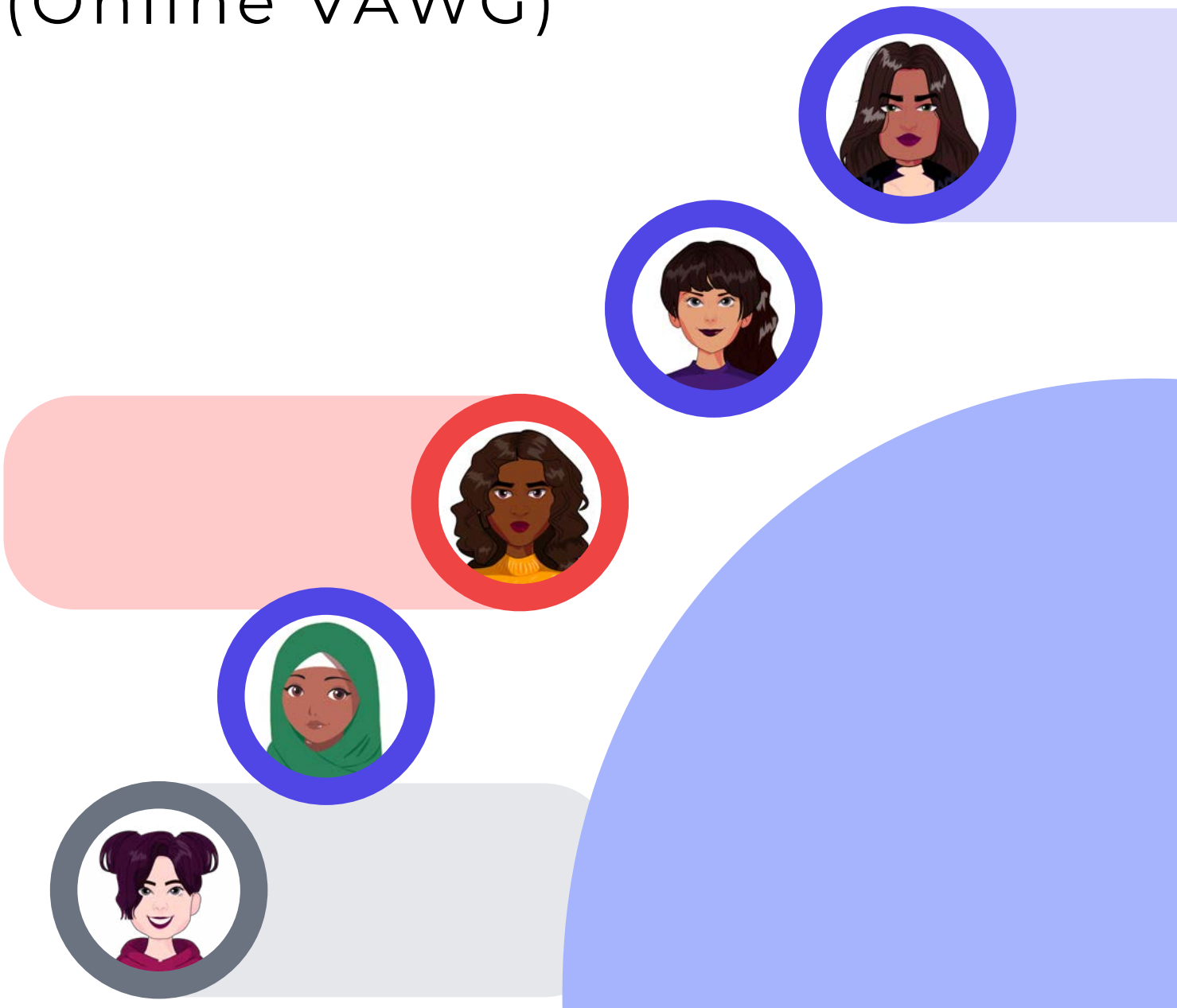
PUBLIC

Department for Science, Innovation & Technology

# Platform Design and the Risk of Online Violence Against Women and Girls (Online VAWG)



# Executive Summary

Online violence against women and girls (online VAWG) is a pervasive problem, impacting the lives of women and girls globally.[1] While the exact prevalence of online VAWG globally is unknown, research suggests that 85% of women and girls have witnessed online VAWG.[2] The Covid-19 pandemic further exacerbated online VAWG as women, girls, and perpetrators increased their time online.[3]

While available research evidences the growing problem of online VAWG, it is important to recognise that online VAWG is chronically underreported, as women and girls are not supported in identifying, documenting, and reporting online VAWG on platforms and other online spaces.[4] This stresses the importance of identifying the gaps in the current understanding of online VAWG, and supporting research and evidence gathering to fill these gaps.

Government, platforms and civil society organisations around the world are increasingly turning their attention to online VAWG. However, there is a lack of understanding around the role of platform design specifically. To fill this gap, the UK Department for Science, Innovation and Technology (DSIT) commissioned this report in late 2022 to investigate the impact of platform design on online VAWG. The research is intentionally geared towards an international audience so as to further support the work of the Global Partnership for Action on Gender-Based Online Harassment and Abuse. As a result, we hope this report will directly contribute towards the Global Partnership's ambition to expand the global evidence base of risk factors towards online VAWG. This project was funded by the Department for Science, Innovation & Technology, however the views expressed in this report reflect the research carried out within the parameters of the project and do not necessarily reflect the views of HM Government.[5]

The research has three objectives, focusing on the current challenges, successes, and future opportunities around platform design, as outlined below.

1. **Understand how** design features of online platforms and services can enable the perpetration of online VAWG
2. **Develop an understanding** of how existing safety by design approaches can protect against the risk of online VAWG
3. **Understand potential** new design approaches to ensure safety for women and girls on online platforms and services

To achieve these objectives, the research took a three-pronged methodological approach:

1. **Landscape mapping** to understand key trends across the online VAWG ecosystem through a literature review and an evidence review of platform features and policies.
2. **Stakeholder engagement** to validate findings from the literature review, generate further evidence, and further refine our understanding of platform design.
3. **Design analysis** to consolidate the findings from the landscape mapping and stakeholder engagement, and consider potential future design features.

## Key Findings

The report findings can be summarised based on the objectives, as outlined below:

### 1. Understand how design features of online platforms and services can enable the perpetration of online VAWG

- There are five key challenges with the perpetration of online VAWG through platform design. First, a lack of platform policies explicitly targeted at online VAWG makes it unclear to victim-survivors as to how they can be supported, as well as making it harder to sanction perpetrators. A lack of focus on women and girls may also contribute to poor data collection on their experiences on platforms, making it more challenging to pinpoint and take action on specific issues. Research also highlighted that many design features put the responsibility on the user to keep themself safe, which is not sufficient and can be re-traumatising for victim-survivors. Another challenge includes the lack of friction on design features, which can enable perpetrators to easily and rapidly target victims. Perpetrator sophistication in bypassing safety measures is also a key concern.
- The research pinpointed 12 design features that are widely available across platforms and pose risks to users, as noted by stakeholders. These design features illustrate that, even with good intentions behind the design, design features can often be infrequently used due to unhelpful design, limited education around features, or belief that the feature will not result in action.

## 2. Develop an understanding of how existing safety by design approaches can protect against the risk of online VAWG

- The research identified three overarching trends in designing to protect against online VAWG. This includes increasing cross-sectoral partnerships to further support and expand online VAWG safety initiatives. This links to building online VAWG specific resources and tools, which is necessary to effectively target the specific challenges and harms within online VAWG. Lastly, platforms should continue designing for online safety, and enabling user safety, privacy, and empowerment.
- With regards to safety by design, the research found that while safety by design is gaining traction across the online VAWG ecosystem, there are inconsistencies across approaches and a lack of online VAWG specific guidance. For safety by design to be effective, the stakeholders felt that shared principles are less important in comparison to dynamic principles, and urged platforms to reconsider their safety by design approach as they would their policies.

## 3. Understand potential new design approaches to ensure safety for women and girls on online platforms and services

- The research drew attention to three future-facing recommendations to ensure safety for women and girls online. First, this includes adopting a user-centric, trauma-informed approach to design, to ensure that safety is designed proactively rather than reactively, with user experience at the forefront. Second, there is a need for increased knowledge sharing across the ecosystem, to best enable collaborative approaches to protecting women and girls through design. Lastly, there needs to be increased awareness and support for victim-survivors through educational resources at multiple points across a user's experience online.

This report aims to help evidence the impact of platform design on online VAWG, as well as support women and girls around the world to safely take part in and benefit from the liberties that online spaces can bring. While this report does not cover all aspects of online VAWG, we hope that the findings and practical recommendations encourage action from online platforms to adopt design approaches that better protect women and girls online.

# Table of Contents

# 1. Introduction

Online violence against women and girls (online VAWG) is a global issue known to harm, exclude and silence women and girls' voices online. In 2014, the European Union Agency for Fundamental Rights reported that one in ten women in the European Union has experienced cyber-harassment since the age of 15.[6] Similarly in Australia, more than one in three women have experienced online violence in their working lives.[7] In 2018, Canada reported one in five women experienced online harassment and in 2016, 40% of female college students in Pakistan reported various forms of online harassment.[8] These are just a few statistics that speak to the scale of this challenge.

As well as the detrimental psychological effects on the individual, online VAWG also poses a societal problem: it undermines women and girl's ability to exercise their rights online.[9] As the offline/online distinction is increasingly blurred, online VAWG is a threat to women and girls that should be treated with the utmost urgency.

Despite substantial multilateral efforts to tackle online VAWG, more evidence is needed to assess its scale and impact. This research, commissioned by the DSIT, investigates a component of online VAWG where there is currently a gap in evidence – the impact of platform design on online VAWG. The report aims to contribute to the evidence-building around this issue for an international audience, including the Global Partnership for Action on Gender-Based Online Harassment and Abuse.

***Online violence against women and girls refers to violence conducted disproportionately against women and girls through or on cyber-enabled devices.[10]***

The research set out three key objectives which have been expanded on throughout the report:

1. **Understand** how design features of online platforms and services can enable the perpetration of online VAWG[11]

2. **Develop an understanding** of how existing safety by design approaches can protect against the risk of online VAWG

3. **Understand potential new design approaches** to ensure safety for Women and Girls on online platforms and services

As our research has focused on platform design, there are many aspects of online VAWG that are not in the scope of this report. Namely the psychological impact on women and girls, the possible motivations of perpetrators, legal interventions, and technology-facilitated domestic abuse, such as the use of the Internet of Things. While these elements are not incorporated directly into the report, they are all important to fully understand the online VAWG ecosystem.

# 2. Glossary

| Acronym | Full Term |
|---|---|
| CSAM | Child Sexual Abuse Material |
| CSO | Civil Society Organisation |
| DSIT | Department for Science, Innovation and Technology |
| EU | European Union |
| GBV | Gender Based Violence |
| G7 | The International Group of Seven (G7) |
| ICFJ | International Center for Journalists |
| LGBTQ+ | Lesbian, Gay, Bisexual, Transgender, Queer + |
| NSPCC | National Society for the Prevention of Cruelty to Children |
| STOPNCII.org | Stop Non-Consensual Intimate Image.org |
| T&S | Trust & Safety |
| UK | United Kingdom |
| UN | United Nations |
| US | United States |
| VAWG | Violence Against Women and Girls |
| WWWF | World Wide Web Foundation |

# 3. Methodology

The project addressed five key research questions:

1. What design features do platforms currently use?
2. How do these design features impact the risk of online VAWG for users of the platform/service? How does the impact differ according to different protected characteristics of the women being targeted?
3. What 'safety by design' features do tech companies currently use to protect against online VAWG?
4. What current design features are the most successful in the protection against the risk of online VAWG? How does the impact differ according to different protected characteristics of the women being targeted?
5. What potential design features would be useful to consider implementing to further protect against the risk of online VAWG?

Recognising that online VAWG is a global, pervasive challenge, the research was scoped to effectively target these questions in three key ways to ensure the biggest impact:

- The research placed an emphasis on large-scale social media and online dating platforms when researching platform design, particularly Facebook, Instagram, Snapchat, TikTok, X (formerly Twitter), YouTube, Bumble and Tinder.
- The research took an international lens, but with a specific focus on the Global North due to the majority of the platforms, as well as their respective design and Trust & Safety teams, being based in the US.[12] In addition, the researchers' existing networks with civil society organisations and platforms are predominantly based in the UK, European Union (EU) and United States (US), which guided stakeholder outreach and interviews.
- The research was not segmented by different types of online VAWG harm types, but rather considered online VAWG as a whole. However, we recognised particularly pervasive harm types in our proto-personas, which fed into our design recommendations.

In scoping the research this way, we recognise that the findings are not representative of all users and victim-survivors across platforms and their respective experiences internationally.[13] Instead, the research is reflective of the perspectives of the 32 stakeholders interviewed for the research and the literature review and may pose limitations due to the scope of their research or experiences. However, we believe that the research provides a strong evidence base on the current challenges, successes and opportunity areas around platform design in preventing online VAWG today.

To capture the lived experiences of victim-survivors, we worked closely with a Civil Society Expert Group. This group consisted of five UK-based and one US-based academics and civil society organisations that work with online VAWG. They were interviewed and provided feedback and advice on some of the material featured in this report, including design short-list. The members were:

- Professor Clare McGlynn, Durham University
- Chayn
- Glitch
- Refuge
- Suzy Lamplugh Trust
- World Wide Web Foundation (WWWF)

This report took a three-pronged research approach, consisting of landscape mapping, stakeholder engagement, and design review. This approach has been briefly described below and a detailed methodology, including an explanation of specific harm types, has been included in the Appendix.

1. **Landscape Mapping:** The purpose of the landscape mapping was to understand key trends across the online VAWG ecosystem and included a literature review and an evidence review. Collectively, the landscape mapping aimed to address research questions 1, 2 and 3.

The *literature review* analysed over 100 academic, civil society and government sources to identify trends impacting the online VAWG ecosystem at large. Each piece of literature was assessed on its methodology, quality of analysis and sources to ensure sufficient consistent quality for inclusion. The findings from the *literature review* have been outlined in the key trends section (page 13).

The *evidence review* investigated and aggregated platform design approaches and respective policies/community guidelines. This informed the design feature longlist, which was a consolidated list of design features found while investigating different platform design approaches to tackle online VAWG. The findings from the *evidence review* have been integrated across the Objectives section (page 25).

**2. Stakeholder Engagement:** The purpose of the stakeholder engagement was to validate findings from the literature review, generate further evidence, and refine the longlist of design features. This phase of the research aimed to address research questions 1, 2, 3 and 4.

First, a *stakeholder mapping* exercise was conducted to identify all relevant stakeholders with expertise in online VAWG and/or platform design. To do so a list of over 200 stakeholders active in the online VAWG ecosystem was consolidated, including academics, civil society organisations, government bodies, platforms, and other stakeholder groups. This list was then prioritised through an interest x influence matrix mapping, to determine which stakeholders were most relevant to engage for stakeholder interviews. The interest x influence scoring matrix is included in the extended methodology in the Appendix section.

The stakeholder map guided *stakeholder outreach*, with priority on social media, online dating and civil society organisations, including support services to victim-survivors. Over the research period, 10 platforms and 16 civil society organisations agreed to be interviewed. Additionally, 6 subject matter experts across academia, online safety tech companies, and think tanks agreed to be interviewed. These stakeholders were based in Belgium, Denmark, India, Poland, Australia, the UK, and the US.

Stakeholder insights and evidence review findings were also leveraged to refine the design feature longlist to a shortlist of 12 features, based on two metrics:

- **Representativeness:** Defined by whether the design features are currently available across multiple platforms. As noted prior, eight key platforms were considered: Facebook, Instagram, Snapchat, TikTok, X, YouTube, Bumble and Tinder.
- **Risk to Users:** Defined by whether the design feature is considered to pose a risk to users, based on stakeholder interviews.

These design features have been further elaborated on in the Objectives section (page 25).

**3. Design Review:** The purpose of the design review was to consolidate the findings from the landscape mapping and stakeholder engagement, and consider potential future design features. This phase of the research aimed to address research question 5. The design review consisted of creating four *proto-personas*, visualising user journeys.

A proto-persona is a fictional representation of a typical user within a segment of the overall user group. The proto-personas were based on challenges women and girls face online as identified in the literature review and informed by interviews with civil society organisations.

To note, this research was carried out between October 2022 and March 2023. As such, the findings reflect the landscape at that moment in time. We recognise that certain aspects may have evolved since then.

# 4. Key Trends Influencing Online VAWG

As this section illustrates, a substantial amount of research has been dedicated to understanding the types of online VAWG experienced today, their impact and potential solutions. However, there is a gap in evidence around platform design specifically, and its potential impact on online VAWG. This report seeks to address this gap. To do so, it is critical to first understand the current online VAWG landscape and the trends impacting it.

This section outlines the findings from the literature review and provides a foundational understanding of the current online VAWG ecosystem. Over 100 pieces of literature were reviewed to generate the findings in this section, covering ten trends across three key themes. These can be found in the bibliography.

**Diagram 1: Key Online VAWG Trends by Theme**

### Landscape Developments

**i.** Governments and intergovernmental organisations across the world are devoting more attention to online VAWG, but with limited focus on design

**ii.** Reports of online VAWG increased significantly during the pandemic

**iii.** The lack of diversity in the leadership of online platforms, which represents the interests of women and girls, poses a systemic challenge to countering online VAWG

### Impact on Women and Girls

**iv.** Women and girls with certain overlapping protected characteristics experience greater prevalence of online violence

**v.** Online VAWG is part of a continuum of violence that women face throughout their lives

**vi.** Victim-survivors are not supported in identifying, documenting, and reporting online VAWG

### Technology

**vii.** Some design features can both increase and decrease the risk of online VAWG depending on context and user intentions

**viii.** Cross-platform online harassment poses challenges to evidencing and intervening against online VAWG

**ix.** The complexities and potential biases of content moderation pose a technological challenge to moderating online VAWG

**x.** Emerging technologies will bring new and unknown harms, requiring adjusted design approaches to mitigate the risk ofonline VAWG

The trends found in the technology theme are most closely related to the research objectives and focus on platform design. However, to fully understand the impact of platform design it is critical to take a wider look at trends impacting the ecosystem at large. As a result, this section gives a comprehensive overview of trends impacting the ecosystem today, including landscape developments, impact on women and girls, and technology.

## The UK's Online Safety Act

The UK's Online Safety Act will place a requirement on services in scope to proactively tackle the most harmful illegal content, much of which disproportionately affects women and girls.

This includes intimate image abuse, harassment, stalking, controlling or coercive behaviour and extreme pornography. In addition, all services will have duties to remove other illegal content when it is flagged to them.

Measures companies can take to tackle this type of content will be set out in Ofcom's codes of practice - Ofcom must consult with the Victim's Commissioner and Domestic Abuse Commissioner ahead of drafting these codes, to ensure the voices and views of women, girls and victims, are reflected.

Services that are likely to be accessed by children are required to take measures to protect them from harmful content which doesn't meet the criminal threshold. This includes preventing them from seeing the most harmful content such as pornography and protecting children from other types of harmful content such as bullying and content that is abusive or incites hate, such as misogynistic content.

The Act will place duties on services over the designated threshold (Category 1 services) to remove certain types of legal content that is prohibited in their terms of service and will require these platforms to put in place user empowerment tools, enabling users, including women to have greater control over their online experience.

These tools, when applied, will reduce the likelihood of women encountering abusive content targeted at them or others on the basis of their sex, or will alert them to the nature of it. Women will also be able to filter out content from unverified users.

Through Ofcom's media literacy duties in the Act, it will need to increase the public's awareness of the ways in which it can protect groups that disproportionately face harm online, such as women and girls.

Additionally, the Act requires Ofcom to produce guidance that summarises in one clear place, measures that can be taken to tackle the abuse that women and girls disproportionately face online. This guidance will ensure it is easy for platforms to implement holistic and effective protections for women and girls, across their various duties. Ofcom will consult on the draft guidance in February 2025.

# Landscape Developments

## i. Governments and intergovernmental organisations across the world are devoting more attention to online VAWG, but with limited focus on design

On a global level, there are a number of initiatives that show an increasing international focus on tackling online VAWG. The annual meeting of UN countries at the 67th session of the Commission on the Status of Women, that took place in March 2023 was focused on the theme of technology and innovation.[14] The *UN Women Technology and Innovation for Gender Equality Action Coalition Blueprint* has outlined one of its visions: "By 2026, a majority of countries and tech companies demonstrate accountability by implementing policies and solutions against online and tech-facilitated gender-based violence (GBV) and discrimination".[15] Similarly, leaders of the G7 have made a series of pledges around violence against women and girls in digital contexts, including mobilising international actors around the issue.[16] Other global initiatives by the Council of Europe, Global Partnership for Action on Gender-Based Online Harassment and Abuse, UN Women and the European Union attest to a general increased interest in online VAWG.[17] Together, these initiatives are indicative of heightened international cooperation and awareness around this challenge, and the importance of this report in supporting this momentum.

National governments are also increasingly looking at online VAWG. A few examples include the UK government's Online Safety Legislation (see box on the UK's Online Safety Act) and commitment to halve VAWG (including online) over the next decade.[18] Australia's eSafety Commissioner has conducted extensive work on online VAWG, with a particular focus on evidencing the challenges of different communities and providing Australian women and girls with guidelines to stay safe online.[19] The US Task Force on Online Harassment and Abuse also calls attention to the disproportionate effect of online abuse on women.[20] Governments have also targeted online VAWG through specific harm types. South Korea's Digital Sex Crime Task Force directs attention to the issue of image-based abuse, as does Ireland through its Harassment, Harmful Communications and Related Offences Act (2020), which created two new offences for intimate image-based abuse based on whether this is done with the intent to cause harm or not.[21] Furthermore, in 2018 Peru introduced legislation to curb offences related to harassment, sexual harassment, sexual blackmail and distribution of intimate images, through digital channels without consent.[22] In the same year, France introduced cyberbullying against women and girls as a new criminal offence.[23] These countries are a small group of governments working to tackle online VAWG.[24] While governments may be targeting online VAWG in different ways, whether community or harm-specific, the momentum alongside intergovernmental organisations is encouraging.

While these early initiatives show signs of promise, few focus on the role of platform design. The connection between design and online VAWG is more frequently explored by civil society organisations through independent research. This includes the World Wide Web Foundation's work on design prototypes to mitigate online VAWG, Chayn and End Cyber Abuse's project *Orbits* on design interventions to online VAWG, and others.[25] This report seeks to further evidence this space and elaborate on how platform design can exacerbate and mitigate the risk of online violence against women and girls.

## ii. Reports of online VAWG increased significantly during the Covid-19 pandemic

Reports conducted by government, intergovernmental and civil society organisations have also highlighted that online VAWG intensified during the pandemic.[26] For example, in the UK, Refuge reported a 97% increase in complex tech abuse cases from April 2020 to May 2021 compared to the first three months of 2020 in the UK.[27] With regards to specific harm types, the Suzy Lamplugh Trust found a significant rise in online stalking during the pandemic, with 100% of cases reported to the National Stalking Helpline involving a cyber element.[28] The pandemic also resulted in an increase in intimate image-based abuse. The Australian eSafety Commissioner noted a 210% increase in reports during the pandemic in Australia.[29] In the UK, the Revenge Porn Helpline reported an increase in cases by over 40% between 2020 and 2021 and reports of sextortion doubled in 2021 compared to the previous year.[30]

It is also critical to acknowledge the disproportionate impact of the pandemic on specific communities of women and girls. For example, in the UK, End Violence Against Women and Glitch found that black women and women from ethnic minorities in the UK were particularly exposed to online VAWG during the pandemic and reported the amount of abuse as being worse than before Covid-19.[31] Although these reports give a clear indication of the rise in online VAWG during the pandemic, it is important to note that both offline and online VAWG is chronically underreported, meaning the prevalence of online VAWG cannot fully be determined.[32] Moreover, many women and girls are unaware of being victims of online VAWG.[33] For example, women and girls subjected to intimate image-based abuse are not always aware of their images being shared.[34] This may contribute to the underestimation of reports of online VAWG. Nonetheless, the existing data illustrates that the heightened challenges in part stem from increased exposure and reliance on technology during the pandemic. Critically the pandemic not only brought more women and girls' lives online but perpetrators' as well.[35]

## iii. The lack of diversity in the leadership of online platforms, which represents the interests of women and girls, poses a systemic challenge to countering online VAWG

The underrepresentation of women from diverse backgrounds in major online platforms, especially in product design, poses a systemic challenge to mitigating online VAWG through design.[36] Reports suggest this underrepresentation results in the embedding of existing biases and inequalities in technology products.[37] Moreover, women and girls suggest their needs are often not heard by platforms, potentially as the lack of diversity makes specific needs easier to ignore or deprioritise.[38] It is important to address these biases and inequalities when we formulate design interventions and continuously work to improve the representation of women on tech platforms. One stakeholder framed it as the need to improve decision-making processes which meaningfully represent the interests of all users, particularly women from different backgrounds. An important part of this is representative workforces and a diversity of perspectives in leadership.

The lack of diversity also runs along geographic lines. Platform headquarters may determine focus areas, resourcing, and which users are considered the "default user". Most platform policies and product changes tend to be designed in the Global North to be applied to a global user base.[39] This puts women and girls in the Global South at a disadvantage, as policy and product changes may not reflect cultural, linguistic, national or other contextual differences or needs.[40] The lack of geographic diversity of headquarters is also reflected in the platforms' allocation of resources to different areas of their business. For example, research has shown that one large platform dedicated the majority of their global budget for classifying misinformation to the United States, despite US users making up a minority of their daily user base.[41] In a report on the online lived experiences of women living in Sub-Saharan Africa, researchers highlighted that: "despite serving a population of 1.2 billion, the number of staff within these companies dedicated to and working from Africa is negligible", resulting in a neglect of processing African languages and cultural contexts.[42] The impact this can have on women and girls is substantial as it can affect how they decipher platform policies, use safety tools or recognise online VAWG when they are experiencing it.[43] To create a truly global approach in tackling online VAWG, the evidence recommends that the current gap in the representation between the Global North and Global South needs to be addressed.[44]

# Impact on Women and Girls

## iv. Women and girls with certain overlapping protected characteristics experience a greater prevalence of online violence

Online VAWG does not impact all women and girls in the same way.[45] This particularly includes young women and girls, the LGBTQ+ community, women with disabilities, black women, women from ethnic minorities and indigenous women, as outlined in further detail below:

- **Young Women and Girls:** Both the impact and prevalence of online VAWG on girls is well documented. In a global research study conducted by Plan International, they reported that more than half of the girls that they surveyed had experienced online VAWG.[46] However, girls may experience online VAWG in different ways based on their age. Research by Thorn indicates that girls aged 9-12 experience an overall potential harm rate of 41% versus girls aged 13-17 with a harm rate of 58%.[47] Between the two age ranges, there is a large jump, over double in fact, in the potential online harm of sexualised interactions: 9-12-year-old girls experience this harm at a rate of 21%, whereas girls aged 13-17 experience this harm at a rate of 46%.[48] It is critical to note that this does not mean that girls outside of this age range (9-17) are not targeted with online VAWG. The Internet Watch Foundation reported that 97% of reported Child Sexual Abuse Material online in 2021 was of girls, inclusive of toddlers and babies (0-2 years old), and saw a threefold increase from the previous year in self-generated imagery of 7-10 year old girls.[49] In comparison to adults, research shows that younger women (born 1981-2002) are more likely to experience online violence versus older women (born 1946-1980).[50] These reports illustrate that girls, even below the set age requirements of platforms (typically 13), are experiencing online VAWG, and need to be considered as a community with specific needs when designing and embedding safety features online.
- **LGBTQ+:** Women and girls in the LGBTQ+ community are also, particularly at risk of online VAWG.[51] 75% of women identifying as LGBTQ+ reported experiencing online abuse, compared to only 33% of non-LGBTQ+ women.[52] Similarly, minors identifying as LGBTQ+ are two times more likely to experience harmful online sexual interaction compared to non-LGBTQ+ minors.[53] This exemplifies the need for platform safety solutions specific to the LGBTQ+ community, and specific groups within the community, while also considering the intersection of LGBTQ+ with other identities which may put users at increased risk of online VAWG.

- **Disabilities:** Women and girls with disabilities have also been reported to experience increased harm, particularly harassment, versus women and girls without disabilities.[54] In an international study, 14% of girls self-identifying with a disability reported experiencing online harassment and said the harassment was due to their disability.[55] Disabled women and girls lack accessible resources and struggle with mistrust in reporting processes where they might face discrimination. As a result, disabled women and girls are more likely to stop using a platform and not report harm, indicating the true scale of online violence towards women and girls with disabilities may be underreported and undocumented.[56]
- **Ethnic minority women and girls:** Women and girls from ethnic minority groups experience an increased risk of online harm, including violence against women and girls. Ofcom's Online National Report highlights women from ethnic minority groups in the UK were three times as likely as white women to have seen or experienced sharing of intimate images without consent and four times as likely to have received an unwanted or unsolicited sexual/nude image or video.[57] Discrepancies in experience based on race are also reported in the US, where a noted 53% of minors identifying as African American and Hispanic/Latino reported experiencing harmful online experiences versus 48% of their white counterparts.[58] The types of harm that women from ethnic minority groups experience also vary, with online gendered disinformation being both greater in volume and more severe in language for women from ethnic minorities than for white women.[59] Moreover, research suggests that women of colour are more likely to modify their online behaviour because of online VAWG. 87.5% of women from ethnic minority groups in a UK survey done by End Violence Against Women and Glitch reported modifying their behaviour online as a result of the abuse.[60]

Women and girls in public positions, such as journalists, politicians, celebrities and activists, are at particular risk of online VAWG according to existing research.[61] Women in public life are also more vulnerable to certain types of online VAWG like dog-piling, cross-platform harassment and gendered disinformation, which seek to undermine women's political and democratic participation. Equally, it may be harder for these victim-survivors to leverage safety interventions, like making their accounts private, as they often rely on their online presence for work.[62]

It is critical to note that the identities included above are not comprehensive, but instead reflective of available research on online VAWG and protected characteristics. Further research and evidence are needed to draw conclusions on the harms experienced by specific identities or intersection of identities that are not listed above.

## V. Online VAWG is part of a continuum of violence that women face throughout their lives

Online VAWG is part of a continuum of violence that women and girls face in their lives and cannot be thought of as separate from offline VAWG.[63] Critically, the psychological impact of online VAWG on victim-survivors is similar to that of offline VAWG.[64] A Women's Aid survey showed that 85% of victim-survivors of domestic abuse experience abuse from a partner or ex-partner online, which is part of a pattern of harm that they also experience offline.[65] In the US, 97% of domestic violence support services report that perpetrators used technology as part of their abusive behaviour.[66] This is not only an issue for the Global North: One in three women experiencing online VAWG in Iraq, Jordan, Lebanon, Libya, Morocco, Palestine, Tunisia, and Yemen report that some or all of their experiences of online VAWG moved offline.[67] The close relationship between online and offline may increase the risk of online VAWG for victim-survivors, as it can open additional avenues for harm and re-traumatisation. While the focus of this report is online VAWG, it is important to remember that this is often only one side of the story of abuse that many women and girls are experiencing.

## VI. Victim-survivors are not supported in identifying, documenting, and reporting online VAWG

Identifying online VAWG can be difficult as there are discrepancies in platforms' policies and languages. Moreover, harms can often overlap, which further complicates the process to identify, document and report online VAWG.[68] With regards to documenting online VAWG, this can be re-traumatising for victim-survivors and sometimes impossible (e.g. with reporting disappearing content), which both deter victim-survivors from reporting.[69] Multiple studies have flagged challenges with regard to reporting. In a global survey on the prevalence of online VAWG, only one in four women who experienced abuse reported this abusive behaviour to the platform, meaning 75% of harm was undocumented and not addressed.[70] One of the major reasons that victim-survivors cite for not reporting online VAWG is not believing that it will help.[71] This is further aggravated by some platforms having a poor record of taking action on reports of abuse and few to no online VAWG specific policies or community guidelines.[72] Another challenge with reporting comes as a result of limited digital literacy. In a survey of Aboriginal and Torres Strait Islander women in remote areas in Australia, victim-survivors often did not know how to identify online VAWG due to a lack of education in online harms.[73]

Due to the difficulty in identifying, documenting and reporting harms, blocking perpetrators is favoured as an alternative response to online VAWG, especially for young users.[74] In a survey of 13-17-year-olds in Denmark, Hungary and the UK, 82% of surveyed young people said they would block the people involved if they experienced online sexual harassment whereas only 39% said they would report it to the social network.[75] A survey on online dating by YouGov on behalf of the Suzy Lamplugh Trust reaffirmed this, stating that the majority of the Match Group's online daters who have had concerns for their safety simply block their perpetrators' profiles.[76] This is partially due to concerns around reporting: 15% of daters felt their report would not be acted upon by the service provider, 12% felt there is not an easy way to report the concern on the dating website and 7% were too embarrassed to report it.[77] Without effective reporting, users may continue to block, which could limit evidence-gathering opportunities on women and girls' experiences online.

# Technology

## vii. Some design features can both increase and decrease the risk of online VAWG depending on context and user intentions

Several academics and civil society organisations argue that design features can have a double-edged impact.[78] While some design features have a clear exacerbating impact, there are many design features that have the potential to both exacerbate and mitigate the risk of online VAWG depending on how they are used.[79] One example is the creation of anonymous accounts, which can offer privacy to women and girls that may wish to protect their identity, such as the LGBTQ+ community.[80] In fact, two in five LGBTQ+ girls and young women said anonymous accounts helped them feel safe online.[81] On the other hand, research also highlighted how anonymity can contribute to the impunity enjoyed by perpetrators of online VAWG and can perpetuate online VAWG as perpetrators can create anonymous accounts to continue their abusive behaviour even after being blocked or suspended.[82] This reiterates that design features must take into account the experience of all users to ensure features are used for their intended purpose and not to harm others.

## viii. Cross-platform online harassment poses challenges to evidencing and intervening against online VAWG

Similar to how violence against women and girls can be acted across online and offline avenues, known as the 'continuum of violence', online VAWG is often carried out across multiple online platforms.[83] For victim-survivors, evidencing online VAWG across different platforms is challenging - both in terms of scale and the psychological impact of being attacked through multiple avenues at once.[84] An example of this is a mob-attack, where the victim-survivor may have to report thousands of individual messages on multiple platforms, which can be traumatising. Despite being a pervasive issue, especially for women journalists, politicians and activists, platforms are often slow to intervene in cross-platform harassment - if they do so at all.[85]

The speed of transmission and ease of sharing content with large numbers of people may contribute to cross-platform online harassment.[86] A study on young women's experiences of online violence in South India showed that 31% of women experiencing online sexual harassment reported experiencing cross-platform harassment, with perpetrators jumping between platforms.[87] For example, when a victim-survivor turned down a perpetrator on a dating platform, the perpetrator then harassed the victim-survivor with repeated requests to connect on another platform.[88] This stresses the need for cross-platform intervention and collaboration to effectively combat online VAWG.

## xi. The complexities and potential biases of content moderation pose a technological challenge to moderating online VAWG

It is reported that artificial intelligence (AI) brings inherent biases toward some communities, reflecting wider inequalities in the offline world, as notably researched and evidenced by the works of Safiya Umoja Noble and Joy Buolamwini.[89] For example, the use of AI in content moderation can lead to the embedding of existing biases against certain women and girls, as research indicates that some algorithms are biassed against women with certain overlapping protected characteristics, such as black women.[90]

The complexities and nuances of online VAWG are also difficult to pick up by content moderation, both human and automated.[91] In particular, this is challenging when perpetrators use coded language (e.g. memes, emojis or other creative use of media) to inflict harm that requires an acute awareness of context and culture.[92] The Wilson Centre coined the term 'malign creativity' to refer to "the use of coded language; iterative, context-based visual and textual memes; and other tactics to avoid detection on social media platforms", used to harass or abuse victims. For example, a perpetrator

threatening a girl or a woman by sending a picture of a film title in which a woman is being murdered.[93] Without knowledge of what this means, content moderators would not be able to recognise it as online VAWG. This points to the need for educational training for human moderators across cultures and contexts and investing an active effort to combat existing biases in algorithmic moderation to capture the complexities of online VAWG.

## X. Emerging technologies will bring new and unknown harms, requiring adjusted design approaches to mitigate the risk of online VAWG

Emerging technologies, including virtual reality, augmented reality, mixed reality and haptics, create new forms of online VAWG, which require a future-facing approach to tackling online VAWG through design.[94] We have already seen this in online gaming and the metaverse where women and girls have reported sexual harassment and abuse.[95] As these technologies become more mainstream, the risk of new types of online harm may increase. In particular, immersive haptic technologies that simulate the feeling of touch, deep fakes (the majority of which are non-consensual, sexual portrayals of women) and deep nudes/nudification tools are all examples of how new technologies can bring new cyber-dependent harms.[96]

Emerging technologies also reveal the need for continuously updated platform interventions. With the rise in video conferencing, the pandemic saw an increase of "zoom-bombing" as an emergent type of online VAWG.[97] Moreover, as live streaming became popular, it brought new considerations for child safety.[98] In 2018, the National Society for the Prevention of Cruelty to Children (NSPCC) found that 6% of children that use live streaming have received requests to change or remove their clothes.[99] However, research shows that platforms rarely have real-time moderation mechanisms for recognising child sexual abuse in live-streamed content.[100] To future-proof responses to online VAWG that reflect emerging technology, some civil society groups and academics recommend that online platforms have a team dedicated to tracking new developments within online VAWG.[101]

This section has identified ten key trends across the online VAWG landscape, which sets the scene as we move into the details of platform design and its impact on online VAWG.

# 5. Platform Design and Online VAWG

As outlined in the Key trends section, there is an evidence gap on the impact of platform design on online VAWG, which this research seeks to fill. This section synthesises the research findings from the landscape mapping and stakeholder engagement, broken down by the three project objectives focusing on the impact of platform design:

1. **Understand** how design features of online platforms and services can enable the perpetration of online VAWG
2. **Develop an understanding** of how existing safety by design approaches can protect against the risk of online VAWG
3. **Understand potential new design approaches** to ensure safety for women and girls on online platforms and services

The focus on these three objectives facilitates a comprehensive overview of current challenges and successes with design features in tackling online VAWG, and future opportunities. Each subsection is split into overarching insights on the objective, before diving into specific findings on design features.

# Objective 1: Perpetration of online VAWG through design

The first objective seeks to understand how design features can enable the perpetration of online VAWG. Through the literature review and stakeholder interviews, five overarching challenges with design features were identified, including a lack of platform policies/ community guidelines addressing online VAWG, user responsibility, friction, perpetrator sophistication and data collection. These have been further detailed below, before elaborating on challenges with specific design features in enabling the perpetration of online VAWG.

It is important to note that while the stakeholders who participated in the research were representative of industry, civil society organisations (CSO's) and academics, the findings from this section are particularly reflective of CSO feedback. This is due to their unique insight into the communities they represent, and specifically the lived experiences of victim-survivors of online VAWG. Throughout this section the distinction between the two stakeholder groups has been noted to ensure transparent evidence.

## Overarching challenges with perpetration of online VAWG through design

### 1. Lack of platform policies guidelines addressing online VAWG

Out of the eight platforms reviewed, none had policies that directly target online VAWG at the time that this research was carried out.[102] Instead, most platforms have policies that may address specific types of online VAWG, for example, hate speech, but do not explicitly state how their policies protect women and girls specifically. Multiple stakeholders interviewed, across both CSO's and industry, noted this is a significant challenge to mitigating online VAWG as it makes it unclear to women and girls how they are supported if experiencing online VAWG, and makes it harder to sanction perpetrators.

### 2. User responsibility

Safety features in use by platforms often place the onus on women and girls to take responsibility for their own safety. Stakeholder interviews with CSO's highlighted that frequently used design features to respond to online VAWG, such as blocking, reporting, and muting, place undue responsibility on users who have been targeted.[103] These burden victim-survivors with their own safety, and are primarily reactive, rather than proactive.

Placing the responsibility for safety on users can have significant consequences for young women and girls who might be unaware of available safety features or how to use them.[104] A child safety expert shared that when children experience online harm "99% of the time they don't know what to do". This reinforces the need for platforms to appropriately safeguard women and girls, while also empowering them to take action when confronted with online VAWG.

Additionally, this reliance on user responsibility can further exacerbate online VAWG in specific circumstances. Stakeholders across sectors noted that some safety features may be ineffective for women in public positions who experience large volumes of online violence. For example, reporting is often designed to cover one or a few instances of content or conduct, in greater detail. However, if a woman receives thousands of harmful messages a day, this design might be less effective and might contribute to the burden of stress on that user.

### 3. Friction

Design features with little friction can enable the ease and speed of online VAWG. 'Friction' on online platforms refers to the elements of design that slow the user down to accomplish their task.[105] Stakeholders, predominately CSO's, noted that minimal friction often entails a smoother user experience and aligns with many platforms' business objectives of increasing engagement. However, design features with little to no friction can also exacerbate the risk of online VAWG occurring, mainly by enabling perpetrators to commit harm more easily, more quickly and with greater reach. For example, low friction in the registration process on a platform may enable a perpetrator to easily set up a fake account or to create new ones if their original profile is reported. This is particularly harmful as perpetrators sometimes use disposable and anonymous 'burner accounts' to post abusive content.[106] Research suggests that making it more difficult to set up 'burner accounts' could reduce the amount of abuse.[107]

Another critical example includes limited friction in direct messaging. This enables perpetrators to easily gain access to users that they are not otherwise connected with, while also allowing them to quickly share harmful images or videos. The latter is particularly relevant for image-based abuse, as images can be shared rapidly across platforms through frictionless sharing. This can exacerbate the reach, scale and escalation of harm.[108]

### 4. Perpetrator sophistication

Perpetrators often find ways to circumvent safety features. An academic expert from the Stanford Internet Observatory argues that regardless of the intention behind design features they will "be misused and abused". This was specifically reinforced in the context of child safety by one of our interviewees who shared "if there is potential for contact, there is potential for harm". This emphasises the need to deter perpetrators while also protecting victim-survivors, and ensuring that design is always user-centric and trauma-informed.

### 5. Data collection

The limited data collection by platforms around the experiences of women and girls poses an obstacle to developing design features that specifically target and mitigate online VAWG. Platforms reportedly collect little to no data on online harms that is sex disaggregated as highlighted by The World Wide Web Foundation, which noted that "platforms don't have independently verifiable data" on online harms against women. Sex disaggregated data on online harms would enable platforms to better understand the experiences of women and girls on platforms to build and measure the effectiveness of design features that mitigate the risk of online VAWG.[109] However, interviewed CSO stakeholders also argued that it is critical that, if actioned, that data is not shared or misused for other, non-safety related purposes, such as profiling users.

## Perpetration of online VAWG through specific design features

As outlined in our methodology, the research focused on five key questions, including *What design features do platforms currently use?* With an expansive ecosystem of social media and online dating platforms, the list of design features available across platforms is extensive. This research identified over 50+ design features through our evidence review exercise.

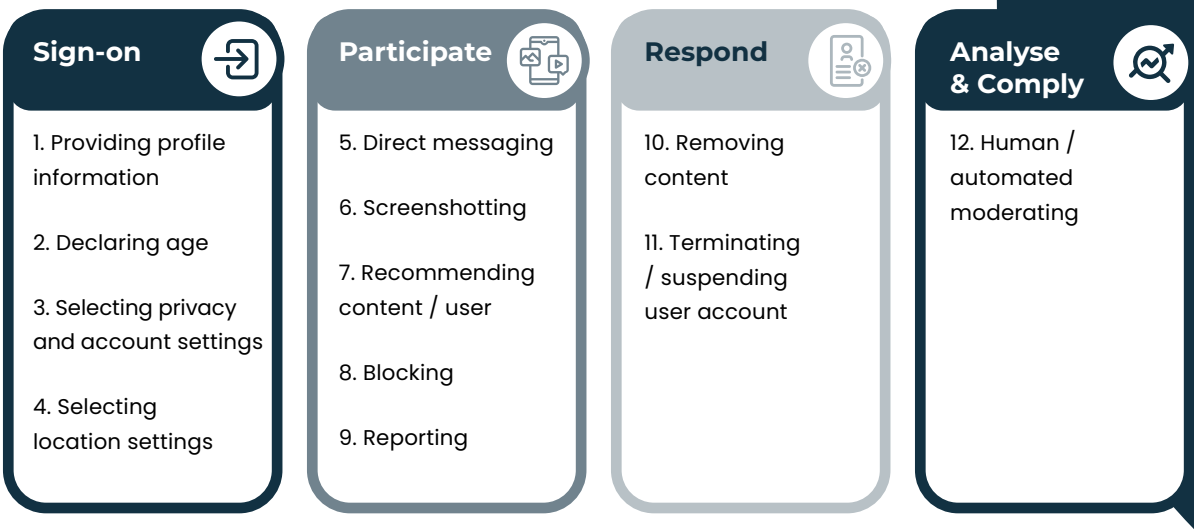
***For the purpose of this report a design feature is defined as a touchpoint which enables the user or platform to take an action at various points in the user journey that affects the user experience.***

For this research, the value in identifying design features is in recognising their risks and being able to better design for future user safety. As a result, we prioritised the 50+ design features and created a shortlist of 12 design features that currently present a risk to users. The shortlisting was based on two metrics:

1. **Representativeness:** Defined by whether the design features were currently available across multiple platforms at the time of the research. As noted in the methodology, eight key platforms were taken into consideration to assess this metric: Facebook, Instagram, Snapchat, TikTok, X, YouTube, Bumble and Tinder.
2. **Risk to Users:** Defined by whether the design feature is considered to pose a risk to users, based on stakeholder interviews.

This resulted in a shortlist of 12 design features that were most prevalent and have the potential to pose risks to users dependent on design across major social media and online dating platforms. It is critical to note that all the design features we identified have the potential to inflict harm if misused. However, the features identified below/in the shortlist were repeatedly flagged by CSO interviewees as currently risky for users. The shortlist of design features is visualised across the user journey in Diagram 2 below, and the findings based on stakeholder interviews are detailed following.

**Diagram 2: Shortlisted Design Features across the User Journey[110]**




It is important to note here that 'algorithms' as a whole were not considered as a design feature as they are tools that may make up a design feature, running in the backend of the platforms. Algorithms have been shown to sometimes promote hateful content, but the diversity of use cases and implementation make them difficult to study as a discrete category.[111] In addition, due to a lack of transparency over how algorithms operate, there was difficulty in evidencing the specific role of algorithms in mitigating or exacerbating the risk of online VAWG. However, we have included 'recommending content/ user' which covers aspects of algorithms but is more user-facing.

## Sign-on

### 1. Providing profile information

'Providing profile information' refers to the identity details users input when setting up their profiles. This includes but is not limited to name, age, gender, bios and images. Stakeholder interviews across sectors flagged that dating platforms in particular struggle with fake and spam accounts. To ensure users are who they say they are, many platforms require users to provide some type of user verification. This can vary from requiring ID to asking for a selfie at sign-up to match with profile details, which have varying levels of reliability. Most platforms also ask for emails and phone numbers, which can be used to track bad actors or individuals with multiple accounts.

However, this was not consistent across platforms: some platforms require limited information, enabling perpetrators to have a frictionless account set-up experience. This is important as online VAWG often stems from fake or 'burner accounts'.[112]One industry interviewee also noted that all trust and safety technologies and tools, including identity verification, can "provide a false sense of security", as bad actors can often find a way around safety measures. Without appropriate ways for users to verify their identities, women and girls may interact with people who may not be who they say they are, exacerbating the risk of harm, specifically impersonations and scams.

Several CSO and industry interviewees also flagged the downsides of requiring identity verification, through official documents like ID, for particular communities of women and girls. In particular, users from the LGBTQ+ community, refugees or displaced persons or sex workers, may be less willing to entrust platforms with identity documents. For the trans community specifically, requiring identification may be sensitive for individuals who are transitioning, particularly when profile information, such as name or sex, cannot be changed once documented. This reinforces the need to safeguard users' privacy and respect individual experiences when implementing safety features. As one industry interviewee put it: "privacy is safety for vulnerable people''.

## 2. Declaring age

> ***"It's [self-declaration] a very weak safety measure."*** *Empowering Children Foundation*

Ideally, age assurance should ensure that users meet platforms' minimum age threshold and that children and young users are not exposed to any age-inappropriate content. Our research showed that the most common form of age assurance is self-declaration, meaning the user inputs their birthday to declare their age or simply ticks a box to 'declare' that they are over a certain age (usually 13 or 18). Across representative interviewed groups, we heard that this form of age assurance is ineffective, as it is easy to circumvent and puts girls at risk. For age assurance with hard identifiers, such as ID, stakeholders across sectors also noted risks similar to those in 'providing profile information', with one academic expert noting that "age verification makes me nervous" given the need to provide sensitive personal data and risk to child privacy. However, CSO stakeholders also emphasised that even if age assurance was effective, it is critical for the entire platform to be designed safely for girls. To appropriately safeguard children online through age verification techniques, CSO's stressed that self-declaration is not sufficient and more advanced technologies, accompanied by age-appropriate design within platforms, are necessary.

## 3. Selecting privacy and account settings

> ***"Some users will just use the settings they get from the platform."***
> *Empowering Children Foundation*

'Privacy and account settings', where users can set filters on who can contact them and see their profile/content, is a major topic in the safety of women and girls. CSO stakeholders highlighted that the absence or complexity of such settings may exacerbate the risk of online VAWG. Making privacy and account settings accessible, both to find and understand, is especially important for girls who may be less aware of the existence of these tools. CSO stakeholders flagged that education around privacy and account settings on platforms is critical and that this is currently lacking. This is evidenced by a consultation conducted by the WWWF, where less than half of young women had set their privacy settings to high, likely as a result of a combination of a lack of awareness, education and language barriers if privacy settings are not available in local languages.[113] Additionally, stakeholders, particularly support services for victim-survivors, noted that if privacy settings are written in complicated language or are badly translated, women and girls who do not speak the primary language of the platform are at a disadvantage. To mitigate this challenge, academic experts and CSO representatives encouraged platforms to set privacy settings to high by default and stressed the responsibility of platforms to explain the implications of removing those settings. CSO stakeholders argued that there is a strong need for accessible and rigorous privacy settings to safeguard women and girls.

## 4. Selecting location settings

'Sharing location', especially sharing live location, was flagged by CSO interviewees as a design feature that exacerbates the risk of online VAWG - especially if platforms automatically share a user's location by default. Location sharing on platforms varies. For example, at the time this research was executed, one big platform showed the user's location globally unless they choose to hide it through using a specific tool. Alternatively, dating apps often have a location feature which allows you to see other users in or close to your location. Online VAWG harm experts and support services noted the risks of sharing locations are especially prevalent for online stalking. Refuge emphasised the risk that location sharing poses on dating apps, as it may put women who live in small, remote communities at particular risk, given it is easier to pinpoint their location, which may in turn facilitate offline VAWG. As with privacy and account settings, if location settings are hard to find or explained in inaccessible language, it can be challenging for girls and young women to understand.

In addition, CSO stakeholders highlighted the risk of location sharing in enabling other offline harms, particularly doxxing, by allowing perpetrators to find and track women and girls.[114] This feeds into the argument that online and offline VAWG exist in a continuum of violence against women and girls. Overall, CSO interviewees stressed the risks of sharing location and the need for enhanced user control and informed consent when sharing their location to ensure that users are kept safe online.

## Participate

### 5. Direct messaging

> ***"The biggest harm comes when it's two people talking in messages."***
> *Interviewed online dating platform*

Online VAWG that occurs through 'direct messaging' can be particularly challenging as it gives perpetrators direct, and sometimes unmoderated, access to abuse, bully and harass women and girls. In a study on the impact of online abuse on women's working lives, the most common type of harm experienced was unwanted private messages.[115] Another study showed that 30% of girls surveyed reported receiving unwanted sexual messages and videos through direct messaging.[116] An additional difficulty is if direct messages are disappearing, meaning that they disappear after a set amount of time. Victim-survivor support services noted that this can make it hard for the victim-survivors to evidence harassment and harder for platforms to moderate. Although, this is contextually dependent given that disappearing content, as outlined below, can be used as a form of additional security.

Overall, CSO stakeholders flagged that direct messaging is a particularly risky design feature, mainly because it enables direct lines of communication between perpetrators and women and girls. Multiple stakeholders, particularly with CSEA expertise and experience, flagged that children are particularly at risk of abuse through direct messages. The NSPCC found that when children are contacted online by someone they don't know, 74% of the time this contact initially takes place via direct message.[117] Refuge backed this up by sharing that disappearing direct messaging is a "really easy tool for groomers to use without leaving any traces". Some CSO stakeholders suggested limiting access between adults and children in direct messaging, in that adults who are not connected with the child should not be able to directly message them. While direct messaging is an important tool to enable private conversation, CSO stakeholders flagged that the lack of safeguarding, specifically around who is able to contact a user, is problematic and can increase the risk of online VAWG.

## 6. Screenshotting

'Screenshotting' is a design feature that has the potential to both mitigate and exacerbate the risk of online VAWG. On one hand, it can be heavily abused by perpetrators in sharing content or using it to abuse, bully, harass or threaten users. On the other hand, both interviewed groups shared that it is the primary way for victim-survivors to provide evidence of online VAWG. Across sectors stakeholders noted that challenges with screenshotting are relevant to all women and girls, and do not put specific communities at higher risk than others.

For perpetrators, screenshotting can be a way to inflict harm. Overall, it limits user control over removing content from online spaces as content can be screenshotted and reposted with ease. Glitch highlighted that this can be used to harass and target women and girls, for example by reposting content as a trigger for them to be abused or on the anniversary of an incident or abusive period. Screenshotting to inflict harm is also particularly relevant with regard to intimate image-based abuse, as perpetrators will screenshot intimate images without consent, often leading to further harm by sharing without consent. As a response, women and girls sometimes use platforms that specifically use disappearing content as it can give a sense of safety when sending intimate images.[118] Nonetheless, this type of content can also be screenshotted. One interviewee specialising in intimate image based abuse said that so-called "stolen snapchats" - that is intimate images originating on Snapchat - are commonly leaked on pornography sites. CSO stakeholders noted that anti-screenshot technology used by some platforms is ineffective.

With regards to evidence gathering, screenshotting can be helpful but also re-traumatising for the victim-survivor. Victim-survivor support services shared that this is particularly relevant if a victim-survivor has experienced multiple incidents of harm. In addition, CSO stakeholders indicated that screenshots are not always recognised by authorities as legitimate, as they come from users and can be edited or manipulated. CSO stakeholders warned that this experience can be very distressing to women and girls and may set a precedent for users to not report in the future or to remove themselves from online spaces completely.

## 7. Recommending content/users

> ***"It's about giving users more influence on what they're being recommended next time."*** *Stanford Internet Observatory*

'Recommending content' speaks to the algorithms that platforms build to recommend content or other users for a user to interact with. Algorithm-based recommendations often work to maximise engagement, which can include extreme content (e.g. hateful, racist and misogynistic content).[119] Across interviewed CSO's, stakeholders shared that this content can be detrimental to users in multiple ways, particularly in promoting online VAWG

(e.g. online threats, gendered hateful speech), showing users content they do not want or are not prepared to see, and enabling harm overall. Across both industry and CSO interviewees, stakeholders also emphasised that algorithm-based recommendations can radicalise men and boys into engaging in harmful behaviour toward women. These stakeholder interviews stressed the need for better safeguards on algorithms for all women and girls, as well as appropriate interventions to mitigate the radicalisation of users into the harmful treatment of women.[120]

One CSO stakeholder noted that users are getting better at understanding how these algorithms work and can sometimes work around them to actively fight against harm. For example, there is a case where a racist hashtag on a platform was spammed with non-harmful content by K-Pop fans looking to defuse online harm.[121] However, the interviewee also noted that if users are getting better at understanding algorithms in ways to prevent harm, perpetrators are also getting better at "game[ing] the system" and enabling harm. Overall, providing users with more choices on what they see on their feeds and timelines was a proposed solution. In addition, interviewed stakeholders across sectors recommended platforms clearly define what content is unacceptable via recommendations.[122] In the words of one academic expert interviewed: "it's about giving users more influence on what they're being recommended".

'Recommending users' was also pointed out as a risk - primarily where the perpetrator is known to the victim. Refuge highlighted that a perpetrator may reconnect with a victim-survivor because perpetrators are recommended the victim-survivor's profile, often because of common friends or followers. This can happen even if a victim-survivor changed their name or profile information. This puts the victim-survivor at further risk of online VAWG.

## 8. Blocking

> ***"People just block them, they don't think reporting will do anything."*** *Imkaan*

'Blocking' is one of the most common ways that users respond to online violence, as it rapidly limits access from the perpetrator. However, stakeholders, across sectors, explained that not all blocking features are effective. Often platform blocking features only allow you to block one user at a time, or do not automatically block any other accounts the perpetrator may use. Moreover, CSO interviewees shared that blocking is often misunderstood as a permanent mechanism to mitigate harm, whereas in reality blocking provides temporary or shallow mitigation efforts. One victim-survivor support service representative highlighted the risk of blocking for children in particular, saying that sometimes children do not understand that perpetrators can create new accounts and therefore blocking features give a false sense of security. In their words: "They don't know what to do… the child believes that because they've blocked the one account, the perpetrator will not try to connect with them". Another challenge CSO stakeholders highlighted was the lack of action after a perpetrator

has been blocked. Indeed, platforms often urge users to block perpetrators after they have reported, but do not always give further information on whether the perpetrator has been sanctioned or not, which may leave the victim-survivor uneasy. This is particularly problematic when women and girls experience cross-platform harassment. While stakeholders across sectors recognised blocking as a quick and easy-to-access safety feature, the limitations of blocking were acknowledged as a key risk to women and girls, specifically victim-survivors. One industry stakeholder noted that blocking has a limited impact in addressing the core source of the harm, and is used as a temporary solution. Lastly, one CSO stakeholder noted that platforms are not very transparent in how the blocking feature is being used. For example, how platforms are dealing with users who have been blocked by many different users. The interviewee suggested that if there were more transparency, it would be easier to identify what needs to change.

## 9. Reporting

> ***"It's like being burgled with no one coming to take the fingerprints."*** *Child Safety Expert*

'Reporting' is the action that most online platforms recommend to users when experiencing harm online. However, as highlighted by the stakeholder interviews across industry and CSO, it is one of the major design features that is ineffective in protecting women and girls online.

One of the key issues with reporting is a poor response by platforms. In one UK-based study of women who reported online abuse, 84% thought that their complaint had not been properly addressed and this increased to 94% for black and women from ethnic minority groups.[123] Refuge found that of those that reported online violence to platforms, 52% of women said the platform handled their report badly, which rose to 56% among women experiencing abuse from a partner or former partner.[124] When women do get a response, it is often delayed, automated and/or contains few details.[125] CSO stakeholders noted that a response that "this does not go against our community guidelines" without further explanation can be defeating for women and girls. To this Glitch specified that "this is often linked to platforms not having specific policies on online VAWG, meaning they are not well equipped to adequately respond to the gendered nature of reports made to them, and they do not always fit neatly into other policies which take a general (gender-neutral) approach to issues like harassment, abuse, disinformation or image-based abuse".

In turn, CSO stakeholders noted that this has led to women and girls being less willing and likely to report. A child safety expert argued that children "don't believe these services do anything to help them after they've made the report" and are less likely to report online harms. This was validated in research on online sexual harassment among children, which indicates that a major blocker in reporting was targets/victim-survivors thinking it would not make a difference.[126] Another interviewed child safety expert echoed

these sentiments, sharing that 60% of young girls experiencing harm will not use the reporting function on platforms, as they do not believe any action will happen on the platform level and express they do not know who to trust, therefore favouring the blocking action to quickly dissipate the risk.

This applies to women as well, as evidenced by a report by UN Women, which stated that the most common reason that women did not report online violence was that they "didn't think it would make a difference" followed by those who "didn't know who to report it to".[127] Ultimately, this may prevent women from reporting, or engaging on platforms, altogether. In the words of Glitch: "if we don't see significant improvements to reporting practices and transparency in relation to online violence against women and girls, we will continue to see women reducing or stopping reporting".

There are specific aspects of reporting that have been pointed out as particularly unhelpful by CSO stakeholders, including not being able to report multiple pieces of content at once, restrictive reporting categories, and a lack of explanation of what those categories mean. Not being able to report multiple pieces of content or accounts at once makes it particularly hard for women and girls who experience high volumes of abuse, which is common for women and girls with public profiles. The rigidness of reporting features, often expressed as a finite list of predefined categories, means that they don't account for the complexity and intersectionality of many online VAWG cases.[128] CSO stakeholders shared that the lack of explanation around reporting categories in practice makes it unclear to the victim-survivor what to choose to 'report' on (e.g. hate speech), especially for those who do not speak the primary language of the platform. They flagged it may also lead to inaction from the platform if a user chooses the wrong category when reporting. The World Wide Web Foundation captured this challenge well, stating "unless you have good feedback loops that are specific to those that are vulnerable to harm you will continue to cause harm". Although reporting is widely used and supported by platforms, the current approach has been extensively flagged as ineffective for women and girls.

## Respond

### 10. Removing content

> ***"Don't prevent digital access, make it safe."*** *European Women's Lobby*

Similar to the point above, CSO stakeholders argued that platforms are ineffective in removing abusive content. This is particularly true for some types of harm. CSO stakeholders shared that the removal of intimate image-based abuse is particularly challenging, which has also been reflected in reports on taking down intimate images.[129] Specifically, victim-survivors may need to find and report all individual images or videos for them to be taken down completely. There are organisations set up to help with the removal of intimate images, but due to limited resources and

geographical reach they cannot meet the needs of all women and girls.[130] This has led some victim-survivors across the world to take matters into their own hands. For example, some victim-survivors in South Korea have turned to private companies and paid them to remove intimate images.[131]

Moreover, CSO stakeholders noted that platforms are specifically ineffective in removing what they referred to as 'legal but harmful' content, of which many forms of online VAWG are included. One interviewee argued that the ineffective removal of harmful content normalises and legitimises online VAWG. Some stakeholders across sectors also identified the harmful cumulative effect of when women witness abuse against other women. As explained in multiple studies, this can have a silencing effect and prevent women and girls from pursuing certain professions like politicians or expressing themselves freely online.[132] One CSO stakeholder also theorised that it is "hard to quantify the loss" of women and girls who leave online platforms because they witness the violence that other women and girls experience. This highlights two points. Firstly, it shows the importance of the effective removal of content as a mechanism to support women and girls and ensure that their online spaces are safe for them. Secondly, it emphasises the importance of preventing online VAWG to reduce the burden on communities to remove the content in an efficient manner - something we explore further in Objective 2.

## 11. Terminating and suspending user accounts

Across industry and CSO stakeholder interviews, we heard that platforms were generally ineffective in sanctioning perpetrators, including terminating and suspending perpetrators' accounts. CSO stakeholder interviews attest to how perpetrators' accounts remain active after having been reported for online VAWG and sometimes resurface using the same handle or profile pictures. Furthermore, it is commonly understood that perpetrators create new accounts to inflict harm. Speaking to this, a report by the eSafety Commissioner Australia on how platforms tackle child sexual abuse highlights discrepancies in how platforms deal with this recidivism - when a perpetrator reoffends after having been suspended or banned from the service.[133] The lack of cross-platform cooperation was shown to be a major blocker to the effective sanctioning of users who commit harm on several platforms, as well as understanding the breadth and depth of the issue.

Some stakeholders across industry pointed out the misalignment in enforcing community guidelines in which perpetrators' accounts are not removed, but at the same time, some users can have their accounts terminated for using keywords that may get flagged by automated content moderation. Overall, poor enforcement of community guidelines and the lack of sanctioning of bad actors contribute to the lack of trust women and girls feel in reporting online VAWG.

## Analyse & Comply

### 12. Human and automated moderating

> ***"The idea that we can solve human problems online without the interventions of real humans is unrealistic."*** - *Unconform*

Several issues around human and automated content moderation were flagged in interviews. Firstly, content moderation for text is largely easier than images, which puts victim-survivors of image-based abuse at a disadvantage as it reduces the likelihood of action being taken. Secondly, online VAWG that uses coded language or that has cultural nuances create challenges for automated moderation. Thirdly, victim-survivors rarely receive any communication about why a certain moderation decision was taken, such as a decision not to remove a post, contributing to confusion and lack of trust in platform moderation.

One academic expert from the Global Network on Extremism and Technology underlined challenges that platforms face in automated moderation. Namely, building content classifiers is difficult because there's no shared understanding of online VAWG or even a shared definition. This contributes to a misalignment between what users want content moderation to do and what platforms can feasibly do. In their words: "what the general population's expectations of what platforms can do to mitigate harm and what a platform can actually do is completely misaligned". Glitch noted that this may be a consequence of features that are implemented without safety by design approaches. An example would be implementing design features where content moderation is harder or impossible so platforms cannot mitigate harm. There is the additional challenge of online VAWG that takes place in end-to-end-encrypted environments, where it is an unrealistic expectation that moderation will be able to take place.

Some platforms have stressed the importance of human moderation. This is especially true when there are additional cultural considerations or nuances which automated moderation cannot pick up. One industry stakeholder made a distinction between bad content and bad behaviour, where bad content is relatively easy to moderate but bad behaviour, which is dynamic and more context-dependent, is much harder. Therefore, platforms have to lean on human moderation to a certain extent to get insight into the context of harm. Interviewees also stressed the need for victim-survivors to have better and more direct communication channels with human moderators to give context and receive support.

A small platform highlighted that human moderation is preferable in tackling online abuse but also recognised it does not work at scale. Another platform argued that having an all-female moderation team allows for better emphasis on and sensitivity to online VAWG - but this is rarely the case on other platforms, nor is it feasible. A challenge to this,

one industry stakeholder pointed out, is that sometimes moderators are outsourced and hired in politically conservative countries which may negatively impact certain women and girls such as those who are part of the LGBTQ+ community. Another academic expert interviewee stressed that outsourced moderation teams are often underpaid and not provided with adequate training and support to make decisions around this content.

Overall, issues around moderation stress the need for more investment and education in automated and human moderation to better respond to the needs of women and girls

.

## Summary: Objective 1

There are a number of overarching challenges that platforms will need to address to mitigate the risk of online VAWG.

This includes a lack of platform policies explicitly targeted at online VAWG, design approaches that put too much responsibility on users to ensure their own safety, little friction enabling the reach and scale of online VAWG, perpetrators bypassing safety measures and poor data collection on women and girls' experiences on platforms. Although any design feature can be misused to cause harm, there are some that present a particular risk to women and girls. We have presented 12 design features prevalent across the eight different platforms, highlighted as posing a risk to users through our stakeholder interviews. Importantly, while we have presented a number of challenges, we remain positive about the potential for safety by design to protect women and girls online, which will be explored in the following section.

# **Objective 2:** Designing to protect against online VAWG

This objective aims to understand current best practices in building safer online spaces, including the use of safety by design. We hope that the findings in this section will support actors across the online VAWG ecosystem to shape their safety practices and encourage the adoption of safety by design. This section will outline current successes in protecting against the risk of online VAWG through design from an overarching perspective, before diving into safety by design.

## Overarching trends in designing to protect against online VAWG

### 1. Increasing cross-sectoral partnerships

There has been an increase in cross-sectoral partnerships to support online VAWG safety initiatives. In 2022, Google's Jigsaw partnered with civil society organisations (CSO) such as Glitch and the European Women's Lobby to build a 'Harassment Manager' which allows users on X, then known as Twitter, to review, sort, and export harmful comments, and mute or block perpetrators.[134] This technology was later outsourced to the Thomas Reuters Foundation to create their 'TRFilter' which additionally helps users create reports to store or share with third parties. [135] The Online Dating Association facilitates networking with other providers in their sector; one of our interviewed organisations reflected that, as a smaller platform, this networking is hugely beneficial, enabling discussion of these issues with larger platforms with sectoral expertise.[136] Increased collaboration has the potential to further refine safety measures and encourage actors across the online VAWG ecosystem to share learnings and prioritise safety.

### 2. Building online VAWG specific resources and tools

Though community guidelines and policies do not yet specifically address online VAWG, platforms are increasingly recognising women and girls are at higher risk of harm and are developing targeted resources. In 2021, Meta launched their *Women's Safety Hub* featuring specific policies, resources, and safety tools.[137] LinkedIn's recent campaign, "Keeping LinkedIn a Safe, Professional Community Where Everyone Can Thrive", updated community guidelines and automation tools to prevent harassment.[138] In addition, civil society organisations are continually developing new resources, including Refuge's 'Digital Breakup' education tool which guides users through

securing their online accounts across multiple platforms.[139] As the sector develops resources that address online VAWG, proper engagement with users is needed to truly create an impact in preventing online VAWG.

### 3. Increased development of design features for online safety

Platforms are developing and rolling out more features focusing on user safety. Over January and February of 2023, we identified over 12 new design initiatives focusing on user safety, privacy, and empowerment. This includes giving users more control over content curation, such as TikTok's "Refresh" feature to reset the video recommendations on a user's feed.[140] Online dating platforms are providing users with greater personalisation for searching, filtering, and connecting with other users. For example, in February 2023 the French online dating platform Happn announced their 'hub' feature, which provides users with greater control over their matches, offering four different mechanisms.[141] Reporting features are also being improved, for example, Tinder announced a new 'Long Press' feature, which allows users in chats to automatically initiate a tailored reporting flow.[142] Across the Match Group, there are now notifications and warning messages focused on specific harms, such as impersonations and scams.[143] Platforms are actioning and prioritising online safety by incorporating more and more tools to build safer online spaces.

## Safety by design practices in the online VAWG Ecosystem

The UK Government has defined safety by design as:

> ***"[...] the process of designing an online platform to reduce the risk of harm to those who use it. Safety by design is preventative. It considers user safety throughout the development of a service, rather than in response to harms that have occurred."*** [144]

Similarly, Australia's eSafety Commissioner's guidance on safety by design also emphasises proactiveness and the need to be user-centric.[145] The *VAWG Code of Practice*, developed by a coalition of civil society organisations, takes safety by design one step further by embedding a focus on online VAWG. It encourages platforms' use of safety by design to "(a) minimise the risk of those harms arising from VAWG content and practices, (b) mitigate the impact of those that have arisen [and] (c) enhance women and girls' freedom online". [146] In this section, we will explore what is working well and not working well with safety by design, based on feedback provided through the literature review and stakeholder interviews.

## What is working well with safety by design?

From our stakeholder engagements, we identified two key trends:

### 1. Safety by design is gaining traction across the online VAWG ecosystem

Interviewed civil society organisations are using their understanding of victim-survivors to develop safety by design principles, and are sharing their work with platform partners. The focus on safety by design has also appeared at international conferences, such as the Global Dating Insight conference which featured Garbo, an online safety technology provider, to discuss best and worst practices of safety by design.

Platforms are also considering safety by design principles when planning product strategy. One interviewed platform shared that after their team reviewed the safety by design framework, they look to include Trust & Safety team members in every product strategy meeting. Another interviewed platform shared that Trust & Safety representatives create weekly newsletters for the leadership board, collating internal and external safety by design news. Companies are also increasingly appointing leaders to develop safety by design into their work, such as the Match Group's recruitment of a safety by design Director.

### 2. Platforms are integrating online safety tech solutions

Online safety tech solutions are defined as: "Safety tech providers develop technologies or solutions to facilitate safer online experiences and protect users from harmful content, contact, or conduct".[147] We are seeing an increase in interest and integration, even acquisition, of online safety tech solutions, from Muzz partnering with Yoti for identity verification to Reddit acquiring Oterlu for content moderation.[148] The sole adoption or integration of online safety tech solutions by platforms is not a safety by design practice unless both parties exemplify the principles; online safety tech companies themselves need to embed and practice safety by design principles. Both buyers and providers of online safety tech solutions need to review and embed safety by design principles separately to ensure effectiveness for users. Platforms recognising this difference, and intentionally acting on both, are set to be more effective in addressing online VAWG.

## What can be improved in understanding and incorporating safety by design?

Although the online VAWG ecosystem is making strides with safety by design, stakeholders highlighted a few challenges which have been outlined below. These are important to bear in mind as governments, organisations, and platforms look to collaborate and agree on what it means to practise safety by design.

From our stakeholder engagements, we identified three key challenge areas:

## 1. Inconsistencies in the application of safety by design principles

Inconsistencies in safety by design principles can lead to discrepancies in their implementation. The UK Government's and Australia's eSafety Commissioner's safety by design principles have aligned definitions. However, the principles themselves vary.[149] In addition, Civil Society and academia in the UK have collaborated to build on the UK's safety by design principles.[150]

Platforms also have their own perspectives and objectives on how their products are best designed and might refer to their own community guidelines rather than safety by design principles. Platform interviewees acknowledged the safety by design principles defined by government and CSOs, but also raised the need to be specific to their own product functionality and user groups. One platform shared that they find some principles more relevant to certain product features, and so principles are applied to differing extents.

Interviewed stakeholders proposed a range of solutions to this challenge. Many interviewees called for an agreed-upon set of principles for the entire online VAWG ecosystem. However, others highlighted that an aligned set of principles is not the answer, as platforms and civil society organisations have different needs and priorities. One civil society organisation encouraged platforms to partner with external safety by design experts to better adapt principles to individual platform needs. Another recommended having dynamic principles that are routinely reviewed, discussed, tested, and adapted, similar to platform policies and guidelines.

## 2. Lack of targeted safety by design guidance to counter online VAWG

There is a lack of safety by design principles for specific at-risk communities, including women and girls, and even more so for those with layered protected characteristics such as black and minority ethnic groups. A women-centric design expert shared, "women are being failed in digital products in the same way they have been failed in offline products". Platform design teams need to recognise that women and girls experience spaces, products, and services differently – and therefore have pain points and needs that must be specifically addressed. A safety by design expert also stated, "Safety by design addresses human interaction", reiterating the point that understanding the differences in user behaviour and what enables interactions between users is crucial.

It is not practical to develop individualised safety by design principles for every socio-demographic group. However, existing guidance could be improved by embedding contextual empathy and research into its approach. For example, it could include case studies of how design teams can include women and girls throughout the entirety of product development, from design ideation to monitoring to the termination. This builds on the

UK Government's second principle: *Platforms should consider all types of users.* Two examples of victim-survivor-informed research include the *VAWG Code of Practice*, and Chayn and End Cyber Abuse's *Orbits* report.[151]

## 3. Smaller platforms face challenges in embedding safety by design

Smaller platforms face additional challenges in implementing safety by design, such as the upfront cost of building new features or integrating third-party solutions. In addition, advanced online safety tech tools can be expensive to adopt, especially during early implementation. This is significant as more advanced online safety tech tools have a greater ability to adapt to platform-specific requirements, enhancing the capability of the platform to respond to harm appropriately. One platform shared that since they have grown, they were able to acquire an online safety tech solution rather than simply integrate it, which then enabled them to train and improve the technology with platform-specific data.

## Summary: Objective 2

> ***"Safety by design approaches have proven effective in preventing and reducing online violence against women and girls, particularly when different stakeholders work together. We can no longer ignore the impact of platform design on online VAWG, so let's work together to tackle it."*** *Glitch*

Safety across the sector is picking up momentum through cross-sectoral partnerships, internal management changes, policy developments, and safety design feature roll-outs.

Efforts to address online VAWG specifically continue to emerge, but platforms are beginning to share resources and tools to best support women and girls' safety. Safety by design is gaining attention and efforts are being made across the online VAWG ecosystem, but shortfalls in clarity of the frameworks and integrating the principles into organisational decisions keep the potential for impact stagnant. As the sector continues to shape these frameworks and apply them specifically to address online VAWG, we can expect a new wave of design features that truly target online VAWG effectively, as they would not only be specific to the needs and pain points of victim-survivors and women and girls but be operating in a system completely embodying and acting on the safety by design approach.

# Objective 3: Potential new design approaches to protect women and girls

The third objective takes account of the findings from Objectives 1 and 2 to identify opportunity areas and recommendations for design to help protect users from online VAWG. This is a critical exercise in ensuring the research does not just evidence the issue area, but focuses on actionable, future-focused solutions for platforms to potentially integrate to better protect women and girls. This section highlights three overarching recommendations to support design, identified through the evidence review and stakeholder interviews, before outlining specific opportunities for better reporting, education, warning and taking action at scale (bulk actions). The overarching recommendations are not new, in fact, multiple stakeholders across the online VAWG ecosystem have been advocating for many of them. The convergence around these recommendations reinforces their importance and shows that there is a pressing need to implement them.

## Recommendations for future design approaches

### 1. Adopting a user-centric, trauma-informed approach is the first step to mitigating the risk of online VAWG through design

Intentionally creating a platform design around the needs of women and girls and victim-survivors is critical to mitigating the risk of online VAWG and also supporting victim-survivors online. Keeping women and girls at the centre of platform design would, for example, mean greater choice in what women and girls can see on platforms, as well as what content is allowed and promoted in relation to online VAWG policies. Glitch stressed that design features that can be effective in mitigating online VAWG often entail a "'move towards more preventative safety measures alongside more transparency and accountability for curation on timelines". Taking a user-centric approach also means recognising that women and girls across the world have different experiences and needs. It is also about recognising that those experiences intersect with other identities. When discussing online VAWG, Glitch described it as: "'it's very much gender and, not just gender. For example, highlighting the experiences of 'misogynoir' directed at black women online, as a mix of misogynist and racist abuse".

Another way to practically centre women and girls in platform design proactively is to monitor risk to women and girls, before any harm has happened through a risk assessment.[152] In the words of one interviewee: "the earlier back you can go, the better". These risk assessments would also pay particular attention to intersecting, protected characteristics. Other stakeholders have pointed out the importance of an industry code of practice on online VAWG to standardise preventative measures for online VAWG.

A trauma-informed approach would ensure that the impacts of online VAWG are acknowledged and considered throughout the platform. Building on Orbits, trauma-informed design means that it *"understands and acknowledges the nature and impact of trauma"*.[153] Trauma looks different to each victim-survivor but centering the role of trauma in design is an important step forward, which was reinforced by our stakeholders. One interviewee stressed that "preventative measures have to be trauma-informed". This could involve working with civil society organisations that represent victim-survivors or working with victim-survivors themselves to design new features.

A women-centric design expert argued that, currently, *"we are still designing for safety reactively rather than proactively"*. Adopting a user-centric, trauma-informed design approach would ensure that platforms do.

## 2. Increasing knowledge sharing across the online VAWG ecosystem to enable collaborative approaches to protect women and girls through design

The need for greater sharing in trust and safety was emphasised by stakeholders across the online VAWG ecosystem. As noted in Objective 2, there are increasing collaborative efforts in the online VAWG space, such as the Stop Non-Consensual Intimate Image Abuse Initiative (STOPNCII.org). However, stakeholders stressed that there should be more knowledge sharing between platforms and also between platforms and civil society organisations who work with online VAWG. One dating platform put it as there should be "collaborative knowledge sharing rather than knowledge siloing". Stakeholders flagged common definitions, collaborative development of safety solutions, and shared lessons learned in design developments as key first steps to enable actionable change.

## 3. Raising awareness and support for victim-survivors through educational resources around online VAWG throughout the user journey

Educating and supporting users before and after harm has occurred is crucial, especially as many women and girls are not aware of available resources to them. While many platforms list educational resources and civil society organisations in their community guidelines, nudges or tip tools can be used to provide users with more targeted resources at the right time in the user journey. One dating platform described this as "how can you give users the information they need at the point they need it?" For example, if a user receives a harmful message the platform could prompt them to report it. Tinder does this by sending users a message "Does this bother you?" when it detects harmful language in a message.[154]

This can also be extended to users who are about to inflict harm, such as sending nudges flagging that the content they are about to send or post is harmful and asking them to reconsider. A women-centric design expert put this as "How do we carve out a role for men and engage men in the safety conversation? How do we engage those who perpetuate harm alongside those who experience it?" Platforms like Tinder are already using these types of behavioural nudges directed at bad actors.[155] Using informative nudges can both support the victim-survivor in taking appropriate action and also work to deter perpetrators from inflicting harm.

Once harm has occurred, victim-survivors should be directed toward appropriate support services. For example, if a user reports cyberstalking they would get a nudge directing them toward support resources and services specifically targeted at victim-survivors of cyberstalking. As shown in our Appendix: 'Safety Features Case Studies" we see how this works in practice with the partnership between Bumble and Chayn where users who report sexual assault or harassment are given access to Bloom - a remote trauma support service for victim-survivors.[156] Stakeholder interviews highlighted that it is critical to make these resources culture- and location-specific, so that women and girls all over the world can benefit from this support. Our research and stakeholder interviews have stressed the incorporation of education throughout the user journey as an essential step in making the overall user experience safer as well as in supporting victim-survivors.

## Summary: Objective 3

This section has suggested actionable, future-focused design solutions that would better protect women and girls on platforms.

By highlighting three overarching recommendations to support design we stress the need for user-centric, trauma-informed design approaches, an increased knowledge sharing in safety by design and better support for women and girls throughout the user journey. Together, we hope that this inspires platforms to adopt potential new design approaches that would safeguard women and girls online.

# 6. Conclusion

***"We must not forget women and girls have every right to take up space online - to meet people, make friends, fall in love, fall out of love, express themselves, build a business, start movements and browse and produce content for entertainment. The reason we need the internet to be safe is so that they can enjoy these liberties."*** *Chayn*

Online VAWG is a global, pervasive issue. While governments, international organisations, and platforms are devoting more attention to this critical topic, there has been less emphasis on the role of platform design and its impact on online VAWG. Our hope is that the findings from this report shows the importance of platform design and that it encourages rapid action to protect and safeguard women and girls online.

Our research presented trends in the online VAWG landscape around prevalence, impact and technology. Paying particular attention to the impact of online VAWG across different protected characteristics, we emphasised the intersectional nature of online VAWG and how this creates different lived experiences for each user. Our interviews with platforms and civil society organisations shed light on the strengths and challenges with current design features tackling online VAWG. This illustrated how, even with good intentions in feature development, design features can be misused and abused. Interviews with Civil Society Organisations helped to voice the lived experiences of women and girls who have experienced online VAWG, emphasising specific pain points of women and girls across platforms. Platform interviews highlighted best practices in building trust and safety into their platforms and how they embed safety by design principles throughout their organisation, as well as some of the challenges in doing so. While our research showed there is a substantial effort across the online VAWG ecosystem to tackle the harm, there is still work to do.

Recognising that online VAWG requires a systems-based approach, **three key recommendations** were developed to enable ecosystem-wide efforts to mitigate online VAWG through design:

## 1. Adopting a user-centric, trauma-informed approach is the first step to mitigating the risk of online VAWG through design

Keeping women, girls and victim-survivors at the heart of platform design would ensure that safety for women and girls is embedded from the start, and not as an afterthought. This also entails consideration of the intersecting needs of women and girls with different protected characteristics. Platforms can engage expert advisory boards and victim-survivor helpline/support services and review recent reports during design ideation sessions to understand victim-survivors' and at-risk populations' experiences and needs, specific to the platform.

## 2. Increasing knowledge sharing across the online VAWG ecosystem to enable collaborative approaches to protect women and girls through design

Interviews with both platforms and civil society organisations highlighted the need for greater knowledge sharing around trust and safety, potentially through a platform coalition around online VAWG. This would avoid platforms having to reinvent the wheel and help smaller platforms with fewer resources to effectively implement online safety tech solutions. Trust and safety should be not seen as a commodity, but rather a necessity to keep women and girls safe.

## 3. Raising awareness and support for victim-survivors through educational resources around online VAWG throughout the user journey

Our research showed that women and girls are generally under-supported by law enforcement, platforms and support services. To educate users about online VAWG and to support victim-survivors, educational and support resources should be embedded across the user journey, not just reserved to the community guidelines. These resources do not have to be developed by platforms independently and should use readily available resources from civil society organisations, government agencies and other expert bodies.

While this report adds to the evidence base of the role of platform design in mitigating and exacerbating online violence against women and girls, there is more work to be done. Nonetheless, we hope our findings and design recommendations can help guide platforms to build safer spaces for women and girls and allow them to engage online without the fear of harm, exclusion or silencing.

# 7. Appendix

## Methodology

The report was broken down into three workstreams, including landscape mapping, stakeholder engagement, and design iteration. The tasks undertaken in these workstreams have been outlined in detail below.

### Landscape Mapping

The purpose of the landscape mapping was to understand key impacts on the current online VAWG ecosystem. It aimed to address research questions

1. What design features do platforms currently use?
2. How do these design features impact the risk of online VAWG for users of the platform/service? How does the impact differ according to different protected characteristics of the women being targeted?
3. What 'safety by design' features do tech companies currently use to protect against online VAWG?

The landscape mapping consisted of a literature review and an evidence review.

The **literature review** aimed to present key trends in the prevalence and impact of online VAWG as well as highlight the experiences of victim-survivors by addressing research question 2.

The literature review included an analysis of over 100 academic, civil society, government and platform sources to identify trends impacting the online VAWG ecosystem at large. Each source was given a RAG rating (red-amber-green) according to methodology, quality of analysis and source to ensure consistent quality.

Multiple literature searches were conducted in Google and Google Scholar using a series of keywords over a time period of two months. Acknowledging that "Online Violence Against Women and Girls" is just one of many terms used in the online VAWG ecosystem, the keyword searches consisted of iterative keyword filtering.

The following list contains the keywords used for the general search:

- "Online violence against women and girls"
- "Online violence against women"
- "Online violence against girls"
- "Online gender-based violence"
- "Online child abuse + girls"
- "Technology facilitated violence against women and girls"
- "Technology facilitated gender-based violence"
- "Online violence against women and girls + [protected characteristics]"
- "Safety by design + women"
- "Safety by design + girls"
- "Online safety + women"
- "Online safety + girls"

For specific harm types we used some of the terms in the online VAWG taxonomy, including "intimate image-based abuse", "deep fakes" and "doxxing".

For platform policies keywords included:

- "[Platform name] + women's safety"
- "[Platform name] + online violence against women and girls"
- "[Platform name] + online abuse women"
- "[Platform name] + online abuse girls"
- [Platform name] + technology facilitated gender-based violence"

The platforms used were primarily eight large-scale social media and online dating platforms: Facebook, Instagram, Snapchat, TikTok, X, YouTube, Bumble and Tinder.

We focused on sources from the last two years (2021, 2022) to capture the current nature and trends of online VAWG, as well as capture the increase in reports of online VAWG during the Covid-19 pandemic. In terms of location, the literature review took an international lens but given the existing networks of the researchers, a majority of them focused on the Global North. To ensure a comprehensive overview of the impact of platform design on online VAWG the sources covered a broad range of online harms as specified in our taxonomy. However, we did not cover sources that focused explicitly on 'technology-facilitated domestic abuse' as it was out of scope of this project.

All sources were included in the literature review, even if their RAG rating was low, as they were often platform announcements or policies that are relevant to the report but score lower on methodology and quality of analysis (e.g. a blog post by Australia's eSafety Commissioner). That said, not all sources assessed in the literature review were used in the report as they may not have been relevant for a specific section.

An **evidence review** was then conducted to investigate platform design approaches and respective policies by addressing the questions of:

1. What design features do platforms currently use?
2. What 'safety by design' features do tech companies currently use to protect against online VAWG?

When researching platform design, we focused on eight large-scale social media and online dating platforms: Facebook, Instagram, Snapchat, TikTok, X, YouTube, Bumble and Tinder. The information was found by looking at the respective platforms' community guidelines, terms and conditions, safety centres or safety guidelines as well as blog articles published by the platforms. All research was conducted and collated on a Miro board to allow for agile working.

From the evidence review, we compiled a long list of design features used across the eight platforms. It is important to note that there is no universally agreed upon definition of 'design feature' but to set the scope for the longlist we formulated the definition: *'Touchpoint in the user journey which enables the user or platform to take action which affects the user experience'.* This captures both user and platform actions.

We compiled a long list of 57 design features by going through the platforms themselves and documenting different features as aligned with our definition of 'design feature'. We complemented this by reviewing websites that detail user interfaces and user flows on major platforms such as mobbin.com, pageflows.com and nicelydone.club. As the longlist generation was ongoing during stakeholder interviews, we also ensured to cross-check the longlist with interview notes and ensure no outlier features were missing.

However, as platforms release and retract features frequently, there are undoubtedly more design features that were not featured in the long list. The long list was reviewed internally by PUBLIC's design team for quality assurance.

A risk assessment comprising two parts was used as the selection criteria for the design feature shortlist of 5-10 features. The two criteria were:

1. **Representativeness**: If the design feature is leveraged by multiple platforms currently.
2. **Risk to users:** If the design feature poses a risk to victim-survivors, based on civil society organisations' (CSO) interviews.

Each feature in the longlist was then given a rating

from 1-3 for the two risk assessment criteria.

- For 'prevalence', a 3 denoted that the feature exists on 6-8 platforms, a 2 if the feature exists on 3-5 platforms, and a 1 if the feature exists on 1-2 platforms.
- For 'risk to users', qualitative feedback was gathered from stakeholders and the design features were ranked according to the frequency with which it was discussed and the level of risk conveyed by the stakeholder.

It is important to note that this risk-based approach is not exhaustive of all the risks that women and girls may face on platforms, but rather provide a snapshot. The short list was tested and reviewed by the internal design team and the Project Advisory Board.

It is worth noting that narrowing down on a series of design features for this project that may contribute to the perpetration of online VAWG is not an exact science. There may be more design features that may exacerbate the risk of online VAWG that are not highlighted in this report. For example, we did not count algorithms as a design feature. Algorithms have been shown to sometimes promote hateful content, but the diversity of use cases and implementation make them difficult to study as a discrete category. In addition, due to a lack of transparency over how algorithms operate, we have had difficulties in evidencing the specific role of algorithms in mitigating or exacerbating the risk of online VAWG. However, as stakeholders noted risks in how users are recommended content by the platforms we included 'recommending content / user', which covers aspects of algorithms but is more user-facing.

## Stakeholder Engagement

The purpose of the stakeholder engagement was to validate our findings, generate insights and test findings. This phase included stakeholder mapping and stakeholder interviews.

**Stakeholder mapping** was conducted to identify a longlist of significant and active online VAWG ecosystem players across both the support sector and online platforms. The support sector organisations were categorised into

a) Academia
b) Government bodies
c) Civil society organisations
d) Industry bodies
e) Online safety tech suppliers
f) Intergovernmental organisations
g) Support services

The online platform ecosystem organisations were categorised into

a) Online dating
b) Social media
c) Adult websites
d) Gaming platforms

Stakeholders were identified through PUBLIC's current networks, online search using the same key terms as used in the literature review and listed as key actors through campaigns, literature, research, support service, technology development, etc. This led to the development of a long list of over 200 stakeholders active in the online VAWG ecosystem.

To target outreach for interviews, the long list was shortlisted using an **influence x interest** scoring matrix and assessed in relation to their appropriate groupings. The **influence x interest scoring matrix** was based on the roles of stakeholders specifically:

| Score | | Influence<br>How has the organisation engaged with the online VAWG ecosystem and impacted/informed key interventions, decisions, policies, developments, etc.? | Interest<br>At what level of priority does the organisation work to address online VAWG through their work? |
|---|---|---|---|
| 3 | **Description** | **High - Top-tier organisation with influence in preventing online VAWG. This is evaluated based on:** | **High - Preventing online VAWG is central to the organisation's mandate and/or activities. This is evaluated based on:** |
| | Support Sector | Partnership and involvement with key stakeholders on a global or regional scale (e.g. government partnership, coalitions etc.). | The key mission is specific to address online VAWG, or the overarching mission is to address VAWG, with a specific strategy/ pillar of work of online VAWG. |
| | Tech Platforms | High (leading) number of international users (platforms): Over 500 million active users. For dating platforms: over 80 million. | Explicitly call out VAWG as an area of concern within policy and/or design. |
| | Academia | Seen as a leading academic contributor to online VAWG based on engagement in the ecosystem through partnerships, events, committee involvement, joint-research pieces etc. | Distinguished research papers in online VAWG; seen as leading academic contributors to this space. |
| | Government | Has enacted policies, programmes, research, etc. related to preventing online VAWG and generally VAWG. | Dedicated policy group or branch of civil service committed to VAWG. |
| 2 | **Description** | **Medium - Second-tier organisation with influence in preventing online VAWG. This is evaluated based on:** | **Medium - Preventing online VAWG is a secondary issue to the organisation but evidence of recent activity in online VAWG. This is evaluated based on:** |
| | Support Sector | Working in VAWG but with limited partnership and involvement. | Hold a pillar of work addressing VAWG, with at least one project or paper in online VAWG. |
| | Tech Platforms | Mid number of users, including at regional scale: Between 100 - 500 million active users. For dating platforms: over 80 - 40 million. | Limited call out VAWG as an area of concern within policy and/or design OR focus on other areas within VAWG space (e.g. image-based abuse) specifically. |
| | Academia | Seen as an academic contributor to VAWG based on engagement etc. | Conducts research focused on VAWG, less specific to online VAWG. |
| | Government | Yet to enact policy, programmes, research, etc. but has in place a clear strategy for online VAWG/VAWG; or works on related areas of women's equality and women's rights. | Interest but no dedicated policy group or branch of civil service focused on online VAWG/VAWG space |
| 1 | **Description** | **Low - No meaningful statutory powers or system influence in preventing online VAWG. This is evaluated based on:** | **Low - Limited evidence of active interest in online VAWG. This is evaluated based on:** |
| | Support Sector | Little to no partnerships or involvement in the VAWG space. | Work relates to gender equality or addressing VAWG but not specifically online VAWG. |
| | Tech Platforms | Low number of users, a niche community or local level of users: Under 100 million active users. For dating: under 40 million. | No mention of VAWG as an area of concern within policy and/or design. |
| | Academia | Limited engagement in the VAWG ecosystem. | Limited focus on VAWG. |
| | Government | Executes little work around VAWG or gender equality and women's rights in place. | Few civil servants are interested, but no clear group working on policy in this area. |

To appropriately conduct prioritisation evidence was pulled from the literature review and across organisations' websites. For the platform prioritisation, the team additionally assessed the online VAWG risk for each platform based on the evidence collected during the literature review to select key outliers of the interest x influence score matrix. Assessing the risk of online VAWG mitigated the risk of biassed results that only platforms which expressed interest in online VAWG were considered. This was due to address the research question to assess increased risk by design features, and not only those that decreased the risk by design features.

The final output was a short list of 165 stakeholders. Due to scoping, the team focused outreach for online platforms to be within the social media and online dating sectors. The team reached out to 130 stakeholders:

- 73 support sector actors
- 57 online platforms

Outreach was facilitated through direct communication channels and all engaged stakeholders were provided with a Participant Information Sheet which reviewed the project aims, outputs, interview logistics, GDPR compliance and more. Of the invited stakeholders to interview, 65 replied to the inquiry, demonstrating the collaborative nature of the online VAWG ecosystem. The team provided additional information and answered questions to facilitate interview scheduling as requested.

In total 32 stakeholders were **interviewed** over multiple weeks. These stakeholders were based out of Belgium, Denmark, India, Poland, Australia, the UK, and the US:

- 10 platforms
- 16 civil society organisations
- 6 subject matter experts across academia, online safety tech companies, and think tanks were interviewed.

Interviews were 45 minutes and held over Google Meet or Zoom depending on the interviewee's technological restrictions. Questions directly related to the project research questions and were tailored to either the support sector, platforms, or subject matter expert's area of expertise. Questions were grouped under the following categories:

1. Online VAWG landscape
2. Company Trust & Safety Initiatives [Platforms only]
3. Victim-survivors lived experiences [CSOs only]
4. Safety by design
5. Challenges and Successes Design Features
6. New Design Features and Strategies

The interview team took notes, but did not record the session, and followed all GDPR compliance for data collection and handling. All evidence was anonymised for GDPR reporting purposes. Follow-up with Interviewees was conducted to provide requested access to data/notes and answer any outstanding questions. Additionally, the final report's relevant sections where case studies, quotes, or other interviewee evidence was used, were provided to interviewed stakeholders for review.

## Design Analysis

The purpose of the design review was to consolidate the findings from the landscape mapping and stakeholder interviews and address research question 5) What potential design features would be useful to consider implementing to further protect against the risk of online VAWG? The design review consisted of creating four proto-personas, visualising user journeys. To understand the design analysis, it is important to highlight some key concepts:

- A proto-persona is a fictional representation of a typical user within a segment of the overall user group.
- Feature set is a group of features that serve a common aim.
- User journey is a path a user may take to reach their goal when using an online platform.

We created a total of four proto-personas with the aim of demonstrating women and girls' different experiences with online VAWG depending on harm types and protected characteristics. Their stories were based on challenges women and girls face online as identified in the literature review and informed by interviews of civil society organisations working with violence against women and girls. After the interviews, we aggregated findings based on feedback on different protected characteristics and then broke it down by pain points, reoccurring problems, needs and additional considerations. Through this, we aimed to capture some of the common lived experiences of victim-survivors.

The proto-personas centred around four different protected characteristics: age, religion, ethnicity and gender reassignment and covered six different types of harms: cross-platform harassment, sexualised photo-shopping, gendered hate speech, dog-piling, intimate image-based abuse and cyberflashing.

# Limitations

This report has aimed to contribute to the evidence-building around the role of platform design in mitigating or exacerbating the risk of online VAWG. In doing so, we wish to acknowledge three key limitations of our research.

Firstly, given that our stakeholder interviews were predominantly conducted with UK, European Union (EU) and United States (US) based organisations and platforms, we recognise that not all of our findings or recommendations will apply to women and girls globally. We encourage more research into the experiences of women and girls in the Global South, who have traditionally been underrepresented with regards to online VAWG research.

Secondly, we aimed to shed light on the experiences of women and girls across protected characteristics but our findings have not spread equally across all. For example, there is less evidence on religion or belief and disability than on race and age. We have not elaborated on pregnancy and maternity or marriage and civil partnership due to a lack of data and information around these protected characteristics and online VAWG. Moreover, women and girls' experiences of online VAWG do not fit neatly into predefined categories of identity and most often are not documented by their protected characteristics. The intersections of identities undoubtedly plays a role in online VAWG, and needs to be recognised as such.

It is also important to note, that the absence of certain characteristics in the report does not mean that women and girls with these characteristics are not at increased risk of online VAWG. However, at this point in time we don't have the evidence necessary to speak to those experiences. This may be because some characteristics are less visible than others, and women and girls may not wish to openly report them.

Thirdly, our choice in large-scale social media and dating platforms allowed us to crystallise our focus. However, it also does not speak to the experiences of women and girls on other types and sizes of platforms. For example, online VAWG is also prevalent in gaming, messaging and adult platforms. Platforms with smaller user bases too have issues with online VAWG and are formulating solutions to combat it.

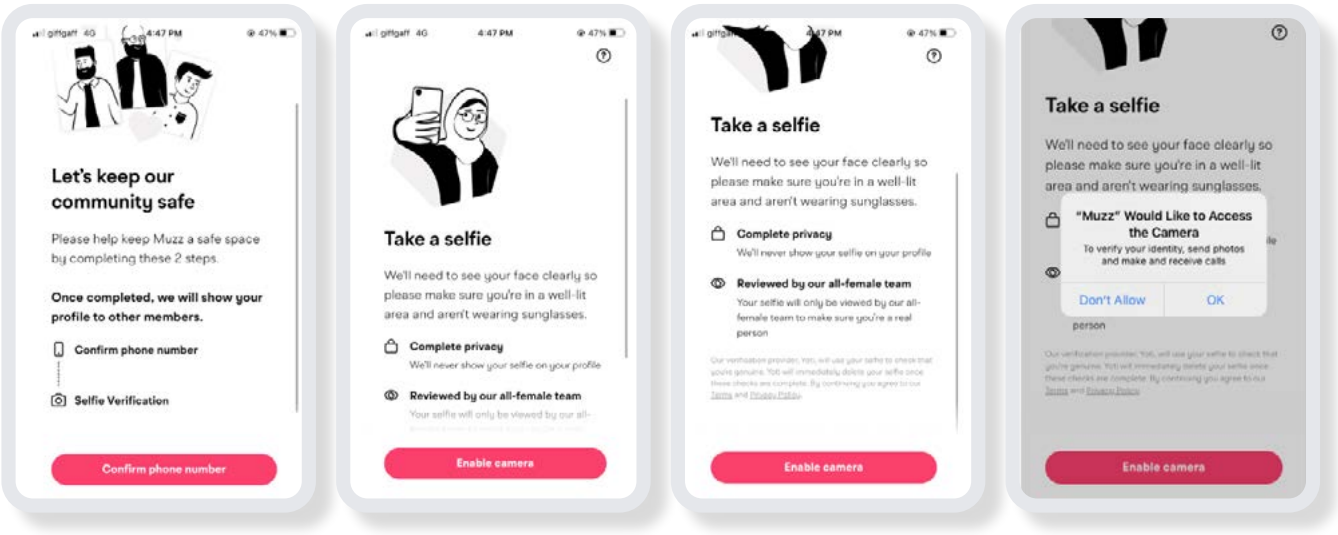
Lastly, whereas the literature review covered a large amount of sources available between October 2022 and March 2023, the online VAWG research is rapidly growing and there are sources that have not been reviewed for this report. Moreover, sources that were more focused on design, such as platform blog posts or announcements from civil society organisations, were not as robust as other more academic sources.

# Safety Features Case Studies

In assessing what is working well across the sector, the team identified three case studies of safety features that are good practice in mitigating against violence against women and girls.

## Case Study 1:
## Identity verification technology provided by Yoti within Muzz

Online marriage platform Muzz has partnered with online safety technology provider, Yoti, to moderate for identity verification. The online safety tech solution is embedded into the sign-up process and is a requirement for users to complete successfully before using Muzz and connecting with others. Additionally, it is free for all users demonstrating the inclusivity of the safety feature. Throughout the entire process, the user is informed of the purpose of the tool and usage of their data, ensuring user privacy and informed consent is properly in place.[157]

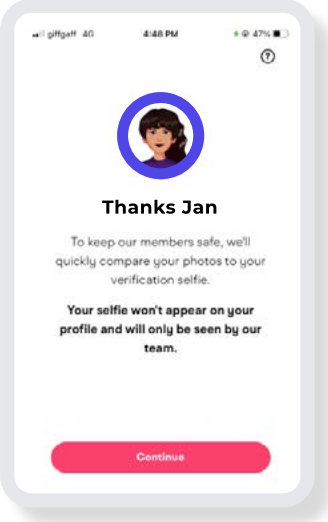



**Steps: 1-3**

1. Users at sign-up are required to take a 2-step identity verification process. The first step is to verify via a phone number to ensure the member is genuine (e.g. not a bot) and to prevent bad actors from re-accessing the app via a different account linked to the same number.

2. After phone verification is complete, the user is asked to conduct an image verification process run by the online safety tech company Yoti, by capturing a selfie to ensure "liveness" and comparing the selfie to the user's profile photo.

3. Before taking the selfie, users are informed that the selfie will not be shared on their Muzz profile and only be handled by Muzz's all-female review team, ensuring user privacy and proper data handling. The platform requests camera control and once access is enabled, a face shape box is shown for the user to place their face to capture a selfie. In real-time, the Yoti technology MyFace will use on-device recognition to automatically attempt to take a photo of your face.

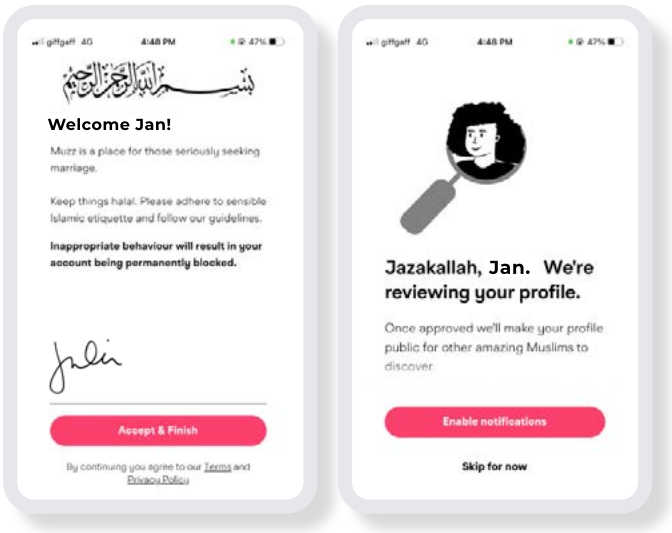



**Steps: 4-6**

4. The captured photo is firstly checked for "liveness", ensuring photos of images or screens, etc. are not approved. If the user tries to use a still photo or mask their face, the MyFace technology will detect the false or defective image and prompt the user to retake the selfie.

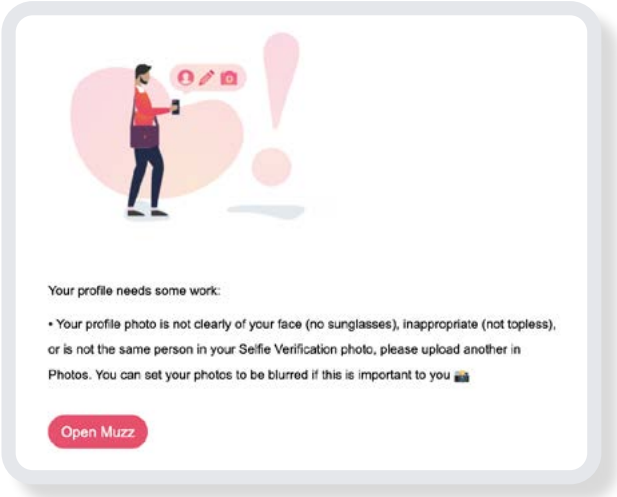
5. Once the user has captured a true Selfie, the face shape outline turns green and the user is notified their verification request will be reviewed by the Muzz team.

6. Once handed over to the Muzz team, the selfe is compared to the user's main profile photo to ensure they are the same person. The main profile photo is checked for inappropriate imagery, the photo is clear, and there is only a single face present. If the submission fails in any of these steps the Muzz team will ask members to resubmit the selfie verification process, suspending their access in the meantime until they have passed the process. If users fail the automation flow multiple times Muzz will send the profile to the all-female community team for a manual review.




**Steps: 7–8**

7. Once the selfie is submitted for review in the verification flow, the users are asked to agree to adhere to the community guidelines, and notified that their profile will only be made public for others to view once the verification process is successfully completed.
8. Once the profile and verification submission is approved, the profile will be shown on the app.
9. Only one selfie is required to complete the liveness check, and the entire process is completed within the Muzz platform ensuring an easy user flow. Within 24 hours, the Muzz review team will notify the user over email if their verification process has been successful, and if not, they will provide guidance on the next steps to resubmit their verification request.
10. Once the user's identity is verified, the profile becomes public to other users, and the user has access to all of Muzz's services.




**Step: 11**

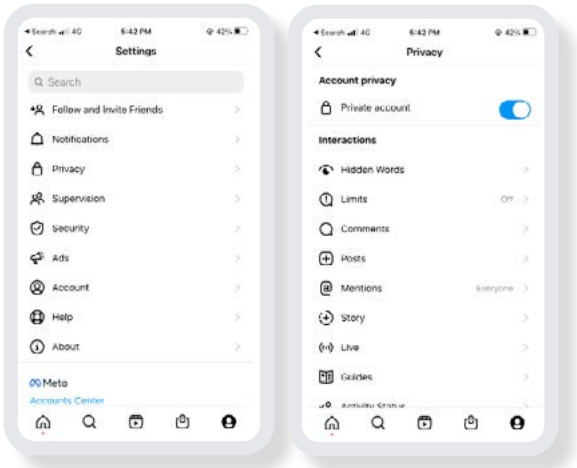
**11** If there are issues with the image uploaded, Muzz will send an email with additional information to the user. The email includes further safety guidelines, including flags against impersonation (e.g. no sunglasses, or this is not the same person in your Selfie Verification) and image abuse (e.g. inappropriate imagery). Critically, Muzz reminds users of their blurring feature to protect user privacy and safety at this stage. Additionally, alongside this process, Muzz offers optional ID verification to further confirm user identity.

The MyFace Technology provided by Yoti has proven to have a success rate of 90% for first attempts and 97% after 3 attempts. Additionally, the technology has a 1% False Positive Rate and a 90.5% True Positive Rate, meaning out of 100 attempts, one false image will be wrongfully approved, and about 91 real photos will be rightfully approved.[158]

## Case Study 2:

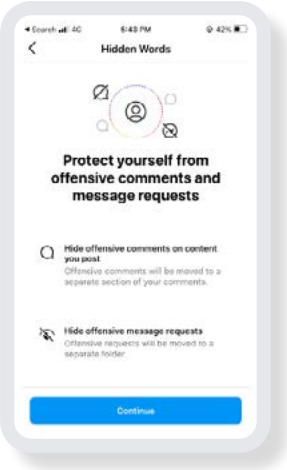
## Personalised moderation of harmful text with Instagram's *Hidden Words*

To address text-based harms, Instagram developed the *Hidden Words* feature where users can create a personal list of harmful terms to hide across the interface. This feature sits within privacy settings and aims to protect the user from harmful text-based content specific to them. The user has control over where the *Hidden Words* filter can be applied across the platform from content comments, connect requests, and soon-to-be extended to the Stories feature.[159] This provides users with the autonomy to decide how the moderation filter is applied across their online experience, and if they wish to apply it all. As one online dating platform put it, "for a lot of people being safe is being able to decide what, where and with who to share information". Instagram recognises that individuals experience harm differently as the user themselves forms the list of *Hidden Words* to monitor once the filter is applied, as explained in the user journey below.



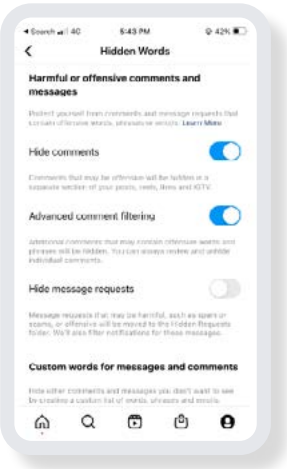

**Step: 1**

**1** The user navigats to settings, clicks privacy settings, and navigates to the Hidden Words option in the list of options.



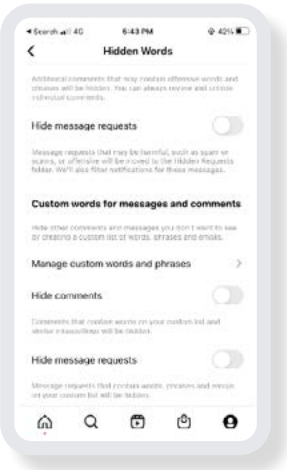

**Step: 2**

Users are presented with an information interface explaining the purpose of Hidden Words. This increases user awareness and understanding of how best to use this feature, encouraging safer online habits.



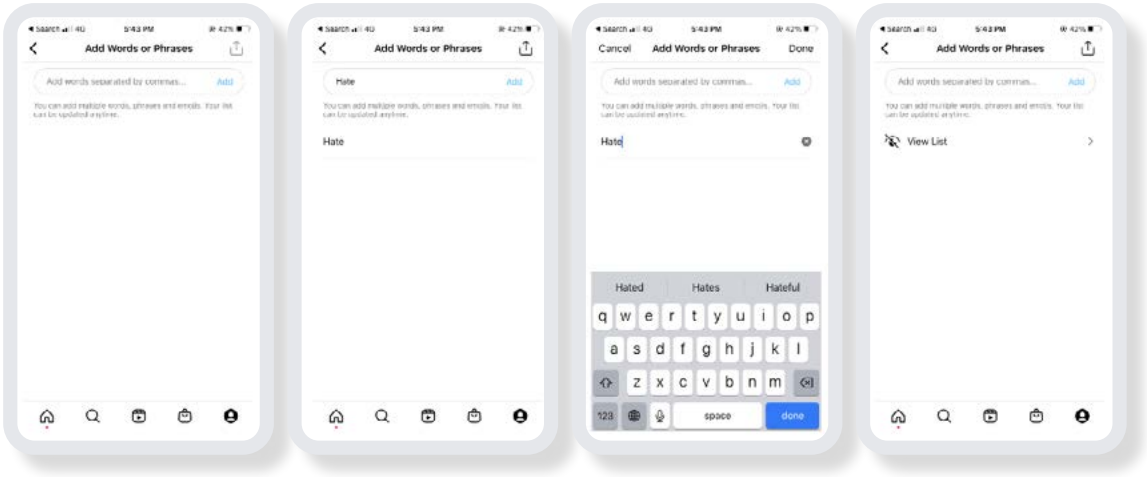

**Step: 3**

3 In the Hidden Words setting, users are presented with informational text as they work through the different settings, including applying Instagram's advanced filter for harmful content.




**Step: 4**

4 In the "Custom words for messages and comments" section users click "Manage custom words and phrases" to create their list of Hidden Words. Here they are able to add, delete, or amend their list of words to be hidden across the platform.




**Step: 5**

**5** The user turns on the Hidden Words personalised filter to comments and message requests by sliding the toggle for each feature, indicating you have the control to choose where the filter is applied. Once they have added a word, the list is not automatically invisible to the user, decreasing the exposure to harm. Users are able to review their list of phrases by clicking the "view list" button. The user is able to amend the list, adding or deleting words at their discretion recognising harm trends change over time and providing the users flexibility applying the moderation technique.

Instagram continues to build out the feature including extending it to different languages such as Farsi, Turkish, Russian, Bengali, Marathi, Telugu, and Tamil. Instagram is also improving the technology to spot intentional misspellings by perpetrators such as replacing the letter "I" with the number "1". Instagram reports, "When people turn on *Hidden Words* for comments, on average, they see 40% fewer comments that might be offensive". [160]

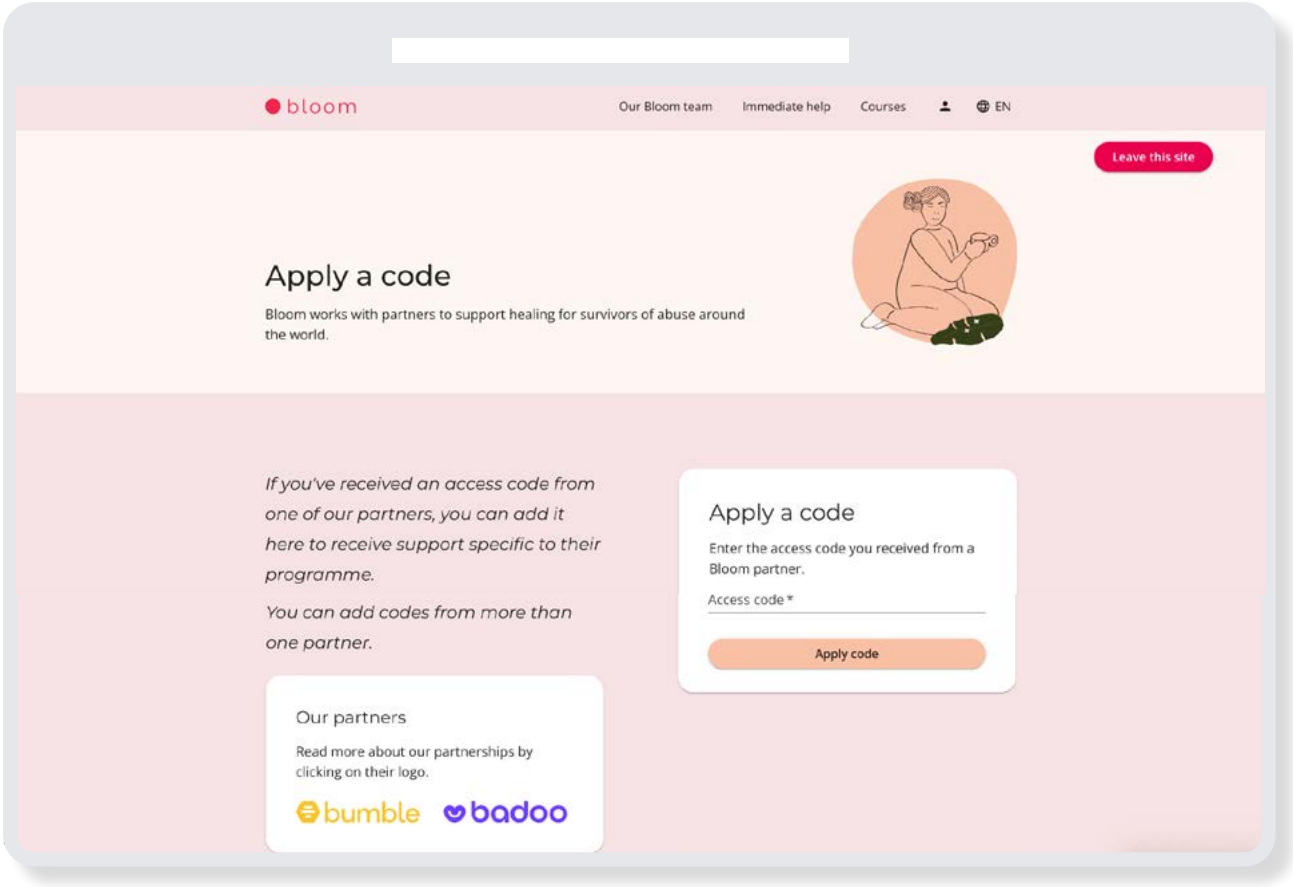
## Case Study 3:
## Providing tailored free support for victim-survivors through Bloom in partnership with Bumble

Stakeholder interviews highlighted the importance of tailored support services for victim-survivors of online VAWG to be provided within or as immediate steps following a report of harm in platforms. A strong case study of this is Bumble's partnership with the victim-survivor trauma support programme Bloom. Bloom is a free web-based support programme for Bumble and Badoo members who experience sexual harassment, assault or relationship abuse. Bloom is run by Chayn, a global non-profit run by survivors and allies from around the world who create online resources to support the healing of survivors of gender-based violence. Bloom aims to "inform and empower survivors by offering remote courses that combine the insights of survivors globally on trauma and gender-based violence with therapeutic practices to heal from trauma". [161]

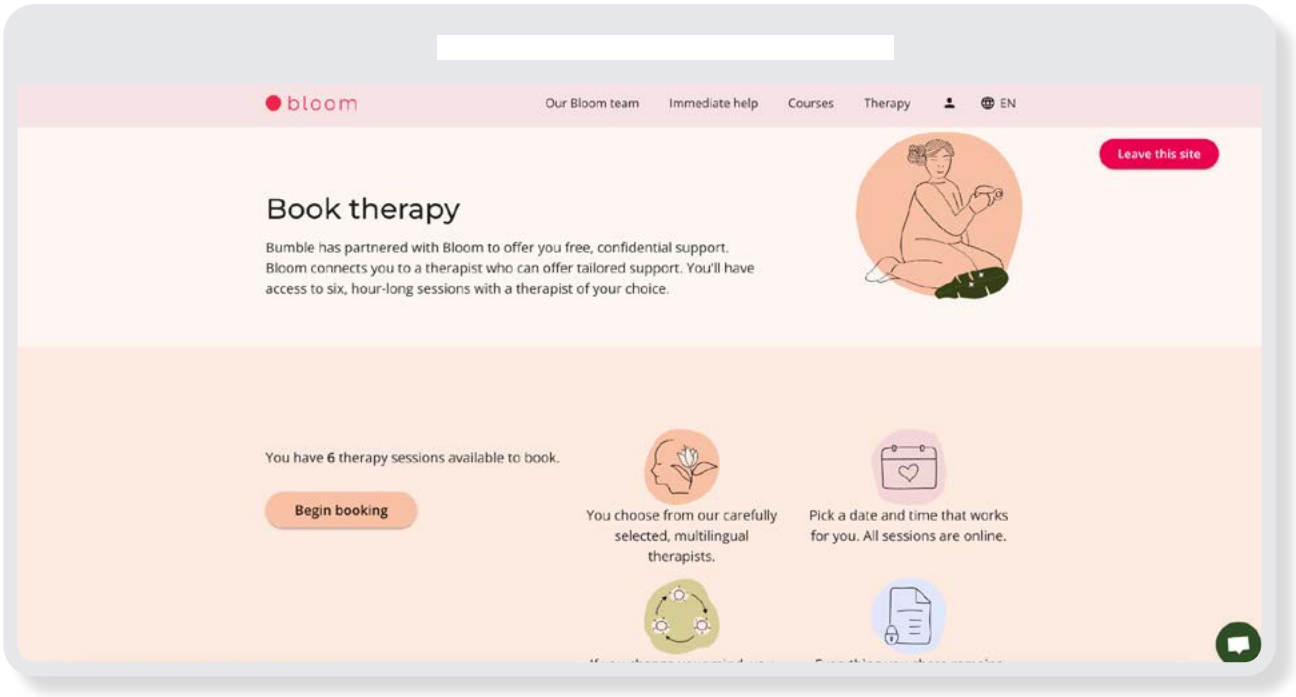
Through this programme, Bloom offers a range of services for Bumble and Badoo members to support their healing. This includes free access to Bloom's self-guided library of courses created by survivors and trauma-informed therapists, and access to Bloom's interactive one-to-one online chat with course facilitators. The self-paced courses include: i) Healing from Sexual Trauma, Society, ii) Patriarchy and Sexual Trauma, and iii) Dating, Boundaries, and Relationships.

In some cases, survivors will be offered access to free and confidential therapy sessions with a trauma-informed therapist. The support service respects user privacy, ensuring every participant's information remains completely confidential and uses end-to-end encryption in Bloom's one-to-one chat service. The user journey is as outlined below to access to Blooms support services.

1. When users report sexual harassment, assault, or relationship abuse in Bumble or Badoo, the platforms will respond to survivors with information about the unique Bloom offer.
2. Users are provided with a code for free access to Bloom's courses and one-to-one chat service.

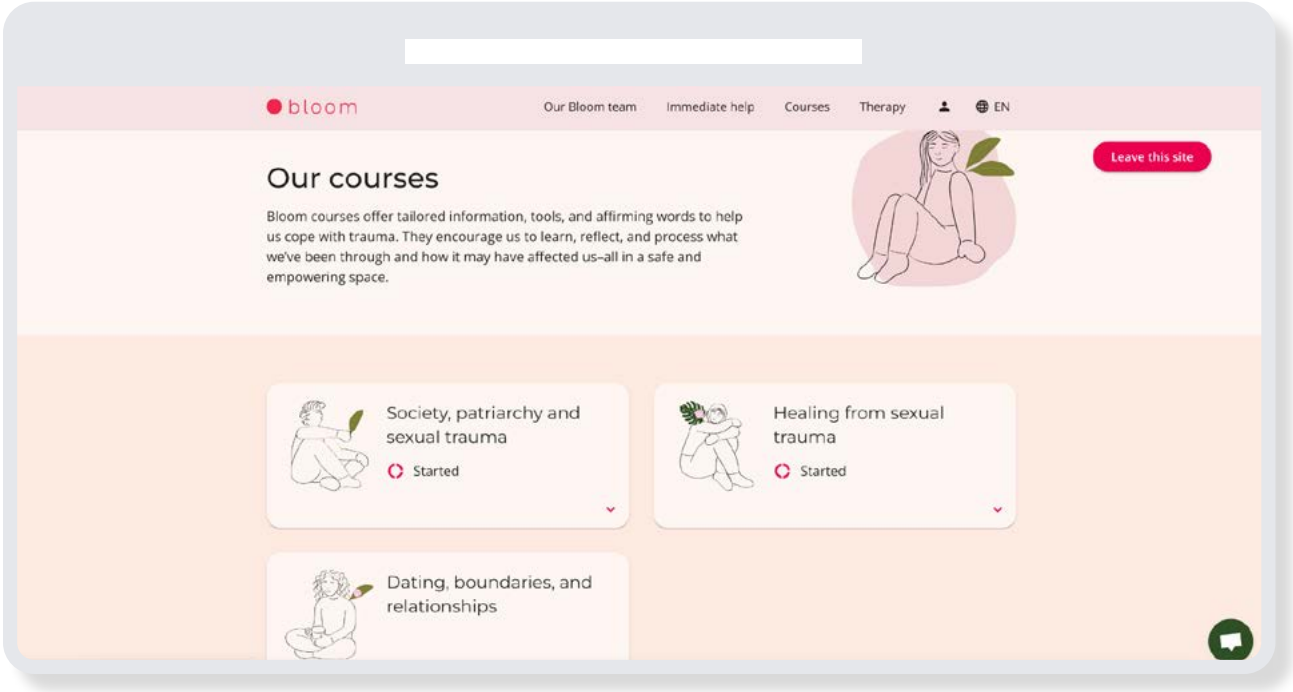



3 Users are then directed to the bespoke Bloom & Bumble/Badoo platform landing page, and asked to firstly enter their access code provided by the platforms, and once granted access, provide their name, email address, and create a secure password for their account.

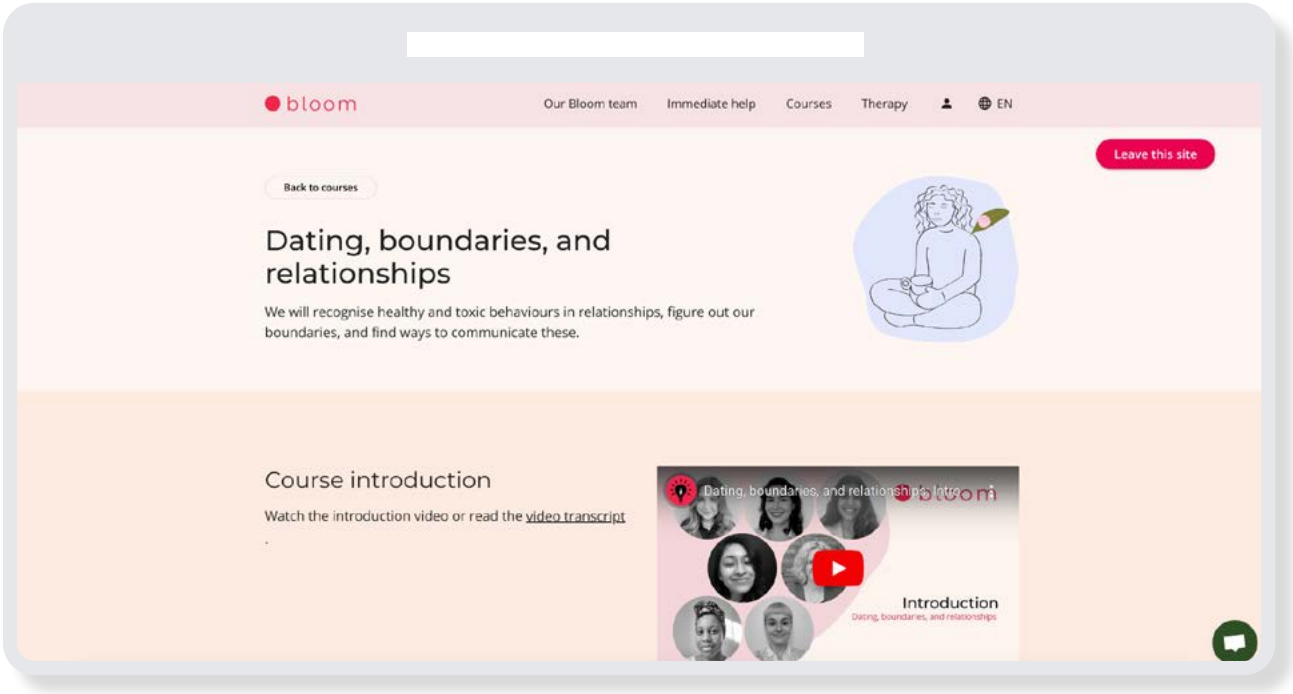


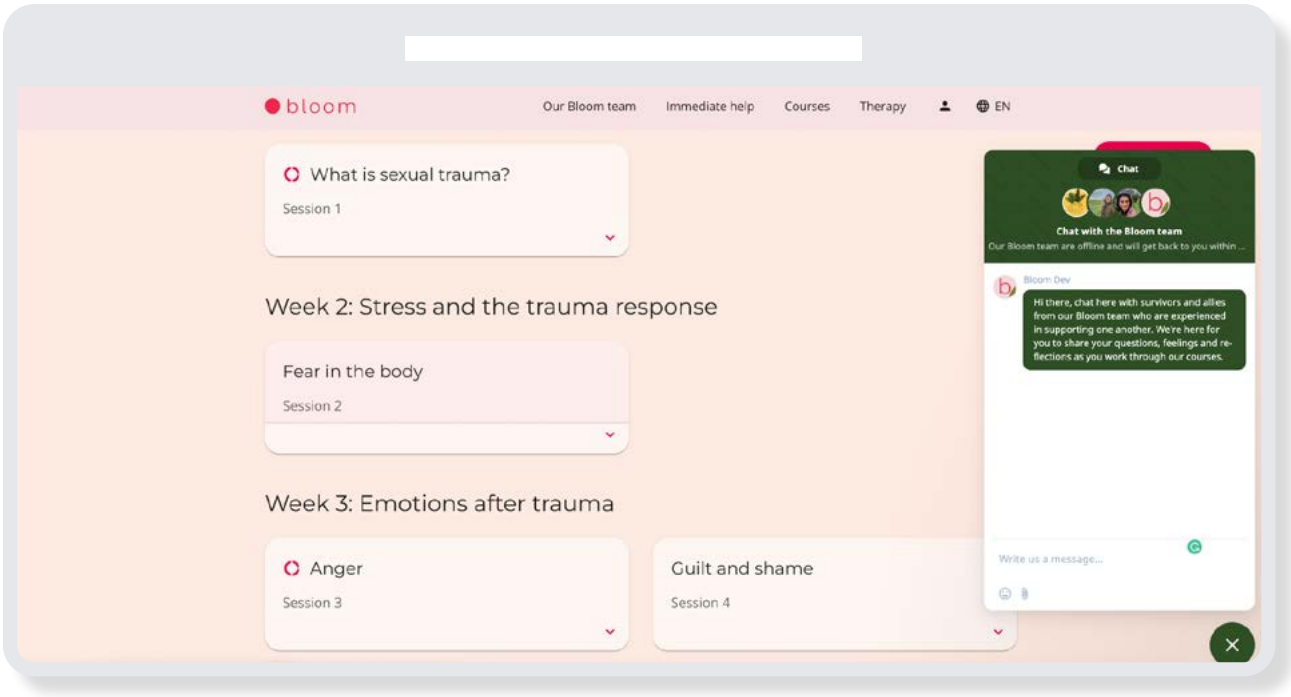

**Step: 4**

- In certain cases, survivors are given an access code to book therapy sessions with a trauma-informed therapist of their choosing, in either English, French, Hindi, Spanish or Portuguese.




**Step: 5**




**Step: 5**

**5** Users can explore the sessions on offer, use the chat at their own pace, and book their free therapy sessions (if applicable).

The support provided by Bloom is not only trauma-informed but it is specific to the experiences of those who have experienced harm on the two platforms, Bumble and Badoo, and therefore provides a very tailored support for recovery and healing. Additionally, the range of services provides victim-survivors the option to choose the best method for their personal healing. This service is a good example of how a simple integration of support services within reporting functions can immediately provide proper support for victim-survivors. The reporting flow now includes not only action for permanent harm prevention and proper legal action, but just as importantly, it paves a clear pathway for victim-survivors' healing journeys through tailored support.

# Endnotes

1 For the purpose of this project, online violence against women and girls (online VAWG) is defined as violence conducted disproportionately against women and girls through or on cyber-enabled devices.

2 The Economist Intelligence Unit, Measuring the prevalence of online violence against women

3 UN Women, Accelerating efforts to tackle online and technology facilitated violence against women and girls and HM Government, Tackling violence against women and girls

4 Plan International, Free To Be Online?- Girls' and young women's experiences of online harassment ; Council of Europe, Protecting Women and Girls From Violence in the Digital Age and The Economist Intelligence Unit, Measuring the prevalence of online violence against women

5 US Department of State, 2022 Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse

6 The European Union Agency for Fundamental Rights, Violence against women: an EU-wide survey. Main results report

7 Australian eSafety Commissioner, Women in the spotlight: how online abuse impacts women in their working lives

8 UN Women, Online and ICT-facilitated violence against women and girls during Covid-19

9 The Global Partnership for Action on Gender-Based Online Harassment and Abuse, 2022 Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse,

10 While there is no universal definition of online VAWG, this report uses the definition from PUBLIC, DSIT, OSTIA and Faculty's previous research for DCMS and the Home Office on Online Violence Against Women & Girls (VAWG). This definition, or versions of this definition, is also used across the international community as seen in the UN Human Rights Council's Report of the Special Rapporteur on Violence Against Women, Its Causes and Consequences on Online Violence Against Women and Girls from a Human Rights Perspective

11 For the purpose of this report a design feature is defined as a touchpoint which enables the user or platform to take an action at various points in the user journey that affects the user experience.

12 Global North refers to the group of countries that are in Europe, North America and the developed parts of Asia. Cambridge Dictionary, Global North.

13 Throughout the report we use the term 'victim-survivor' to denote women and girls who have experienced online VAWG, to reflect agency as well as victimisation.

14 UN Women, CSW67

15 UN Women, Technology and Innovation for Gender Equality Action Coalition Blueprint

16 The International Group of Seven (G7) is an intergovernmental political forum consisting of Canada, France, Germany, Italy, Japan, the UK, and the US. The EU is a "non-enumerated member". It discusses a number of topics and shapes political responses to global challenges. Council on Foreign Relations, Where is the G7 Headed? Government of Canada, Charlevoix commitment to end sexual and gender-based violence, abuse and harassment in digital contexts

17 Council of Europe, Protecting women and girls from violence in the digital age ; The White House, Launching the Global Partnership for Action on Gender-Based Online Harassment and Abuse ; UN Women, Accelerating efforts to tackle online and technology facilitated violence against women (VAWG) ; European Commission, Violence against women: the European Union established an EU-wide helpline number and calls to end violence against women worldwide

18 HM Government, Tackling Violence Against Women and Girls

19 Australia eSafety Commissioner, eSafety Women

20 The White House, FACT SHEET: Presidential Memorandum Establishing the White House Task Force to Address Online Harassment and Abuse

21 Ministry of Justice Korea, Digital Sex Crimes task force team expert committee; Council of Europe, No space for violence against women and girls in the digital world and Citizens Information, Sharing of Intimate Images Without Consent

22 World Wide Web Foundation, Women's rights online score cards - Peru

23 Council of Europe, No space for violence against women and girls in the digital world

24 World Wide Web Foundation, Women's rights online score cards - global summary Country score cards

25 World Wide Web Foundation, "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing, and Chayn & End Cyber Abuse, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TBGV

26 UN Women, Accelerating efforts to tackle online and technology facilitated violence against women and girls and HM Government, Tackling violence against women and girls

27 Refuge, Refuge Anniversary: 50 years on domestic abuse is getting smarter

28 Suzy Lamplugh Trust, Unmasking stalking

29 UN Women, Intensification of efforts to eliminate all forms of violence against women and girls Report of the Secretary-General

30 UK Safer Internet Centre, Helpline Reports 2021

31 End Violence Against Women and Glitch, The Ripple Effect: Covid-19 and the epidemic of online abuse

32 Chayn & End Cyber Abuse, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TBGV

33 Australian Research Council , Shattering Lives and Myths: A Report on Image-Based Sexual Abuse and Durham University, 'Devastating, like it broke me': Responding to image-based sexual abuse in Aotearoa New Zealand

34 Australian Research Council , Shattering Lives and Myths: A Report on Image-Based Sexual Abuse and Durham University, 'Devastating, like it broke me': Responding to image-based sexual abuse in Aotearoa New Zealand

35 Gender Based Violence AoR, Release of GBV AoR HelpDesk Learning Brief Technology Facilitated Gender-Based Violence, Learning Brief 1: Understanding technology-facilitated gender-based violence

36 UN Women, Intensification of efforts to eliminate all forms of violence against women and girls Report of the Secretary-General; Chayn & End Cyber Abuse, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TBGV

37 UN Women, Intensification of efforts to eliminate all forms of violence against women and girls Report of the Secretary-General

38 Plan International, Free to be online? Girls and young women's experiences of online harassment

39 World Wide Web Foundation, Glitch & Social Finance, Strengthening Accountability for Online Gender-Based Violence - one year later

40 Global South refers to a group of countries that are in Africa, Latin America and the developing parts of Asia. Cambridge Dictionary, Global South

41 International Center for Journalists, The chilling: a global study of online violence against women journalists

42 Pollicy, Alternative Realities, Alternate Internets - African Feminist Research for a Feminist Internet

43 World Wide Web Foundation, "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing and Carnegie UK, End Violence Against Women, Glitch, National Society for the Prevention of Cruelty to Children, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

44 World Wide Web Foundation, Glitch & Social Finance, Strengthening Accountability for Online Gender-Based Violence - one year later

45 Equality Human Rights Commision, Protected Characteristics; UN Women, Accelerating efforts to tackle online and technology-facilitated violence against women and girls

46 Plan International, Free To Be Online?- Girls' and young women's experiences of online harassment

47 Thorn, Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021

48 Thorn, Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021

49 Internet Watch Foundation, The Annual Report 2021

50 The Economist Intelligence Unit, Measuring the prevalence of online violence against women

51 Plan International, Free To Be Online?- Girls' and young women's experiences of online harassment, End Violance Against Women, Violence against women and girls Snapshot report 2021-2022 , The Global Partnership for Action on Gender-Based Online Harassment and Abuse, 2022 Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse, NSPCC, Delivering a Duty of Care: An assessment of the Government's proposals against the NSPCC's six tests for the Online Safety Bill, Durham University, 'Devastating, like it broke me': Responding to image-based sexual abuse in Aotearoa New Zealand and Amnesty International, Troll Patrol Findings

52 EVAW, Violence against women and girls Snapshot report 2021-2022

53 Thorn, Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021

54 EVAW, Violence against women and girls Snapshot report 2021-2022

55 Plan International, Free To Be Online?- Girls' and young women's experiences of online harassment

56 Stay Safe East and Inclusion London, Response to Online Harms White Paper for London DDPO Hate crime partnership 2019

57 Ofcom, Online Nation 2022

58 Thorn, Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021

59 Wilson Center, Malign Creativity - How gender, sex and lies are weaponised online

60 End Violence Against Women and Glitch, The Ripple Effect: Covid-19 and the epidemic of online abuse

61 UN Women, Accelerating efforts to tackle online and technology-facilitated violence against women and girls, and Equality and Human Rights Commission, Protected Characteristics

62 ICFJ, The Chilling: A global study of online violence against women journalist

63 Council of Europe, No space for violence against women and girls in the digital world, Global Partnership for Action of Gender-Based Online Harassment and Abuse and UN Women & Wilton Park, Building a Shared Agenda on the Evidence Base for Gender-Based Online Harassment and Abuse, and Australian Research Council, Shattering Lives and Myths: A Report on Image-Based Sexual Abuse

64 UN Women, Accelerating efforts to tackle online and technology-facilitated violence against women and girls

65 Women's Aid, Online and Digital abuse

66 Gender-Based Violence AoR, Release of GBV AoR HelpDesk Learning Brief Technology Facilitated Gender-Based Violence

67 UN Women, Violence against women in the online space

68 Plan International, Free To Be Online? and Council of Europe, Protecting Women and Girls From Violence in the Digital Age

69 Chayn, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TGBV; Australian Research Council , Shattering Lives and Myths: A Report on Image-Based Sexual Abuse and Council of Europe, Protecting Women and Girls From Violence in the Digital Age

70 The Economist Intelligence Unit, Measuring the prevalence of online violence against women

71 UN Women, Violence against women in the online space; Childnet, Young people's experiences of online sexual harassment

72 Centre for Countering Digital Hate, Hidden Hate: How Instagram fails to act on 9 in 10 reports of misogyny in DMs

73 Australia eSafety Commissioner, Can I just share my story? Experiences of technology-facilitated abuse among Aboriginal and Torres Strait Islander women from regional and remote areas

74 Thorn, Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021

75 Childnet, Young people's experiences of online sexual harassment

76 Suzy Lamplugh Trust, Personal safety and online dating

77 Suzy Lamplugh Trust, Personal safety and online dating

78 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

79 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

80 DSIT, Anonymous or multiple accounts: improve the safety of your online platform and Revealing Reality, Abuse and Anonymity

81 Girlguiding, Girls Attitudes Survey 2022

82 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

83 UN Women, Accelerating efforts to tackle online and technology facilitated violence against women (VAWG)

84 WWWF, "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing

85 ICFJ, The chilling: a global study of online violence against women journalists

86 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

87 UN Women, Accelerating efforts to tackle online and technology facilitated violence against women and girls and IT for Change, Born Digital, Born Free? A socio-legal study on young women's experiences of online violence in South India

88 IT for Change, Born Digital, Born Free? A socio-legal study on young women's experiences of online violence in South India

89 Noble, S.U Algorithms of Oppression and Buolamwini, J. Poet of Code

90 Heinrich Böll Stiftung Institute, Algorithmic misogynoir in content moderation practise

91 EU Disinfo Lab, Gender-based disinformation: advancing our understanding and response

92 Wilson Center, Malign Creativity - How gender, sex and lies are weaponised online

93 Wilson Center, Malign Creativity - How gender, sex and lies are weaponised online

94 Augmented reality (AR) overlays your view of the actual world with digitally-generated real-time sound and vision. Virtual reality (VR) uses computer hardware and software to create an artificial environment that looks and sounds as if you are really there. Mixed reality (MR) combines elements of both AR and VR by blending digital content into the physical environment so you see and hear the virtual elements as an extension of reality. Australia eSafety Commissioner, Immersive technologies - position statement

95 MIT Technology Review, The Metaverse has a groping problem already and Centre for Counter Digital Hate, Facebook's metaverse

96 Australia eSafety Commissioner, Immersive technologies - position statement; Centre for International Governance Innovation, Deepfakes and Digital Harms: Emerging Technologies and Gender-Based Violence; Sensity, The State of Deepfakes; Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice and Wilson Center, Malign Creativity - How gender, sex and lies are weaponised online

97 Zoom-bombing refers to someone showing up uninvited to a Zoom meeting and causing disturbance, such as sharing pornographic content; The Economist Intelligence Unit, Measuring the prevalence of online violence against women

98 NSPCC, Livestreaming and video-chatting - snapshot 2

99 NSPCC, Livestreaming and video-chatting - snapshot 2

100 Australia eSafety Commissioner, Basic Online Safety Expectations - summary of industry responses to the first mandatory transparency notices

101 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

102 As outlined in the methodology, the eight key platforms are Facebook, Instagram, Snapchat, TikTok, Twitter, YouTube, Bumble and Tinder.

103 Policy, Alternate Realities, Alternate Internets African Feminist Research for a Feminist Internet; Childnet, Young people's experiences of online sexual harassment

104 WWWF, "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing

105 Seam Gen, friction in UX design can be beneficial

106 Revealing Reality, Abuse and anonymity

107 Revealing Reality, Abuse and anonymity

108 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

109 Government of Canada, Charlevoix commitment to end sexual and gender-based violence, abuse and harassment in digital contexts

110 User journey inspired by Ofcom's A-SPARC model, The A–SPARC model of online platforms

111 Centre for Countering Digital Hate, Deadly by Design and Digital Hate

112 Refuge, Unsocial Spaces

113 WWWF, "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing

114 For a definition of "doxxing", please see the online VAWG taxonomy in the Appendix

115 Australian eSafety Commissioner, Women in the spotlight: how online abuse impacts women in their working lives

116 Childnet, Young people's experiences of online sexual harassment

117 NSPCC, Delivering a Duty of Care An assessment of the Government's proposals against the NSPCC's six tests for the Online Safety Bill

118 Revealing reality, Not just flirting - the unequal experiences and consequences of nude image-sharing by young people

119 Ultraviolet, Putting the onus on women is a PR stunt – the platforms are the problem and Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

120 Recognising that radicalisation can happen to all users, stakeholders had heightened concerns around men and boys.

121 The New Yorker, K-Pop Fans Defuse Racist Hashtag

122 Twitter, Our approach to recommendations

123 End Violence Against Women and Glitch, The Ripple Effect: Covid-19 and the epidemic of online abuse

124 Refuge, Unsocial Spaces

125 End Violence Against Women and Glitch, The Ripple Effect: Covid-19 and the epidemic of online abuse

126 Childnet, Young people's experiences of online sexual harassment

127 UN Women, Violence against women in the online space and Suzy Lamplugh Trust, Personal safety and online dating

128 Refuge, Unsocial Spaces

129 Australian Research Council, Shattering Lives and Myths: A Report on Image-Based Sexual Abuse

130 Examples include StopNCII.org, TakeItDown and

Revenge Porn Helpline

131 Human Rights Watch, 'My life is not your porn' digital sex crimes in South Korea

132 EU Disinfo Lab, Gender-based disinformation: advancing our understanding and response

133 Australia eSafety Commissioner, Basic Online Safety Expectations - summary of industry responses to the first mandatory transparency notices

134 Business Wire, Google's Jigsaw Unit Open Sources Code for Harassment Manager Tool in Partnership with Twitter, Thomson Reuters Foundation

135 Thomas Reuters Foundation, TRFilter.org

136 The Online Dating Excellence Association aims to convene dating platforms and leaders through sharing leading news and host events see their website at www.internetdatingexcellence association.com

137 Meta, Women's Safety Hub

138 LinkedIn, Keeping LinkedIn a Safe, Professional Community Where Everyone Can Thrive

139 Refuge Tech Safety, Digital Breakup Tool

140 TechCrunch, Instagram and Facebook introduce more limits on targeting teens with ads and The Verge, TikTok is testing a way to reset your For You page

141 Global Dating Insights, happn Launches New 'Hub' Feature

142 TechCrunch, Tinder rolls out new safety features, including an Incognito Mode

143 TechCrunch, Tinder and other Match dating apps will offer in-app tips on avoiding romance scams

144 UK Government, Principles of safer online platform design

145 Australia eSafety Commissioner, Safety by design

146 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

147 DCMS, The UK Safety Tech Sector: 2022 Analysis

148 Yoti, Muzz introduces selfie and identity verification from Yoti for safer Halal dating and Reddit, Oterlu Team Joins Reddit to Accelerate Safety Efforts

149 DSIT, Principles of safer online platform design ; Australia eSafety Commissioner, Safety by design

150 Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., Violence Against Women and Girls (VAWG) Code of Practice

151 Chayn and End Cyber Abuse, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TGBV

152 Refuge, Unsocial Spaces and National Democratic Institute, Interventions for ending online violence against women in politics

153 Chayn and End Cyber Abuse, Orbits: a global field guide to advance intersectional, survivor-centred, and trauma-informed interventions to TBGV

154 Tinder, Tinder Introduces Are You Sure?, an Industry-First Feature That is Stopping Harassment Before It Starts

155 Instagram, Our commitment to lead the fight against online bullying and Tinder, Tinder Introduces Are You Sure?, an Industry-First Feature That is Stopping Harassment Before It Starts

156 Chayn, We're partnering with Bumble to bring Bloom to their users

157 Yoti, Muzz introduces selfie and identity verification from Yoti for safer Halal dating

158 Yoti, Yoti MyFace liveness white paper

159 Instagram, New Updates to Hidden Words

160 Instagram, New Updates to Hidden Words

161 Bloom, Welcome to Bloom

# Bibliography


## A

Amnesty International, 2018, "Troll Patrol Findings"

Australia eSafety Commissioner, 2022, "Basic Online Safety Expectations - Summary of Industry Responses to the First Mandatory Transparency Notices"

Australia eSafety Commissioner, "Safety by design"

Australia eSafety Commissioner, "eSafety Women"

Australia eSafety Commissioner, 2022, "Women in the Spotlight: How Online Abuse Impacts Women in their Working Lives"

Australia eSafety Commissioner, 2020, "Immersive Technologies - Position Statement"

Australian Research Council, 2019, "Shattering Lives and Myths: A Report on Image-Based Sexual Abuse"

## B

Bloom, "Welcome to Bloom"

Bumble, "Community Guidelines"

Buolamwini, J., "Poet of Code"

Business Wire, 2022, "Google's Jigsaw Unit Open Sources Code for Harassment Manager Tool in Partnership with Twitter, Thomson Reuters Foundation"

## C

Cambridge Dictionary, "Cisnormativity"

Cambridge Dictionary, "Global North"

Cambridge Dictionary, "Global South"

Carnegie UK, EVAW, Glitch, NSPCC, Refuge, 5Rights Foundation, McGlynn, C., Woods, L., 2022, "Violence Against Women and Girls Code of Practice"

Center for Countering Digital Hate (CCDH), 2022, "Deadly by Design - TikTok Pushes Harmful Content Promoting Eating Disorders and Self-harm into Users' Feeds"

Center for Countering Digital Hate (CCDH), 2022, "Digital Hate - Social Media's Role in Amplifying Dangerous Lies About LGBTQ+ People"

Centre for Counter Digital Hate (CCDH), 2021, "Facebook's Metaverse"

Centre for International Governance Innovation, 2020, "Deepfakes and Digital Harms: Emerging Technologies and Gender-Based Violence "

Centre for International Governance and Technology, 2020, "Technology-Facilitated Gender-Based Violence"

Chayn & End Cyber Abuse, 2022, "Orbits: A Global Field Guide to Advance Intersectional, Survivor-Centred, and Trauma-Informed Interventions to TBGV"

Chayn, 2021, "We're Partnering with Bumble to Bring Bloom to their Users"

Childnet, "Online Sexual Harassment"

Childnet, 2017, "Young People's Experiences of Online Sexual Harassment"

Citizens Information, 2021, "Sharing of Intimate Images Without Consent"

Council of Europe, 2022, "No Space for Violence Against Women and Girls in the Digital World"

Council of Europe, 2021, "Protecting Women and Girls from Violence in the Digital Age"

Council on Foreign Relations, 2022, "Where is the G7 Headed?"

Crown Prosecution Service, 2019, "Extreme Pornography"

## D

DSIT, 2021, "Anonymous or Multiple Accounts: Improve the Safety of your Online Platform"

DSIT, 2022, "The UK Safety Tech Sector: 2022 Analysis"

Due Diligence Project, 2017, "Due Diligence and Accountability for Online Violence Against Women"

Durham University, 2022, "'Devastating, like it broke me': Responding to Image-Based Sexual Abuse in Aotearoa New Zealand"

## E

End Violence Against Women and Glitch, 2020, "The Ripple Effect: Covid-19 and the epidemic of online abuse"

End Violence Against Women, 2021, "Violence Against Women and Girls Snapshot report 2021-2022"

Equality Human Rights Commission, "Protected Characteristics"

EU Disinfo Lab, 2021, "Gender-Based Disinformation: Advancing our Understanding and response"

European Commission, 2022, "Violence Against Women: the European Union Established an EU-wide Helpline Number and Calls to End Violence Against Women Worldwide"

## F

Facebook, "Child Sexual Exploitation Abuse and Nudity Policy"

Forbes, 2022, "Bumble Partners With Bloom To Launch Complimentary Trauma Support For Sexual Assault Survivors"

## G

Gender Based Violence AoR, 2021, "Release of GBV AoR HelpDesk Learning Brief Technology Facilitated Gender-Based Violence, Learning Brief 1: Understanding Technology-Facilitated Gender-Based Violence"

Girlguiding, 2022, "Girls Attitudes Survey 2022"

Glitch, 2021, "A Little Means A Lot - How You Can Be an Online Active Bystander"

Global Dating Insights, 2023, "happn Launches New 'Hub' Feature"

Global Network on Extremism & Technology, 2022, "The Cissexist Assemblages of Content Moderation"

Global Partnership for Action of Gender-Based Online Harassment and Abuse, Wilton Park and UN Women, "Building a Shared Agenda on the Evidence Base for Gender-Based Online Harassment and Abuse"

Government of Canada, 2018, "Charlevoix Commitment to End Sexual and Gender-based Violence, Abuse and Harassment in Digital Contexts"

Grindr, 2022, "Best Practices for Gender-Inclusive Content Moderation"

## H

Heinrich Böll Stiftung Institute, 2021, "Algorithmic Misogynoir in Content Moderation Practise"

HER, "Community Guidelines"

HM Government, 2021, "Tackling Violence Against Women and Girls"

HM Government, 2021, "Identifying the challenges and solutions for tackling online Violence Against Women and Girls"

Human Rights Watch, 2021, " 'My life is not your porn' Digital Sex Crimes in South Korea"

## I

International center for journalists (ICFJ), 2022, "The Chilling: A Global Study of Online Violence Against Women Journalists"

INHOPE, 2023, "LinkedIn"

Internet Watch Foundation, 2021, "The Annual Report 2021"

Instagram, "Community Guidelines"

Instagram, 2022, "New Updates to Hidden Words"

Instagram, 2019, "Our commitment to lead the fight against online bullying"

IT for Change, 2019, "Born Digital, Born Free? A socio-legal study on young women's experiences of online violence in South India"

## L

Law Commission, 2021, "Modernising Communication Offences"

LinkedIn, 2022, "Keeping LinkedIn a Safe, Professional Community Where Everyone Can Thrive"

## M

Match, "Community Guidelines"

McGlynn, C. et al, 2017, "Beyond 'Revenge Porn': The Continuum of Image-Based Sexual Abuse"

Meta, "Women's Safety Hub"

Ministry of Justice Korea, 2022, "Digital Sex Crimes task force team expert committee"

MIT Technology Review, 2021, "The Metaverse has a groping problem already"

Muzz, Behaviour Guidelines

## N

National Democratic Institute, 2022,"Interventions for ending online violence against women in politics"

NSPCC, 2021, "Delivering a Duty of Care: An assessment of the Government's proposals against the NSPCC's six tests for the Online Safety Bill"

NSPCC, 2018, "Livestreaming and video-chatting - snapshot 2"

Noble, S.U., 2021, "Algorithms of Oppression"

NotYourPorn, "Our Story"

## O

Ofcom, 2021, "The A–SPARC model of online platforms"

Ofcom, 2022, "Online Nation 2022"

Oxford University Press, 2017, "A Dictionary of Gender Studies"

## P

Plan International, 2020, "Free to be online? Girls and young women's experiences of online harassment"

PEN America, 2021, "Defining "Online Abuse": A Glossary of Terms"

Pew Research Centre, 2017, "Online Harassment 2017"

Policy, 2020, "Alternative Realities, Alternate Internets - African Feminist Research for a Feminist Internet"

## R

Reddit, 2022, "Oterlu Team Joins Reddit to Accelerate Safety Efforts"

Refuge, 2021, "Refuge Anniversary: 50 years on domestic abuse is getting smarter"

Refuge, 2021, "Unsocial Spaces"

Refuge Tech Safety, "Digital Breakup Tool"

Revealing Reality, 2022, "Abuse and anonymity"

Revealing reality, 2022, "Not just flirting - the unequal experiences and consequences of nude image-sharing by young people"

Revenge porn helpline, 2021, "Intimate image abuse, an evolving landscape"

## S

Safiya Umoja Noble, 2018, "Algorithms of Oppression"

Seam Gen, 2019, "Friction in UX design can be beneficial"

Sensity, 2019, " The State of Deepfakes- Landscape, threats and impact"

#Shepersisted, 2023, "Monetizing misogyny - gendered disinformation and the undermining of women's rights and democracy globally"

Snapchat, "Community Guidelines"

SoSyncd, "Community Guidelines"

Stay Safe East and Inclusion London, 2019, "Response to Online Harms White Paper for London DDPO Hate crime partnership 2019"

StopNCII.org, "Industry Partners"

Suzy Lamplugh Trust, 2020, "Personal safety and online dating"

Suzy Lamplugh Trust, 2021, "Unmasking stalking"

## T

TechCrunch, 2022, "Bumble open sourced its AI that detects unsolicited nudes"

TechCrunch, 2023, "Instagram and Facebook introduce more limits on targeting teens with ads"

TechCrunch, 2023, "Tinder and other Match dating apps will offer in-app tips on avoiding romance scams"

TechCrunch, 2023, "Tinder rolls out new safety features, including an Incognito Mode"

Tinder, 2023, "Tinder Introduces Are You Sure?, an Industry-First Feature That is Stopping Harassment Before It Starts"

The Economist Intelligence Unit, 2021, "Measuring the prevalence of online violence against women"

The European Union Agency for Fundamental Rights, 2014, "Violence against women: an EU-wide survey. Main results report"

The Global Partnership for Action on Gender-Based Online Harassment and Abuse, 2022, "2022 Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse"

The New Yorker, 2022, "K-Pop Fans Defuse Racist Hashtag"

The Verge, 2023, "TikTok is testing a way to reset your For You page"

The White House, 2022, "Launching the Global Partnership for Action on Gender-Based Online Harassment and Abuse"

The White House, 2022, "FACT SHEET: Presidential Memorandum Establishing the White House Task Force to Address Online Harassment and Abuse"

Thomas Reuters Foundation, TRFilter.org

Thorn, 2021, "Responding to Online Threats: Minors' Perspective on Disclosing, Reporting, and Blocking in 2021"

Tiktok, "Community Guidelines"

Tinder, "Community Guidelines"

Twitter, Our approach to recommendations

## U

UK Government, 2021, "Principles of Safer Online Platform Design"

UK Safer Internet Centre, 2021, "Helpline Reports 2021"

Ultraviolet, 2021, "Putting the onus on women is a PR stunt – the platforms are the problem"

UN Human Rights Council, 2018, "Report of the Special Rapporteur on Violence Against Women, Its Causes and Consequences on Online Violence Against Women and Girls from a Human Rights Perspective"

UN Women, 2023, "CSW67 (2023)"

UN Women, 2022, "Accelerating efforts to tackle online and technology facilitated violence against women (VAWG)"

UN Women 2022, "Intensification of efforts to eliminate all forms of violence against women and girls Report of the Secretary-General"

United Nations Entity for Gender Equality and the Empowerment of Women (UN Women), "Online and ICT-facilitated violence against women and girls during Covid-19"

UN Women, 2021, "Technology and Innovation for Gender Equality Action Coalition Blueprint"

UN Women, 2021, "Violence against women in the online space - insights from a multi-country study in the Arab States"

## W

Wilson Center, 2021, "Malign Creativity - How gender, sex and lies are weaponised online"

Women's Aid, "Online and Digital abuse"

Women's Media Center, "Online Abuse 101"

World Wide Web Foundation, 2022, "Women's rights online score cards - global summary Country score cards"

World Wide Web Foundation, 2022, "Women's rights online score cards - Peru"

World Wide Web Foundation, 2021, " "Women shouldn't be expected to pay this cost just to participate" - Online Gender-Based Violence and Abuse: Consultation Briefing"

World Wide Web Foundation, Glitch & Social Finance, 2022, "Strengthening Accountability for Online Gender-Based Violence - one year later"

## Y

Yoti, 2023, "Muzz introduces selfie and identity verification from Yoti for safer Halal dating"

Yoti, 2023, "Yoti MyFace liveness white paper"

Youtube Help, 2019, " YouTube test features and experiments"

Youtube, "Child Safety Policy"

Youtube Kids, "About Youtube Kids"

# Acknowledgements

DSIT and PUBLIC are grateful to the following people and organisations for engaging with us during the development of this report, as part of our workshops, user research and best practice sharing. The findings in this report do not represent the views of any individual stakeholders.

We would especially like to thank the **Project Advisory Board** for their meaningful feedback on our work and demonstrated commitment to supporting victim-survivors of online VAWG.

The Project Delivery Team is especially grateful to:

- Chayn
- Glitch
- Refuge
- Suzy Lamplugh Trust
- World Wide Web Foundation
- Professor Clare McGlynn, Durham University

We are grateful to the following people from academia, civil society, private sector and government who demonstrated their commitment to tackling the issues around online VAWG and gave us their time to conduct user research interviews and provide meaningful feedback on our work.

## Project Contributors

#NotYourPorn, Against Violence & Abuse (AVA), Cybertrauma Online Harm Clinical Researcher and Psychotherapist, Empowering Children Foundation, European Women's Lobby, Demos, Garbo, Global Network on Extremism & Technology , Imkaan, Internet Watch Foundation, Muzz, Pinterest, REDDI, Reddit, Sexual Violence Research Initiative, SoSyncd, Stanford Internet Observatory, The Equality Institute, Two Trust & Safety Policy Advisors, Unconform, and WebPurify

This report was authored by an all-women research team at PUBLIC.

PUBLIC

Department for
Science, Innovation
& Technology

**Website:** public.io

**X:** @PUBLIC_Team

**Email:** contact@public.io