Written evidence submitted by Glitch (OSB0097)

About Glitch

Glitch is a UK charity (no. 1187714) that exists to end online abuse and to increase digital citizenship across all online users. We believe that our online community is as real as our offline one, and that everyone should work together to make it a better place. We work to promote good digital citizenship and address online harms such as online abuse, online hate speech and information disorders, and have developed bespoke training programmes covering Digital Citizenship, Online Active Bystanders and Digital Self Care and Self Defence. As part of this, we have delivered training to women in public life.

We are submitting evidence to the Joint pre-legislative scrutiny Committee on the Draft Online Safety Bill's inquiry because we believe that the Online Safety Bill has the potential to make a significant difference to the prevalence of online abuse experienced by internet users in the UK. However, for it to appropriately serve those disproportionately affected by online abuse – women, and especially Black women, and racialised and minoritised people – Glitch believes that the Online Safety Bill needs to be strengthened, with their lived experiences at the core of this legislation.

Summary

Q1 Will the proposed legislation effectively deliver the policy aim of making the UK the safest place to be online?

No. As women are not named in the primary legislation we cannot see how it will appropriately address the disproportionate impact of online abuse faced by women - both as individuals and as a group. We do not believe the Online Safety Bill (the Bill) is appropriately intersectional to protect those most affected by online abuse.. The Bill currently fails to proportionately hold tech companies accountable for the online gender based violence and racial abuse that is prevalent on their platforms. The Bill also fails to make adequate provisions for effective educational interventions that could help prevent high levels of online abuse now and in the future. We outline below our priorities in following areas:

i. Disproportionate Impact on Women and Girls
ii. Intersectionality
iii. Tech accountability
iv. Effective Education Interventions

Q2 Does the draft Bill make adequate provisions for people who are more likely to experience harm online or who may be more vulnerable to exploitation?

No. Abuse of women online is endemic and we do not believe the Online Safety Bill can deliver meaningful change around gendered online abuse without explicitly naming this abuse as a form of violence against women. There is no mention of the disproportionate impact online abuse has on Black women and other marginalised and racialised communities, nor are the specific harms that women face disproportionately online currently named. We call for the inclusion of collective harms.

Q3 How does the draft Bill differ to online safety legislation in other countries (e.g. Australia, Canada, Germany, Ireland, and the EU Digital Services Act) and what lessons can be learnt?
As well as paying close attention to the Digital Services Act, lessons could be learned from the Australian e-Safety Commission, which focuses both on women and diverse groups, recognised to be most at-risk online. As other countries develop new legislation in this area, we will be noting where women are appropriately included. The UK Government should also consider its pre-existing commitments in the area of all forms of gender-based violence and the forthcoming Law Commission recommendations on hate crimes.

Q4 Does the proposed legislation represent a threat to freedom of expression, or are the protections for freedom of expression provided in the draft Bill sufficient?
We do not see freedom of expression and online abuse in opposition with each other. Rather, ending and mitigating online abuse is an integral part of supporting freedom of expression of those who are frequently and disproportionately silenced online. We believe that our online community is as real as offline one and everyone has a right to thrive both online and offline. The current status quo is driving women and particularly marginalised and racialised women and non-binary people to censor themselves online or remove themselves completely through intimidation or experience of online violence against women, which is a violation of their freedom of expression.

Q5 Are the definitions in the draft Bill suitable for service providers to accurately identify and reduce the presence of <u>legal but harmful content</u>, whilst preserving the presence of legitimate content?
No. The approach needs to be holistic, truly intersectional and acknowledge online abuse as a serious social harm and a form of violence against women that contributes towards gender inequality, systemic racism and discrimination of people with intersecting protected characteristics. The vagueness of the Bill does not suggest that this regime will deliver a seismic improvement when it comes to legal but harmful content on category one platforms. We do not agree with the calls to remove legal but harmful from the Bill in place of further criminalisation but implore harmful behaviours that are legal to be appropriately covered by the Bill.

Q6 Are the distinctions between categories of services appropriate, and do they reliably reflect their ability to cause harm?
No. Harmful online behaviours occur both on large (category 1) platforms and smaller platforms. We do not believe the current draft Bill appropriately addresses that. There is not enough emphasis on risk of harm related to platform design and the systems and processes that platforms put in place. Coordinated or collective harassment needs to be addressed as a specific offence.

Q7 What role do algorithms currently play in influencing the presence of certain types of content online and how it is disseminated? What role might they play in reducing the presence of illegal and/or harmful content?
Current social media business models use algorithms that amplify hate and harmful content to fit an attention-based business model. Far more transparency is needed from tech companies around how they intend to use AI that promotes content that violates their terms

and conditions on users' feeds. There also needs to be more transparency around the number and nature of reports received by tech companies of content flagged as violation platforms' terms of service, and why content moderation decisions are made, especially when found to not violate their rules.

Q8 Are there any foreseeable problems that could arise if service providers increased their use of algorithms to fulfil their safety duties? How might the draft Bill address them?
We are concerned about algorithmic bias and over-reliance on technology to detect online abuse. While we support the use of human moderators, who have a fair greater understanding of the nuance of online abuse, people need to be supported holistically in these roles, as moderating abusive and violent content comes at a human cost with high risks of vicarious trauma. We call for improved algorithms and more conscientious thinking around the social impact of technologies.

Q9 Are the media literacy duties given to Ofcom in the draft Bill sufficient?
No. We believe that digital citizenship is key to addressing online abuse and should be made available for people of all ages. This means not only equipping people with digital skills, with better understanding of how to identify misinformation and disinformation, which is important in itself, but also enabling them to become good online users, who can support one another to create an online environment where all can engage free from discrimination. We need a public health approach to ending online abuse which drives awareness around online behaviour while supporting victims of online abuse and violence.

Full evidence submission

Q1 Will the proposed legislation effectively deliver the policy aim of making the UK the safest place to be online?

No, we do not believe that the draft Online Safety Bill will deliver this aim, for several reasons. Firstly, by omitting any references to women in the primary legislation we cannot see how it will appropriately address the disproportionate impact of online abuse faced by over half the population - both as individuals and women as a whole. Secondly, while the Bill makes loose references to people with different characteristics, we do not believe it is appropriately intersectional. Thirdly, in its current form, the Bill fails to hold tech companies accountable for the online gender based violence and racial abuse that is prevalent on their platforms. Finally, it also fails to make adequate provisions for effective educational interventions that could help prevent high levels of online abuse now and in the future.

i. Disproportionate Impact on Women and Girls

Since its formation in 2017, Glitch has been campaigning to make online spaces safe for all, particularly women, girls and people with intersecting identities who are disproportionately affected by online abuse. Globally, women are 27 times more likely to be harassed online than men. A poll conducted by Amnesty International (2017) across eight countries including the UK and USA showed that nearly a quarter (23%) of the women surveyed across the eight countries said they had experienced online abuse or harassment at least once, including 21% of women polled in the UK. A 2018 report by Amnesty International found that in the UK and the US, Black women are 84% more likely to experience online abuse than white women.

The COVID-19 pandemic has exacerbated online abuse targeting women and marginalised communities. In Glitch's most recent report 'The Ripple Effect', we found that almost 1 in 2 (46%) women and non binary people reported experiencing online abuse since the beginning of COVID-19 and 1 in 3 (29%) of those who had experienced online abuse prior to the pandemic reported it being worse during COVID-19. Online abuse not only violates an individual's right to live free from violence and to participate online but also undermines democratic exercises.

Online gender-based violence affects our whole society both online and offline. It is also a huge digital threat to our democracy in the UK and in democracies across the world. Research has revealed that online abuse is one reason many women MPs choose not to run for re-election. More diverse political representation at all levels of politics makes for stronger democracy that should serve all, not just those who are in the same demographic as the vast majority of politicians (older, white, males) who are not systematically driven out of political careers through campaigns of online abuse.

We call for the UK Government to include specific recognition of the disproportionate impact of online abuse on women. Women have a fundamental human right to live free from fear and threat of violence and abuse. Women have fundamental rights to be able to use the internet without having to decide how to keep themselves safe. Women do not just need extra tools on platforms to support themselves or report abuse they have already experienced, and the solution is not for women to come off the internet. The UK Government must accept and acknowledge the high levels of online violence against women and girls in the legislation, not in a subcategory of harms, to be decided later in the legislative process through Ofcom's recommendations.

The latest Girls' Attitude 2021 survey from Girlguiding, published in September 2021 highlights several important issues relating to girls and young women's experiences online. It finds that 71% of girls and young women surveyed between the ages of 7 and 21 had experienced some form of harmful content while online in the last year - this comprises misinformation and hate speech, appearance pressure, harassment and bullying. In the older category of 11-21 year olds, 82% of girls and young women had experienced some form of online harm in the last year: 50% sexist comments; 45% images that made them feel insecure about their appearance; 28% harassment; and 21% bullying. 23% of girls aged 11-16 reported experiencing online harassment, which rose to 33% in young women aged 17-21.When age is disaggregated we see this exposure to harmful content increasing with age: 49% for those between 7 and 10; 73% for 11 to 16 year olds; and 91% of 17-21 year olds.

Exposure to hate speech and hateful content online is unfortunately the norm for many of these young women and girls - 48% 11-16 year old girls have seen such content in the last year. However, this rises to 59% of young women aged 17-21. This increased exposure with age is consistent throughout this report, which suggests that the current approach in the draft Online Safety Bill, which treats children and adults differently - for example in relation to legal but harmful content for adults - will do little to increase the safety of those younger online users once they reach their 18th birthday.

Public opinion strongly supports amending the Bill to require social media platforms to prevent the circulation of content that is harmful to adults as well as children, with 69% of adults in agreement, and just 8% disagreeing.

A much more holistic, intersectional approach is needed to ensure that, if this legislation can deliver a positive experience for children's online, this is not nullified by the exposure to harmful content and hate once that young person reaches adulthood.

ii. Intersectionality

The UK Government has stated that the Online Safety Bill will bring a stop to racist online abuse and tackle misogyny, yet it is unclear how it will achieve that in its current form. We are encouraged that intersectionality is in a sense included in section 21 6b (through "combined characteristics") of the Bill, though without a gendered lens, this does not go far enough and we are yet to understand how the UK Government intends to implement it in practice.

While we welcome the inclusion of "combined characteristics" in the Bill by way of addressing the intersectional nature of the disproportionate way online abuse affects people with different protected characteristics, we would like to see this further developed to reduce ambiguity as to how this will be impactful to those who are experiencing online abuse.

We note the use of "combined characteristics" rather than "protected characteristics", as is used in the Equality Act. We ask whether there is deliberate misalignment between existing equalities legislation and the Online Safety Bill. Online abuse is an equalities issue and should be treated as such through alignment with pre-existing legislation and commitments to tackling inequality. We also urge that the Online Safety Bill be aligned with the Home Office's Tackling Violence Against Women strategy, the Domestica Abuse strategy and that online harms against women are seen for the online gender-based violence that they are. We recommend that the Bill incorporates "combined protected characteristics" in alignment with existing equalities commitments.

Online abuse does not occur in a vacuum and this work to increase online safety would be strengthened by appropriate alignment with existing legislation and policy. Violence against women is a cause and consequence of gender inequality - online gender-based violence is no different and can be just one of a plethora of tactics used to discriminate against women as a group, as well as individuals. It is imperative that the definition of harms for both women and children includes online gender based violence.

Further discussions on the Online Safety Bill and violence against women can be found in the Centenary Action Group's submission to this enquiry, which Glitch helped to coordinate, and that of the Violence Against Women specialist sector, coordinated and submitted by the End Violence Against Women (EVAW) coalition. More from the specialist VAWG sector can also be found in this briefing and web piece. We reiterate a primary call of both of these joint submissions that the Online Safety Bill must recognise and explicitly name online VAWG in all its forms.

We are concerned that without a definition that includes an intersectional understanding of the gendered and racialised nature of online abuse in the primary legislation of the Bill (which take into account the combined and intersecting identities of those affected), we risk the chance of this Bill failing to make real changes to those who are most frequently affected by online abuse.

In 2020 Glitch and the End Violence Against Women coalition (EVAW) conducted the largest investigation into gender-based and intersectional online abuse during the pandemic in the UK with findings presented in The Ripple Effect: Covid-19 and the Epidemic of Online Abuse. The survey was limited to women and non-binary individuals and received 484 responses.

In addition to speaking about intersectionality, Glitch also refers to the concept of "multiple identities," acknowledging that "women who face multiple and intersecting forms of discrimination offline because of their different identities (i.e. race, ethnicity, religion, sexual orientation, gender identity, disability, etc.) are also likely to be targeted with discrimination that targets their multiple and intersecting identities online.

The findings from the survey show that Black and minorities women and non-binary people were more likely experience online abuse during COVID-19 and more likely to report the abuse being worse during the pandemic. This highlights the need to implement responses that include an intersectional lens.

Respondents were asked what aspect(s) of their identity the abuse they faced was related to.

- Gender was the most often cited reason for online abuse.
- Some 48% of respondents reported suffering from gender-based abuse,
- 21% of respondents reported suffering from abuse related to their gender identity and sexual orientation, followed by 18% for their ethnic background and 10% for their religion and 7% for a disability.
- Black and minoritised women and non-binary people were almost as likely to be abused based on ethnicity as they were to be abused based on gender, with 46% of Black and minoritised respondents of colour reporting abuse based on gender and 43% based on ethnicity.
- Black and minoritised respondents were also more likely to be abused for their religion than white respondents.
- While the sample of nonbinary respondents was too small to draw statistical conclusions, anecdotal evidence shows they overwhelmingly experienced abuse related to gender identity and gender expression (7 out of 11).

While this research is the most ambitious attempt to document online abuse against women and non-binary people in the UK during COVID-19, more research is needed into gender-based and intersectional online abuse, as well as the impact of online abuse on Black and minoritised communities. There is an urgent need for greater financial investment from government, tech companies and employers in digital education programmes and research.

Girlguiding's Girls Attitude Survey 2021 found that disabled girls and young women aged 11-21 were more likely to be harassed online (40%) compared to non-disabled girls and young women (25%). For Lesbian, Gay, Bisexual and Queer and Questioning (LGBQ) girls and young women, the survey recorded increases in experienced harassment (42% of LGBQ girls and young women aged 11-21 compared to 24% straight respondents) which included unwanted messages and receiving threats. This group also experienced an increase in online bullying (29% of LGBQ respondents said they'd been bullied online in the last year compared to 20% of straight girls and young women). There was also a high prevalence of sexist comments or 'jokes' seen - 72% of LGBQ girls and young women aged 11-21 compared to 44% of straight girls and young women.

iii. Tech accountability

Glitch's The Ripple Effect Report found that most of the reported abuse took place on mainstream social media platforms (Twitter, Facebook, Instagram) despite tech companies' commitments to making their platforms safe and addressing gender-based and intersectional abuse.

Technology companies cannot fulfil their duty of care to online users without addressing the disproportionate gendered impact of online abuse. We acknowledge that whilst increased accountability for technological companies - including annual transparency reports - is a positive step, there are limitations for how we can truly make the online space safe for all without an increase in digital citizenship education.

There is a lack of trust amongst young people and children around reporting online abuse to social media platforms, with Childnet's Project deSHAME finding that the main reason for children to not report abuse on social media was because they did not think it would help.

The collaborative report Free Speech For All report, which Glitch partnered on, found that there are high levels of distrust amongst adults too regarding the ability of social media platforms to deal with the problem of hateful and harmful content on their platforms, suggesting there is a public desire for legislative regulation. 74% of respondents agreed with the statement 'I do not trust social media companies to decide what is extremist content or disinformation on their platforms' and 71% agreed that 'social media companies should be held legally responsible for the content on their sites.'

Further Public Opinion on the Online Safety Bill research from Compassion in Politics, Clean Up the Internet, Fair Vote and Glitch found that 69% agree the Bill should require social media platforms to prevent the circulation of content that is harmful to adults as well as to children. Just 8% oppose extending the requirement to prevent the circulation of harmful content to include content harmful to adults. 62% of respondents believe social media companies should be preventing the spread of hate on their platforms. It also found that 57% of people want action to stop the posting of personal insults and high numbers of people think social media companies need to prevent the spread of content that is homophobic (56%), xenophobic (54%), ableist 50%, sexist 50%, transphobic 49%, and ageist (42%).

We do not believe that there is not enough emphasis on risk of harm related to platform design and the systems and processes tech companies put in place. We believe that the current suggestion of a risk assessment-based approach for category one companies and no provisions for those that are not category one when it comes to legal but harmful content is flawed. A large proportion of online abuse falls into this category of legal but harmful, and its exclusion from non-category one companies requirements is very disappointing. As such, we do not believe that the current approach will bring meaningful change with regards to the disproportionate impact of online abuse on women and girls, as well as those from racialised and marginalised communities.

We believe that big technology platforms (category one) can afford to comply with the regulations in a way that may allow them to be compliant with the set regulations while allowing them to not do enough to prevent harm to users. This system does not challenge the current business model of tech giants, which prioritises the attention of users at any cost - i.e. interactions based on targeted racism, sexism, ablism, transphobia, homophobia, anti-Semitism, islamophobia etc are as beneficial to the company as the sharing of a harmless comment. Recent reporting of Facebook's leaked internal research about the harm caused by Instagram (for example, which stated in no uncertain terms 'We make body-image issues worse for one in three teenage girls' has come under criticism for evidencing that the company puts profit before harm, even in the case of harm caused to children and young people.

Existing reporting and moderation mechanisms on platforms can add to the emotional and psychological burden when it comes to reporting abuse, where users are asked to report each individual piece of abuse to the platform, which then reportedly frequently respond in a less-than-timely manner by stating why such content does not violate their policies. Even when content has been judged to violate platform's policies, the content has often been left on the site while such an assessment is made, meaning that much of the intended damage has been done.

Despite platforms' growing investment in content moderation, we have to recognise that moderation policies are not achieving good enough results and are not properly enforced.

By comparison, we have seen what is possible on platforms that have made commitments to appropriately address Covid-19 misinformation, where tech companies have acted with urgency due to the seriousness of the public health threat relating to the pandemic. Online abuse also has huge public and individual health implications - and can be a matter of life and death - therefore it should be treated with similar urgency, resourcing and intervention by tech companies.

We also have concerns that the self-regulation of platforms is hugely undermined when for example it is reportedly the case that Facebook does not apply its own rules consistently to all users, but rather grants freedoms and extra privileges to an elite group of users including influential public figures, politicians and journalists, which of whom have been found to be abusing these exemptions of accountability.

As online abuse continues to thrive on social media:

- Platforms should ensure their policies are properly enforced and constantly reviewed to reflect changes in language and take into account mechanisms that allow abusers to bypass their detection mechanisms
- Platforms should make their content moderation policies as clear and understandable as possible
- Platforms should improve their reporting mechanisms. In particular, platforms should acknowledge all reports of inappropriate behaviour and notify the user of the steps being taken to address the issue. They should review those reports within 24 hours, send a warning to the flagged users, and then, if the problematic user persists, remove them from the platform
- Existing features and optional tools for personalising user experience on the platforms should be made available and more obvious to all users, rather than relying on platforms signposting to their own tools in news articles once perpetrators of online abuse have already targeted victims (as was the case following the England Football European Championship 2020 campaign) or charities such as Glitch raising awareness of pre-existing features to small cohorts of online users at a time

While policies are in place, they are still not always properly enforced and reporting mechanisms are complicated to navigate for victims of online abuse. Beyond content moderation policies, social media companies' content moderation processes need to be changed to provide greater transparency, including:

- Algorithmic transparency: the independent regulator - Ofcom - should be able to audit tech platforms' content moderation algorithms
- Transparency about the number and nature of reports received and why content moderation decisions are made

The design of social media platforms has allowed harmful behaviours to thrive, by allowing content to go viral, unchecked. Platforms' business models are also closely linked to the attention economy, with recommendation algorithms presenting social media users with ever more extreme and sensationalised content to capture our attention. We need to recognise that platforms' business priorities cannot take precedence over the online safety of users. The Online Safety Bill has set out a 'duty of care' for platforms towards their users. Platforms therefore need to change their processes to ensure they do not fuel online abuse - for example by reducing virality mechanisms or making sure repeat offenders who have been banned from platforms cannot create new accounts. We also want to see the enforcement powers of Ofcom strengthened, to ensure that the regulator can meaningfully hold companies to account.

iv. Effective Education Interventions

We also advocate for digital citizenship education, aiming to change behaviour of individual internet users by upskilling them to understand that the online space is as real as the offline space. We do not wish to create a pipeline for perpetrators of online abuse to be given custodial sentences, particularly as we acknowledge that good online behaviour was never taught, and therefore the educational piece to encourage better behaviour online has not been widely delivered in the UK. We also support systemic changes to the way that platforms run their services, which currently reward, rather than punish online abuse.

While financial sanctions against technological companies and the enforcement of the duty of care are essential in addressing online harms, more investment in impactful digital citizenship is needed to make the Internet a safer space. We believe that much more is needed in this Bill by way of prevention: through the promotion of digital citizenship education and building on the newly published Media Literacy Strategy.

The 2021 Girlguiding survey suggests that 74% of 7-16 year olds surveyed across the UK had been taught about online safety during the pandemic, which is a positive step. Worryingly, 18% of respondents said they had not. The newly released data from the APPG on Social Media - Selfie Generation - reports that when assessing 15,000 schools in the UK, SWGfL found that in over 40% of cases, the schools do not have any professional development for staff about online safety. The APPG report also found that 52% of young people are too embarrassed to disclose online abuse.

A major tension point is that as a global society, we have not drawn up the rules or social norms online for the line between political accountability and online abuse. While some offline perpetrators of abuse towards public figures cross the line of what is widely considered socially acceptable, the majority of people would not perpetuate this level of violence towards public figures in the offline world, partly because good bystanders would likely intervene, in addition to the in-built security provisions.

Social norms of this kind are less clear in the online space. In the UK's education systems, there is a deficit in political education, and curricula that are in dire need of being decolonised. We must teach young people about racism, sexism and other forms of systemic oppression, or we cannot expect good online digital citizens.

As the UK Government looks to introduce new laws to make the UK the safest place to be online, we are urging the Chancellor of Exchequer to ring fence 10% of the new digital services tax to help achieve this. The Digital Services Tax of 2% on tech giants like Facebook, Google and Twitter is expected to raise an additional £400m a year (£70m (2019/20); £280m (2020/21); £390m (2021/22); £425m (2022/23); £465m (2023/24); £515m (2024/25).

It is essential that the UK Government provides funding to the specialist violence against women sector and online abuse organisations to support victims of online abuse and helps fund the vital work of ending online violence and abuse, such as through training on good digital citizenship and online safety, providing resources and awareness raising and supporting survivors of online abuse and violence. By ring fencing at least 10% of this new tax annually for ending online abuse, the UK Government can commit £4m+ to further establishing online standards which are fair and necessary to the growing digital economy, funded by the tech giants where these societal harms are pervasive. Through no negative deficit, using money from tech giants, the UK Government can take decisive action.

Q2 Does the draft Bill make adequate provisions for people who are more likely to experience harm online or who may be more vulnerable to exploitation?

No. Abuse of women online is endemic and the Online Safety Bill must recognise this both in the rhetoric around the legislation, as we have seen from DCMS ministerial announcements, and explicitly within the legislation itself. It is important that women are explicitly named in the primary legislation and our inclusion not left to chance in the secondary legislation process, relying on the assumption that women's inclusion will be recommended by the regulator Ofcom, and subsequently agreed and passed by parliament. we do not believe the Online Safety Bill can deliver meaningful change around gendered online abuse without explicitly naming this abuse as a form of violence against women.

We know that those who are disproportionately affected by online harms are women, with women of colour and especially Black women exponentially impacted, as too are other marginalised women and non binary people, yet the Online Safety Bill makes no mention of this, nor the types of harm that women and minoritised groups both witness and are subjected to, for example death threats and rape threats, hate speech and the sharing of non-consensual photographs. Not only is the prevalence of online violence against women a violation of human rights, and an attack on both democracy and gender equality, there is also a financial and economic cost to both the individual woman through reputational damage, and to society, which suffers socially as well as financially from all forms of racial and gender inequality.

We do not support the proposal in the draft bill, which focuses solely on harms to individuals and places no emphasis on collective harms to certain groups. Online abuse against women is a form of violence against women, perpetrated in high volumes towards women not just as individuals but specifically because they are women. The high prevalence of racist abuse is perpetrated for example towards a Black woman MP because she is Black and a woman. The issue is systemic and should not be seen outside of the pervasive power structures of white supremacy and patriarchal attitudes of male-superiority over women. We call for the inclusion of collective harms, which are damaging to our society, to be reintroduced into the draft bill, having been removed from the Online Harms White Paper.

We do not see how this draft of the Bill will adequately address the disproportionate impact of online abuse on women, on racialised people and on marginalised communities.

Q3 How does the draft Bill differ to online safety legislation in other countries (e.g. Australia, Canada, Germany, Ireland, and the EU Digital Services Act) and what lessons can be learnt?

While several countries have passed legalisation in this area already, some legislative developments abandon a gender-neutral stance by acknowledging that women are disproportionately affected by online abuse and that online abuse is part of a continuum of violence against women that it is in the best interest of the State to mitigate, minimise and work towards ending. For example, the e-Safety Commission in Australia is the world's first government agency committed to keeping its citizens safer online, and has a programme - Women in the Spotlight - to specifically elevate and protect women's voices online. The Commission is also becoming increasingly intersectional in its approach, with 'diverse groups' recognised as part of at-risk groups online, i.e. Aboriginal and Torres Strait Islander people; culturally and linguistically diverse people; people living with disabilities; lesbian, gay, bi, trans, intersex and queer people; as well as women, older Australians, children and young people.

The European Parliament and the Council currently have a unique opportunity to ensure that the Digital Services Act is strengthened to better tackle gender-based violence against women. Glitch is working with partners in Europe such as AWO and the European Women's Lobby to make this case. We support the call for the DSA to be strengthened regarding the need for clear obligations to be imposed on platforms to particularly identify, prevent and mitigate the risk of gender-based violence not only being perpetrated on their platforms but also amplified by their product. Through the DSA as well as the Online Safety Bill, legislation should ensure that platforms must take into account the ways in which design choices and operational approaches can both influence and increase risks of gender-based online violence and abuse.

We believe that a gender-responsive approach to both the Digital Services Act and the Online Safety Bill is key in ending gender based violence online that is so prevalent across Europe (74% of women reporting experiencing some form of online violence in the EU in 2020) and the world.

There are also existing frameworks and conventions the UK Government has signed up to and could utilise to strengthen this policy area, as well as aligning this legislation to the Equality Act. For example, now that the Domestic Abuse Act has passed, we urge the UK Government to ratify the Istanbul Convention, as well as incorporating the Convention on the Elimination of All Forms of Discrimination against Women into domestic law. While the UK Government has signed up to both of these important frameworks, the lack of implementation to date has meant that less progress has been made in terms of ending violence against women and discrimination against women than if these were fully ratified and fully incorporated into UK law.

We also believe that the recommendations of the Law Commission with regards to hate crime and online violence against women and marginalised people should be included in the Online Safety Bill to ensure these policy recommendations are legislated as soon as possible, with advice from experts in each specific field.

Q4 Does the proposed legislation represent a threat to freedom of expression, or are the protections for freedom of expression provided in the draft Bill sufficient?

In our view, the current narrative around freedom of expression online has created a false trade-off between 'free speech' and online violence. Online abuse disproportionately affects women and in particular Black women and other racialised women. Since we started our work at Glitch in 2017, we have documented the scale of abuse and online violence targeting women and girls in the UK, as well as marginalised communities and have seen how this abuse undermines free speech by attempting to silence marginalised communities and women and persons of colour who are involved in public life.

Freedom of expression is not the freedom to abuse, or commit hate speech online or offline, and we should be careful about framing 'freedom of expression' and 'harassment' in opposition to one another. Without safeguards against harassment or hate speech, freedom of expression is undermined, and diverse political representation is stifled. We do recognise that there are difficult legal questions to answer in relation to, for example, what constitutes 'gross offensiveness' online but these questions should not distract from the problem at hand: the sheer scale of online abuse targeting women and girls, and marginalised communities in the UK and across the world.

While some argue that increased regulation threatens freedom of expression, there is often little attention given to the status quo, where communities and groups of people are disproportionately affected by 0nline abuse and have their own rights to freedom of expression impinged, through tactics that deliberately threaten, silence and drive users from particular demographics out of the online 'public square' altogether. We believe that our online community is as real as offline one and everyone has a right to thrive both online and offline.

Despite the vast majority (96%) of participants of Glitch's workshops stating that post-workshop, they feel that they now have the skills to be safer and more resilient online, 69% of participants have told Glitch that they will continue to censor themselves online due to anxiety or fear of how others will respond.

Amnesty International has deemed Twitter's response to violence and abuse against women to be inadequate, leading to women self-censoring when posting, interacting with online content or leaving platforms altogether - their survey found that 31% of women in the UK who had experienced abuse or harassment online had stopped posting content that expressed their opinions on certain issues since the abuse. They also highlight the serious and harmful repercussions this can have on young women and women from marginalised communities, as well as future generations, who feel unable to exercise their rights to participate in public life and express themselves freely online.

The perceived and/or actual threat of violence towards women and particularly Black women, racialised women and minoritised women both online and offline is a behaviour of oppression that is part of and reinforces the systems of white superiority and patriarchy. It is no surprise then that women, who are disproportionately affected by online abuse, have their rights to freedom of expression compromised deliberately within the same power structures that we are subjected to offline.

Public opinion research found that 60% of survey respondents believe freedom from abuse is more important than freedom of speech - this was supported by 54% of men compared with 66% of women and 57% of Conservatives and 68% Labour supporters.

Freedom of expression has become a smoke screen for perpetrators of online violence. We do not see freedom of expression and online abuse in opposition with each other. Rather, ending and mitigating online abuse is an integral part of supporting freedom of expression of those who are frequently and disproportionately silenced online.

Q5 Are the definitions in the draft Bill suitable for service providers to accurately identify and reduce the presence of legal but harmful content, whilst preserving the presence of legitimate content?

No, we do not believe so. A large proportion of online abuse against adults falls into this legal but harmful content. Platform's self-regulation to-date has not provided safe spaces for women, Black communities and minorities. Anecdotally, we often hear that content that does violate service providers' own policies is reported by those targeted with abuse and then deemed to be not harmful. We believe much more needs to be done to shift the identification of legal but harmful content away from the current status quo.

The approach needs to be holistic, truly intersectional and acknowledge online abuse as a serious social harm and a form of violence against women that contributes towards gender inequality, systemic racism and discrimination of people with intersecting protected characteristics. The vagueness of the Bill does not suggest that this regime will deliver a seismic improvement when it comes to legal but harmful content on category one platforms.

There are campaigners who argue that the 'legal but harmful' provision should be removed from the Bill and that, for example, the racist abuse experienced by England's footballers after the Euro final that is not currently illegal should be criminalised. We do not support this belief. When talking about gender-based online abuse, the vast majority of online abuse against women falls into the 'legal but harmful' category. We believe that the removal of the 'legal but harmful' regulations in the Bill would weaken the legislation with regards to ending online violence against women. We do not support the calls to make more legal but harmful behaviour illegal as we do not wish to create a pipeline for perpetrators of online abuse to be given custodial sentences - this does not support attitudinal changes as effectively as we believe good digital citizenship and a public health approach to online abuse could.

Furthermore, the report Free Speech For All: Why legal but harmful content should continue to be included in the Online Safety Bill, which Glitch worked on in partnership with Hope Not Hate, Reset, Antisemitism Policy Trust, Impress, Clean Up the Internet, DEMOS and Catch 22, investigated current public attitudes towards harmful content on social media. It found that there is considerable public concern for extremist and hateful content on social media, with 73% of respondents agreeing, while 80% agreed with the statement 'I believe in free speech, but there must be limits to stop extremist content and hate speech on social media.' 73% respondents agreed that social media companies should be made to remove harmful and hateful content being displayed on their platforms

The public opinion research from Compassion in Politics, Glitch, Free Vote and Clean Up the Internet found that 65% of people believe sharing intimate images without someone's approval constitutes harmful behaviour, as does intimidating or threatening someone. 63% agree that the spreading of false information about individuals is harmful. 58% of people say that being transphobic, ageist, ableist, racist and/or homophobic is harmful.

Q6 Are the distinctions between categories of services appropriate, and do they reliably reflect their ability to cause harm?

No, we do not believe that there is not enough emphasis on risk of harm related to platform design and the systems and processes they put in place. Online abuse is prevalent on many online platforms and consists of both legal but harmful, and illegal content, yet the Bill proposes additional duties to just the larger companies like Facebook and Twitter. Some of the "alternative" or less "mainstream" platforms that are unlikely to be classified as category 1 under the current draft (for example Gab, Telegram, BitChute and 4chan) have been highlighted in a recent report as “relatively unregulated online spaces for extremists to share content that celebrates and encourages hatred and murder.”

Coordinated collective harassment, which can take place on these smaller platforms or forums, is a growing problem online, and affects women and minoritised communities disproportionately. Research has shown that malign groups and individuals are using social media to launch coordinated harassment campaigns against public figures and social media users with no public profile alike. An investigation by BBC Newsnight in April 2019 showed that female politicians across Europe were targeted with threatening and misogynist content and coordinated harassment campaigns ahead of the European Parliamentary election.[1] Research has consistently shown that women are more likely to be targeted by coordinated or uncoordinated harassment. In 2017, the European Institute for Gender Equality described cyber harassment against women and girls as a growing problem. Legislation in the UK has not kept up with this growing threat. Coordinated or collective harassment needs to be addressed as a specific offence.

Q7 What role do algorithms currently play in influencing the presence of certain types of content online and how it is disseminated? What role might they play in reducing the presence of illegal and/or harmful content?

We support systemic changes to the way that platforms run their services, which currently reward, rather than punish online abuse due to algorithms amplifying hate in an attention-based business model, rather than one that actively does not promote hate. A better, more effective approach would be to enact effective ways to reduce online abuse in the first place, by investing in good digital citizenship education, as well as changing the way that social media platforms’ current business model champions the volume of interactions, whether these are of positive, negative or neutral impact to the user, and thus profits from online abuse as generated content, data and attention given to the platform.

Social media platforms, for example Twitter, have announced that they will do more to address the way in which AI promotes content on users’ feeds that is in violation of their own platforms. While this is a positive step, more transparency around this is needed. In the past year, the world has witnessed how easy it is for misinformation relating to the global pandemic and in particular the Covid-19 vaccine can spread. We have seen the efforts that tech companies have taken to address this as a public health concern calling grave societal harm. We are concerned that the same emphasis is not being placed on ending online gender-based violence and online racism, with is itself life and death

Beyond content moderation policies, social media companies’ content moderation processes need to be changed to provide greater transparency, including:

- Algorithmic transparency: the independent regulator - Ofcom - should be able to audit tech platforms’ content moderation algorithms
- Transparency about the number and nature of reports received and why content moderation decisions are made, especially when found to not violate their rules

[1] A web of abuse: How the far right disproportionately targets female politicians - BBC News

The design of social media platforms has allowed harmful behaviours to thrive, by allowing content to go viral unchecked. Platforms' business models are also closely linked to the attention economy, with recommendation algorithms presenting social media users with ever more extreme and sensationalised content to capture our attention. We need to recognise that platforms' business priorities cannot take precedence over the online safety of users. The Online Safety Bill has set out a 'duty of care' for platforms towards their users. Platforms therefore need to change their processes to ensure they do not fuel online abuse - for example by reducing virality mechanisms or making sure repeat offenders who have been banned from platforms cannot create new accounts.

Q8 Are there any foreseeable problems that could arise if service providers increased their use of algorithms to fulfil their safety duties? How might the draft Bill address them?

We believe that over-reliance on AI technology to identify toxic and harmful content has limitations - for example, active bystander interventions or discussions about issues such as racism from good digital citizens can be flagged as offensive by AI software which blocks content based on words used rather than sentiment. Online abuse can be personally targeted in a way that would be too nuanced for AI systems to highlight, particularly by a perpetrator known to the victim. Perpetrators of online abuse adapt strategies to bypass AI systems, for example, by using different or special characters or spaces in offensive words to get past detection.

In addition, while platforms are increasingly reliant on algorithms, there are also many examples of issues of platforms applying their own policies appropriately. When AI systems deem reported content to not be in violation of the platform's terms of service, there is additional need for additional reporting on the part of the victim of abuse and human moderator intervention that is far more nuanced than current algorithmic technology is able to be.

While we support the use of human moderation, the role of a content moderator takes an incredibly heavy toll on the person's psychological and mental wellbeing, as they are tasked with reviewing potentially harmful and upsetting content for hours at a time, therefore the risk of vicarious trauma is high. While moderators themselves are not the primary target of the abusive content they are reviewing, they may be from the protected class being targeted or be psychologically triggered in other ways. Such work tasks must be supported in a holistic manner, with good psychological support for the moderator's wellbeing. Human moderation is far more advanced and nuanced than AI, but it comes at a human cost that we cannot underestimate.

We have concerns around the dataset used and the inherent biases that are prevalent in algorithms and AI due to biases held by their creators, which has been evidenced through the work of MIT graduate Joy Buolamwini and the Algorithmic Justice League, amongst other researchers. We support Joy Buolamwini's calls for training sets that create algorithmic learning that need to be more inclusive; existing algorithms that should be audited to root out existing biases and more conscientious thinking around the social impact of technology that exists and is being developed.

Q9 Are the media literacy duties given to Ofcom in the draft Bill sufficient?

Digital citizenship is key in addressing online abuse and should be made available to young people and adults alike in settings such as workplaces, industry organisations and trade unions. All individuals have a right to engage in all online spaces safely and freely without discrimination. Digital Citizenship is respecting and championing the human rights of all individuals online, and encompasses three key elements: individual, social and institutional responsibilities. Glitch believes that digital citizenship is an essential solution to ending all forms of online abuse. Our approach and perspective on digital citizenship is built on definitions from the Council of Europe and Australian Curriculum. *For further introductory reading on digital citizenship from Glitch, please see Appendix 1*.

We need a public health approach that addresses the root cause of bad online behaviour, such as trolling, online harassment and stalking, and all other types of online harm and abuse. A public health approach should also support victims of online abuse as well as those disproportionately impacted.

We also advocate for digital citizenship education, aiming to change behaviour of individual internet users by upskilling them to understand that the online space is as real as the offline space. We need to equip all online users to become good digital citizens, both those in formal education and those who are not to ensure that all can navigate the online space freely and safely. We do not just want online spaces free from direct abuse and harmful content. We want spaces where all users can thrive, not just exist, and survive.

Appendix 1

Introduction to Digital Citizenship

This document is a brief introduction to digital citizenship. It explores definitions of digital citizenship, the success of Glitch's digital citizenship education efforts so far and the challenges to achieving good digital citizenship education.

How Glitch defines Digital Citizenship

At Glitch, we define digital citizenship as the following:

All individuals have a right to safely and freely engage in all online spaces without discrimination. Digital Citizenship is respecting and championing the human rights of all individuals online, and encompasses three key elements: individual, social and institutional responsibilities.

We all have a responsibility as individuals to improve our digital literacy and digital safety skills, understanding our digital footprints and engaging in digital self-care.

Our social responsibility as digital citizens includes enacting online active bystander interventions, practising respectful online etiquette and responsible and positive engagement with digital technologies and not misusing them to disadvantage others.

At Glitch, there are 4 main pillars of digital citizenship:

- **Digital Self-Defence:** Using online tools to protect ourselves and others in online spaces
- **Digital Self-Care:** Creating boundaries in digital spaces to look after our wellbeing
- **Online Active Bystander:** What to do when you see someone else experiencing online abuse
- **Tech Accountability:** Understanding how to hold tech companies accountable

Governmental institutions and tech companies also have a role to play in digital citizenship, ensuring that individuals can exercise their online rights whilst protecting the rights of those with multiple and intersecting identities.

Governments must also prioritise digital citizenship education for all, using a public health approach that examines the wider impact of online abuse in a community, rather than treating online abuse as single incidents.

Governmental institutions also have a responsibility to ensure that tech companies prioritise digital citizenship and outline clear roles and responsibilities, accountability, regulation and investment in education and resources.

Tech companies need to create technology and online platforms that are safe and non-discriminatory for all users. This includes designing online spaces, systems, rules and tools that encourage digital citizenship from their users. They should also consult their users, act transparently, and implement robust safety mechanisms.

At Glitch we advocate for a public health approach via digital citizenship education. We want governments and tech companies to understand how online abuse impacts communities and how incidents create a ripple effect that impacts friends, classmates, communities, families etc. and to create educational content from this. The role of the whole community should be considered to create a robust structure to deal with online abuse. This approach works with key institutions with responsibilities such as governments, tech companies, schools, and youth organisations to create safe online spaces for young people. Adult voices need to reinforce the messages given to young people regarding online abuse and not take a perpetrator/victim approach when responding to online harms. Online abuse also impacts mental health with some high profile cases in recent years demonstrating the extent to which online abuse can harm our wellbeing.

What is Digital Citizenship?

Organisations have differing definitions for Digital Citizenship with some listed below. Most definitions cite a need to engage positively with digital technologies, learning how these technologies work, and understanding social rights and responsibilities in online spaces

Council of Europe:

The competent and positive engagement with digital technologies (creating, working, sharing, socializing, investigating, playing, communicating and learning); participating actively and responsibly (values, skills, attitudes, knowledge) in communities (local, national, global) at all levels (political, economic, social, cultural and intercultural); being involved in a double process of lifelong learning (in formal, informal and non-formal settings) and continuously defending human dignity.

The Australian Curriculum:

An acceptance and upholding of the norms of appropriate, responsible behaviour with regard to the use of digital technologies. This involves using digital technologies effectively and not misusing them to disadvantage others. Digital citizenship includes appropriate online etiquette, literacy in how digital technologies work and how to use them, an understanding of ethics and related law, knowing how to stay safe online, and advice on related health and safety issues such as predators and the permanence of data.

Diana Graber, Cyberwise:
The curriculum starts with digital citizenship, which is the safe and responsible use of digital tools, moves into information literacy — knowing how to find, retrieve, analyze, and use online information — and then finishes with media literacy, which is learning to use critical thinking skills to analyze media messages.

Success of Digital Citizenship education

Glitch has been delivering Digital Citizenship education since 2017. These workshops focused on the following aims:

- Digital rights and digital responsibilities
- Online bullying, prejudice-based bullying and its impact
- Democracy and law as they pertain to online spaces and interaction
- How to be responsible digital citizens within society
- Digital health, wellbeing and critical thinking

In 2019 we refined our monitoring and evaluation framework to closely align to behavioural change. We focused on delivering workshops to Year 7 students in one diverse London borough to measure the impact of our intervention.

When surveyed after the workshop:

- 75% more students knew what digital citizenship was and how to foster positive online engagement following our workshop.
- 86% of students surveyed in School 1 said they would behave differently online as a result of the information they learned during our workshop.
- There was a 40% increase in young people's willingness to take positive action if they or someone they know experienced online harms.
- There was a 49% increase in students' knowing how and why to flag social media content.

Glitch's workshops

As well as our digital citizenship workshop for schools, we offer the following workshops:

Digital Self Defence and Self Care:

Our most popular workshop covers 6 top Digital Self Defence strategies and 3 Digital Self Care Methods. By the end of this two-hour session, you will gain the skills to be safe when using online spaces, be able to use different digital tools to protect your information and privacy, understand how to set your boundaries in online spaces and begin building your digital resilience action plan.

Hosting Conversations with our Toolkit 2.0:

Our Toolkit 2.0 was created to help end Online Gender-Based Violence (OGBV) for Black Women and to help allies provide better support online. This 90 min workshop is aimed to help you feel prepared to use the Toolkit with your community or network and looks at how Online Gender-Based Violence is an intersectional issue that disproportionately impacts Black Women and gives you the skills and knowledge to host your own conversations about OGBV to help end it.

Challenges to Digital Citizenship

- **Funding**: There is a lack of funding for research and development of digital citizenship education programmes.
- **Time with young people:** One workshop is only the beginning of digital citizenship education, a more comprehensive and long-term programme is needed.
- **Young people not in formal education:** It is difficult to reach young people who are not in formal education
- **Adults:** Workshops with young people is only one part of digital citizenship education. The adults in these young people's lives need to understand digital citizenship and work together to create positive digital communities.
- **Longitudinal studies**: To understand the long-term impact of digital citizenship education, studies need to be conducted with participants 2-3 years after undertaking digital citizenship education.
- **Intersectionality**: Online harms impact women and marginalised communities to a greater degree than others. Digital citizenship education and legislation must take into account the disproportionate impact of online abuse on these communities and design solutions to meet their needs. Not every solution is going to work for every community and it is important to build this into any digital citizenship strategy.