and words that can be harsh, such as pictures of a man punching down a cowering woman, or a woman crying or covered with bruises. This risks transporting survivors to times when they felt unsafe and, therefore, should be avoided. We should create visual design that uplifts the mood of survivors, and soothes them. Online spaces should feel as warm as possible when someone is feeling unsafe in their physical world.

Application examples:

★ Using an empathetic tone in written and vocal communications.

★ Ensuring visual assets are not retraumatising.

★ Displaying simple, soothing, and visually appealing UX.

★ Prioritising ethical considerations in corporate decision-making over shareholder priorities.

★ Sharing the work of activists, civil society groups, and innovators working to tackle challenges.

★ Providing realistic information about reporting processes. (For example: 'we respond to requests in 2 to 48 hours, with 70% of reports getting an answer within 10 hours').

★ Thanking survivors for their decision to report through repeated automatic messaging by the individuals who are handling their reports.

★ Taking proactive and communicative steps to stop tech abuse (For example: flag and/or blur offensive content and create digital fingerprints to block uploading of flagged content).

Case Study:

Bloom by Chayn - using tech to support healing

Bloom is a remote trauma support service developed by Chayn. In 2020, as COVID-19 lockdowns were introduced around the world, many survivors were trapped at home with their abusers and/or unable to access in-person support systems. Bloom was created as a response to these circumstances, which also filled an existing, serious gap in online, scalable services that survivors anywhere can access for free.

How Bloom works

Bloom delivers trauma support via online courses. Course participants receive access to pre-recorded videos with grounding exercises, information and guidance to support healing, 'homework' activities to do in their own time, and access to 1-2-1 chat with the Bloom team. The courses are designed to be taken over three to eight weeks, but participants can take the course at their own pace. The 1-2-1 chat can be accessed via web browser, WhatsApp or Telegram, and is a space where participants share their reflections and questions on the course content and activities, as well as talk about their experiences of gender-based violence, their recovery journey, or even just how they are feeling.

The aim of Bloom is to 'inform and empower.' To inform, the courses include information on topics such as the fear response and how the body can repeat this response after trauma, and how our sense of self, as well as relationships with others, can be affected by trauma. To empower, it includes practical tools for grounding ourselves in the present, assertive communication techniques for healthy relationships, and a variety of journaling techniques for exploring our own stories and healing. All of this is grounded in an intersectional feminist worldview, that takes a critical look at the ways society enables predators and abusers. Bloom clearly communicates that abuse is never the survivor's fault. The course content is developed and written by survivors in collaboration with a trauma-informed therapist.

In 2021, Bloom ran five courses: Creating Boundaries, Managing Anxiety, Healing from Sexual Trauma, Recovering from Toxic and Abusive Relationships, and Reclaiming Resilience in Your Trauma Story. Bloom also launched an industry-first partnership with dating app Bumble, by providing a customised version of Bloom to Bumble users who report sexual abuse or assault. By the end of 2021, Bloom had supported over 1,000 survivors from over 60 countries. 97% of Bloom users would recommend the programme to someone in their position.

> *"Through Bloom, we see the kind of deep impact that comes from people understanding how trauma has impacted them, and how sexism shapes even the way you deal with it. 40% of survivors who take our course have never been to a therapist due to lack of affordability, stigma, or fear of being seen."* - Hera Hussain, Founder & CEO, Chayn

Our principles in practice

Bloom prioritises **privacy** by making all courses completely anonymous - participants do not have to share their real name or any personal information to take part. Participants do not interact with each other or find out who else is doing the course, but they work alongside other survivors and are continuously reminded through the courses that they are not alone and 'are in this together'. In this way, they benefit from group learning, without compromising on **safety**. The **safety** of Bloom is further supported through safeguarding processes, including mandatory safeguarding training for all Bloom team members.

To ensure the **agency** of survivors, the courses are made to be flexible - participants can learn at their own pace. They can watch the videos and complete the activities whenever it is convenient for them. This adaptability responds to a **plurality** of survivor experiences and needs. Moreover, participants actively shape the course - the course content is continuously adapted and improved by feedback received during the courses and from regular user research interviews. In this way, Bloom practises **power redistribution**, too.

Bloom also promotes **equity** by ensuring the course content is relevant for all survivors, and uses examples which particularly highlight the experiences of marginalised groups. Since the service is completely free, no-one is priced out. To improve accessibility, transcripts are available for all course sessions, in addition to the videos, and all videos have captions which are edited for accuracy.

Hope is central to Bloom - the foundational message of all courses is that healing from trauma is possible for every survivor. Moreover, Bloom seeks to inspire **hope** in each participant through inviting, soothing UX and by starting each video with a grounding exercise. These grounding exercises are designed to help participants mentally distance themselves from their daily lives and physical surroundings, and feel physically and psychologically present in Bloom's online space.

In response to the growing rate of tech abuse, Chayn has started working on a new Bloom course, focused on image-based abuse.

Case Study:

Tech Policy Design Lab - co-creating tech policy solutions to end online GBV

The Tech Policy Design Lab, an initiative of the Web Foundation, aimed to create innovative tech-policy solutions for building a safer and more equitable internet, free from GBV. From March 2020 to February 2021, the Web Foundation hosted a series of four multi-stakeholder consultation workshops to explore and build understanding about online GBV on women activists, women in public life, and young women. The findings from these consultations were used to develop three policy design workshops in April 2021. Partnering with service designers Craig Walker and Feminist Internet, the Web Foundation brought together the world's largest tech platforms, policymakers, academics, and civil society organisations to co-create solutions for tackling online GBV through multi-stakholder workshops. This project especially focused on women in highly public-facing roles (such as politicians, journalists, and activists) leading active online lives. Based on the insights from the consultation workshops, policy design was concentrated on two areas of great importance for creating a safer internet for women: curation and reporting.

Curation: Greater control over who can comment or reply to posts, as well as more choice over what women see online, when they see it, and how they see it.

Reporting: Improved reporting systems so women can be better supported when they do receive violent or abusive content.

Policy design method

The Tech Policy Design Lab used design thinking and co-creation methodologies to generate potential policy solutions around these two themes. Participants worked in small multi-stakeholder groups and were given a specific scenario to design for, including a fictional persona, app, and problem. While the scenarios were hypothetical, they were based on the real, lived experiences of women facing online GBV. The personas were chosen to represent intersecting identities (for example, race, sexuality, and gender identity) to encourage solutions to take an intersectional approach. Using this methodology, participants were able to design solutions based on the needs of survivors, rather than being limited by currently available tech solutions.

"While we can't quickly unwind the sexism that drives abuse, we can redesign our digital spaces and change the online environments that allow this misogyny to thrive." - Azmina Dhrodia, **Safety** Policy Lead, Bumble (formerly Senior Policy Manager, Web Foundation)

Prototypes

The workshops generated 11 promising prototypes for tackling online GBV. For example, Reporteroo is a prototype that affords transparency for users in the reporting process by allowing simple, real-time access to information about follow-ups, and also providing the option of reporting in local languages along with the provision to add context-specific information of the incident. Another prototype, Com Mod, allows users to appoint trusted users who can then moderate comments on the user's behalf. The actions taken by trusted users can be approved or reversed by the original user if needed. This prototype reduces the burden of trauma experienced by women facing abuse by reducing the amount of abuse they see and allowing delegation of removal/ blocking/restricting of abusive comments to someone they trust. These collaborative solutions explore the scope for community intervention and prioritise the **safety** of vulnerable users.

Recommendations

The final report on Online Gender-Based Violence and Abuse was released by Tech Policy Design Lab in June 2021. Based on the workshop discussion and prototypes developed, the report includes user-centric recommendations, design suggestions about how recommendations could be achieved, illustrative examples of what the recommendations could look like in practice, and other considerations that should be taken into account when introducing these measures, such as technical challenges, required policy changes, and the possibility of misuse.

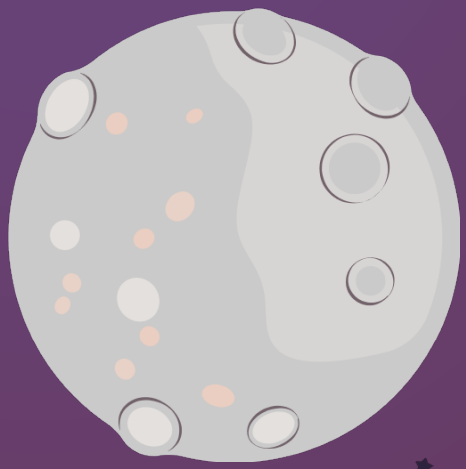
Curation	Reporting
1. Offering more granular settings (e.g. who can see, share, comment, or reply to posts) 2. Using simple and accessible language throughout the user experience 3. Providing easy navigation and access to **safety** tools 4. Reducing the burden on women by proactively reducing the amount of abuse they see	1. Offering users the ability to track and manage their reports 2. Enabling greater capacity to address context and/or language 3. Providing more policy and product guidance when reporting abuse 4. Establishing additional ways for women to access help and support during the reporting process

The Tech Policy Design Lab not only generated concrete suggestions for how to design technology that addresses online GBV, but also demonstrated how survivor-centred, trauma-informed, and intersectional policies can and should be developed. By clearly detailing their process as well as their findings, the Web Foundation offers a blueprint for technology companies on how they can work together with civil society, academia, and survivors to co-create policy and design solutions that effectively tackle GBV on their platforms. The participation of representatives from big tech companies like Facebook, Google, Twitter, and TikTok in the workshops means they now have first-hand experience of this process. The Tech Policy Design Lab acts as a benchmark against which the tech companies' progress can be measured.

Our principles in practice

The Tech Policy Design Lab supported **power redistribution** by creating multi-stakeholder spaces where everyone worked together to create solutions. Moreover, it encouraged **accountability** from the world's most powerful tech platforms by involving them in the process. By adopting a design thinking methodology, and creating personas with intersecting identities, **plurality** and **equity** are prioritised.

Tech Policy Design Lab's recommendations promote **agency** (by focusing on curation of content by survivors, and more oversight and control in the reporting process) and **safety** (by recommending how to restrict the amount of abuse women see online and offer more support throughout the reporting process). By initiating this project, sharing their process and insights openly, and making concrete recommendations to tech platforms, they offer **hope** for a better, safer, and more inclusive internet.



Case Study:

Pex - fighting IBA with technology

Pex is a digital rights technology company enabling the fair and transparent use of copyrighted content on the internet. Founded in 2014, Pex has developed a copyright solution for the creator economy known as Attribution Engine, which enables content identification on digital platforms so that creators and rightsholders can be acknowledged and credited for their work. When building their Attribution Engine, the Pex team recognised that it could be used for another purpose too: helping to prevent the spread of toxic content, including image-based abuse.

> *"Technology alone isn't going to solve the problem, but it needs to be a massive part of the solution. The internet is still the wild west and we have so much opportunity to make it a better place for everyone."* - Chanelle Murphy, Product Manager of Trust and **Safety** Division, Pex

Pex's Trust and **Safety** division has developed a feature designed specifically for preventing the publication of known toxic content on platforms. Built with Pex's leading fingerprinting technology, Attribution Engine can scan videos and images for known abusive content and send information about the content automatically to the appropriate digital platforms so that it can be flagged for removal or blocked before it gets published. Pex partners with trusted non-profit organisations who are provided a user-friendly software development kit that creates fingerprints locally. The fingerprint is then sent to Pex and compared against user-generated content, or UGC, fingerprints in real time. If a match is identified, the content-sharing platform is notified and Image-Based Abuse (IBA) is blocked from the platform before it is ever posted. These results are communicated back to a Pex dashboard, which shows non-profits where the content has been uploaded or blocked. Pex does not store the content in its original form, and digital fingerprints cannot be re-programmed to derive original images.

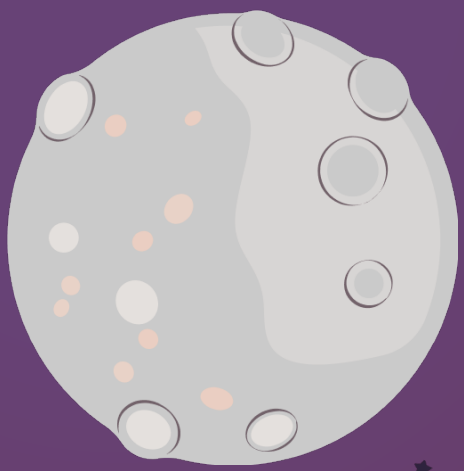
Alongside creating this tech, Pex has also begun community engagement work on the issue of IBA. Since IBA is a reflection of societal attitudes and prejudices, Pex sees a role for facilitating conversations to raise awareness about this topic, build solidarity and empathy for survivors, and shift the narrative. For this, Pex has started an initiative called the Trust and Safety Internal Community, in which Pex staff meet to talk and learn about different kinds of IBA, its prevalence, and the implications on survivors' lives. They **hope** these discussions will motivate employees to speak to their families and friends, and to become advocates against IBA in their communities.

> *"This is a fundamental-societal problem, and it's going to take a lot of voices coming together, in addition to heavy tech solutions."* - Chanelle Murphy

Our principles in practice

The capabilities of Pex's technology improve **privacy** and **safety** for survivors, by providing an effective route to report and remove IBA, without needing to continuously share or engage with it. Pex prioritises the emotional **safety** of survivors too, by collaborating with trusted non-profits to deliver this tool so that survivors know they can trust the process. Simple design with step-by-step guidance on reporting abuse makes removal of IBA content easier for the non-profit staff, reducing the risk of vicarious trauma.

Pex's Trust and **Safety** team have worked extensively with survivor advocates and non-profits to develop the technology, showing a commitment to **power redistribution**. By enabling non-profits to report their IBA content and have it not only removed but also blocked from future uploads, Pex provides a beacon of **hope** for survivors.



Case Study:

Digital Rights Foundation - Cyber Harassment Helpline

Digital Rights Foundation (DRF) is a feminist, not-for-profit organisation based in Pakistan. Founded in 2013 by lawyer Nighat Dad, DRF defends digital freedoms and rights through awareness-raising, research, and policy advocacy. One of their priority aims is protecting women and other marginalised groups from online harassment.

In 2016, after running an awareness campaign about online harassment and digital safety, the DRF team found themselves inundated with messages from women looking for guidance and help with cases of cyber harassment. DRF recognised the need for a dedicated channel to deal with these enquiries and later that year, established the Cyber Harassment Helpline - the region's first helpline for these kinds of cases. Today, the helpline receives an average of 212 calls per month.

> *"And we have seen that the number of such complaints never decreases at the helpline. It always increases. Even though there is a lot of awareness. Despite the fact that we have a "cyber crime law" that aims to protect women online."*- Nighat Dad, Executive Director, Digital Rights Foundation

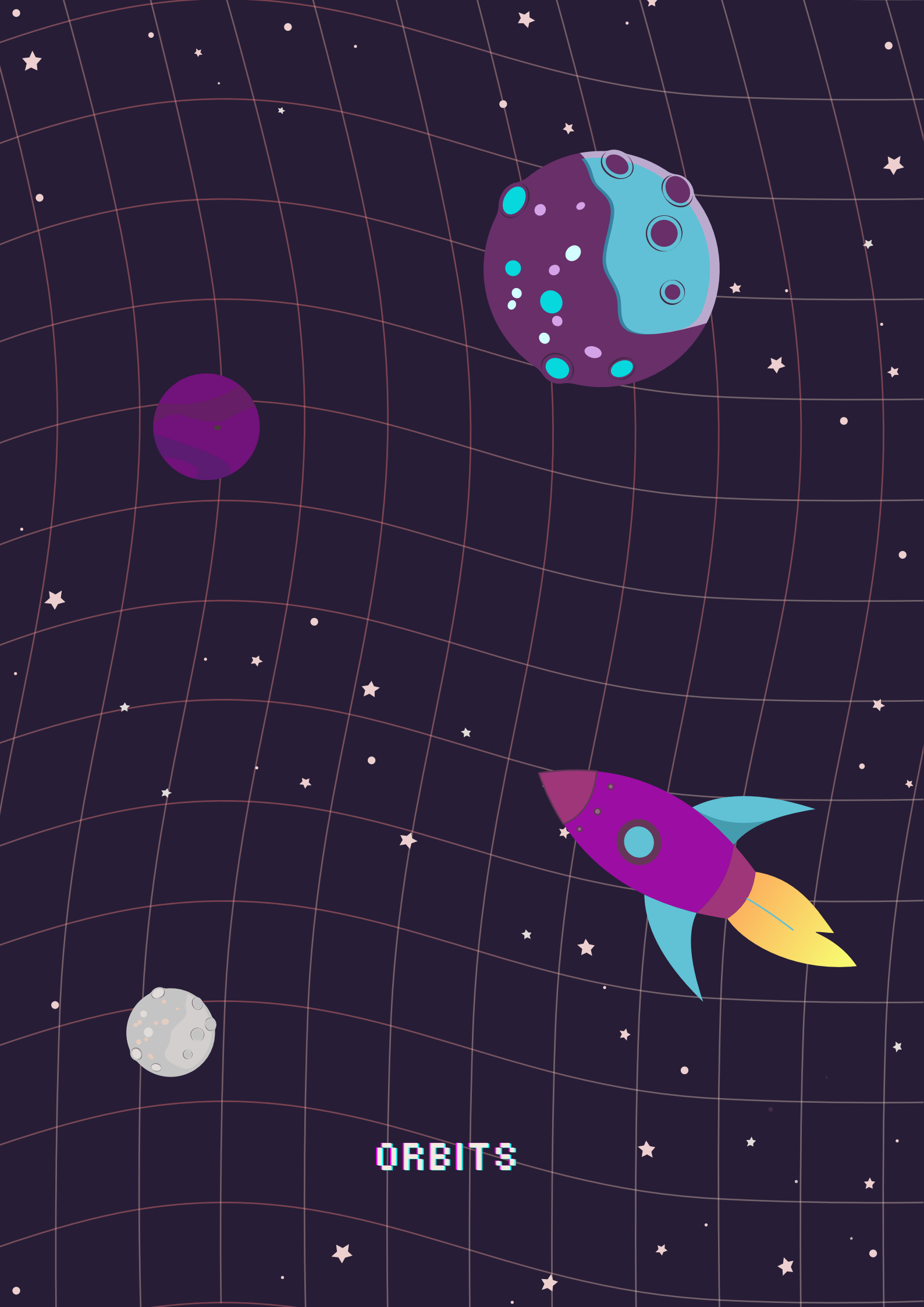
The helpline receives calls on many different types of online violence, including hacking, online stalking, doxxing, impersonation, and abusive language. However, their most common cause of complaint (around a third of overall calls to the helpline) relates to blackmailing: when threats and demands are made based on sharing an individual's personal information and/or photos without their consent. This presents particular dangers in Pakistan, where cultural and religious norms mean information and photos shared online can be the cause of great shame and backlash. This can therefore restrict a survivor's ability to exist online, as well as have serious offline risks for survivors including mental health implications, punishment from family, restriction of other freedoms (for example, the opportunity to go to university or work), and violence.

While the helpline was originally set up to provide digital security support, the service has now expanded to offer psychological counselling and legal assistance to keep up with the demand. Over a quarter of callers require legal assistance, and DRF has a network of lawyers who offer pro bono legal support to callers. Helpline support staff are all trained in psychological support and can assess distressed callers against mental health indicators, referring them to DRF's in-house psychologist if they are found to be at risk.

Our principles in practice

Privacy is foundational to how the helpline operates. DRF prioritises caller confidentiality and does not collect any information which is personally identifiable. If it's assessed that the call might be cut off, phone numbers are temporarily stored so DRF can contact the caller, but numbers are never collected in permanent records. Prioritising the **agency** of survivors, the DRF team is very careful about if and when they use survivor stories in their advocacy or awareness-raising work. When they do, they work with survivors whose case has been resolved or come to some sort of conclusion, and/or those they have a long-standing relationship with. They are also careful to inform survivors about exactly how and why the information will be used, ensure they are providing remedial resources throughout the process, and protect the survivors' anonymity.

Learn more about Nighat Dad's work and life story in this Digital Rights & Feminist Future zine.


ORBITS