



## REPORT HARMFUL CONTENT

# Pilot year evaluation

Prepared by Elena Sharratt

# Contents

04

Table of figures

05

Executive summary

10

Introduction and background to service

What is Report Harmful Content? 10  
What is the purpose of this report? 11  
Who is this report for? 11

12

Top level statistics

Helpline use and growth 12  
Service response 12  
Following initial service response 13  
Nature of cases 13  
Referral routes 15  
How were clients helped? 16  
Client demographics 18

20

Exploring cases in more depth

Emerging trends 20  
Trend one: impersonation + bullying and harassment + privacy violation 20  
Trend two: abuse + threats + hate speech 21  
Trend three: perpetrator-less harms 22  
Emerging concerns and challenges 22  
Mental health 22  
Criminal/legal issues 24  
Cultural and religious issues 24  
Clash of protected characteristics 25  
Relationships with platforms 26

27

Discussion and Recommendations

The value of the service 27  
Responding to emerging trends 27  
Tackling emerging issues 30

32

Conclusion

33

Bibliography

Figure 1: Volume of cases per month 12

Figure 2: Proportion of service response by type 13

Figure 3: Proportion of harms by type 14

Figure 4: Frequency of harms by type per month 14

Figure 5: Proportion of other online issues by type 15

Figure 6: Frequency of associated offline issues and contexts by type 15

Figure 7: Frequency of signposting according to service/agency 16

Figure 8: Frequency of assistance channels 16

Figure 9: Age of clients 18

Figure 10: Gender of clients 18

Figure 11: Location of content 18

Figure 12: Proportions of content location by age group 18

Figure 13: Proportion of content location by gender 19

Figure 14: Proportion of content location by month 19

Figure 15: proportion of hate speech by type 21

Figure 16: Number of trend three cases by month 22

Figure 17: Frequency of mental health impacts by type 23

# Executive summary

Report Harmful Content (RHC) is a national reporting centre provided by the UK Safer Internet Centre and operated by South West Grid for Learning (SWGfL).

RHC has been designed to assist everyone in reporting harmful online content by providing up to date information on social media community standards and direct links to the correct reporting facilities across multiple platforms. The service also takes on a mediatory role for clients who have already submitted a report to industry and would like outcomes reviewed or escalated. RHC have trusted flagger relationships with a number of social media platforms, which enables them to act in this mediatory role. RHC reviews reports associated with eight types of online harm: abuse, bullying and harassment, threats, impersonation, unwanted sexual advances, violent content, self-harm/suicide content and pornographic content. RHC also offers advice on additional issues faced online and signposts to other support services and the police when necessary.

**RHC has been in pilot phase since January 2019.** This report presents results of mixed-methods research carried out on all cases dealt with in the first year of operation (January 2019-December 2019). In the year analysed, the RHC website received 9,282 visitors and practitioners dealt with 164 unique cases.

The service's popularity rapidly increased in September, following the official service launch, and continued to grow until the end of the year.

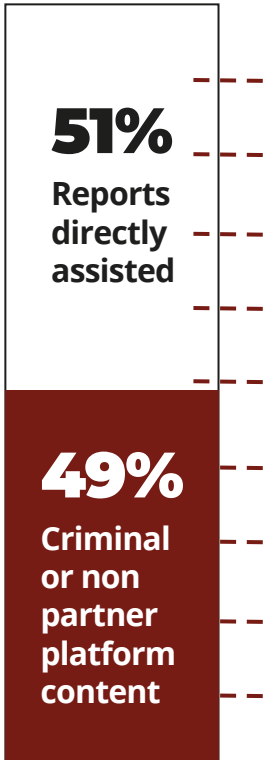
**9282**  
Visitors to the RHC website

**164**  
Unique cases



**The results of this study found that, in the majority of instances (51%), practitioners were able to directly assist clients in reporting harmful content online.** This was done either by directing them to the correct industry reporting links or by escalating content to industry on their behalf.

In the remaining 49% of cases, content was deemed to be either criminal or it was found to be located on platforms with which RHC do not have partnerships. In these instances, practitioners provided advice and onward signposting. Of the content escalated to industry, 92% was successfully actioned (e.g. removed/ restricted/ regained access to) and 62% was done so within 72 hours, demonstrating a high level of service speed and efficiency.



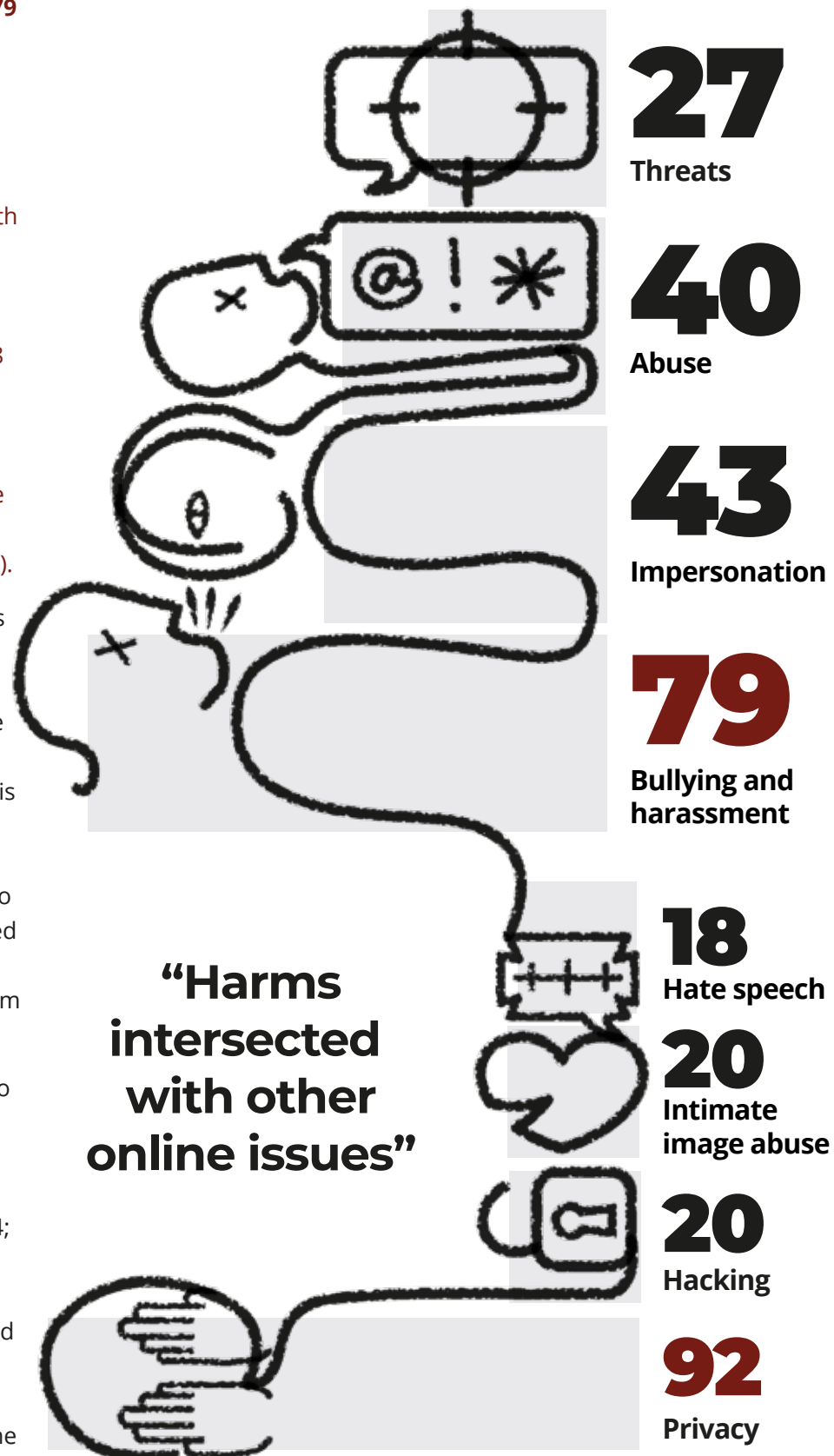
**92%** Escaled content sucessfully actioned

**62%** Removed within 72hrs

**Out of the main eight online harms, cases involving bullying and harassment were most common (79 cases).** This was followed by impersonation (43 cases), abuse (40 cases) and threats (27 cases). It was rare for a client to report just one online harm; harms frequently overlapped. They also intersected with other online issues (most commonly privacy violations (92 cases), hacked accounts (20 cases), intimate image abuse (20 cases) and hate-speech (18 cases)) and with offline concerns (including offline bullying and harassment (25 cases), domestic abuse (17 cases), business/workplace disputes (14 cases), honour based abuse (6 cases) and stalking (2 cases)).

In analysing these overlapping harms and issues in more depth, three common trends were identified. The first involved a cluster of three online issues: impersonation, bullying and harassment and privacy violation. This cluster significantly intersected with offline abuse, in particular domestic violence, and was much more likely to affect women, in particular those aged 31-50. The perpetrator of this abuse was almost always known to the victim (e.g. partner or family member).

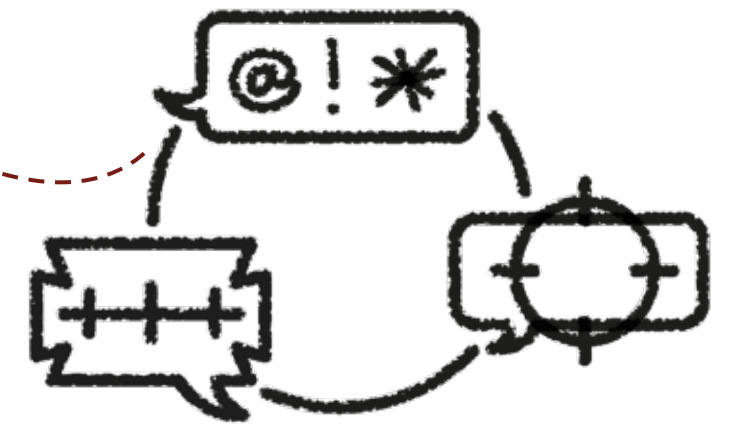
Significantly, this trend can be seen to confirm existing knowledge surrounding the gendered nature of intimate/domestic abuse and harassment (Dobash & Dobash, 2004; ONS, 2019; Women's Aid, 2019), alongside the growing propensity for this type of violence to be perpetrated through technology (Drouin, Ross & Tobin, 2015; Refuge, 2020; Thacker, 2017). These findings thus support the need for a range of tools and a



multi-agency, technology-centred approach to be developed in order to tackle all aspects of gender-based abuse.

The second trend identified also contained a ‘cluster’ of three online issues: abuse, threats and hate-speech. Within this, the most common type of hate-speech reported was racism/xenophobia. Unlike trend one, clients reporting trend two were overwhelmingly bystanders as opposed to victims and perpetrators of this abuse were usually unknown individuals on social media, as opposed to partners or family members. This type of abuse was reported as an isolated event, as opposed to an ongoing pattern, and it did not significantly intersect with offline abuse, harassment and/or domestic violence. Clients reporting this trend were also more evenly distributed across gender and age.

The identification of this trend can be seen to confirm recent research documenting the growth of online hate speech, in particular racism, extremism and far right ideology (Williams & Mishcon de Reya, 2019). These findings thus support the recognition that social media companies need to develop more consistent and rigorous policies for identifying and responding to hateful conduct (Williams & Mishcon de Reya, 2019). They also support recommendations which outline the need for an independent regulator to enforce this duty of care (Department for Digital, Culture, Media and Sport, 2019; Williams & Mishcon de Reya, 2019).



The third trend identified differed from trend one and two in that a perpetrator was not evident. Instead, this trend involved clients inadvertently viewing harmful content rather than being victim of or witnessing targeted, harmful behaviour. The types of harmful content viewed included graphic/violent content, self-harm/suicide content or pornography, often overlapping (e.g. violent pornography, sexualised self-harm). Men were most likely to report this type of content (64%), with men in the 19-30 age group reporting a significant proportion (39%). Reports of this nature began to increase towards the end of the year, peaking in December. The growing increase in these types of reports is encouraging, given that it demonstrates the broadening reach of the RHC service. If, however, reports of this nature continue to grow, it might become necessary for the RHC service to introduce additional tools to deal with this demand, including building trusted flagger relationships with a broader range of platforms.



**Alongside trends, a number of emerging issues and challenges were identified.** One particularly concerning issue arose regarding law enforcement. 19% of RHC clients reported content which was deemed to be criminal and thus referred to law enforcement. Of that 19%, however, 47% got back in touch with RHC, often reporting that the police had dismissed them and incorrectly informed them that their issue was non-criminal. These findings thus support previous recommendations regarding the need for better training of law enforcement on issues of online crime and abuse (Bond & Tyrrell, 2018; Home Office, 2018).

Some issues with industry partners were also identified. At times practitioners experienced inconsistent responses from industry platforms when escalating content and there was often a lack of clarity around what type of content would be removed. On top of this, industry community guidelines were not always comprehensive enough to deal with more complex cases. One particular area of confusion lay in relation to cases involving a clash of characteristics protected under the equality act (Equality Act 2010), in particular gender reassignment and sex.

Alongside issues with service operations, issues regarding online culture more broadly were identified. RHC dealt with a number of clients from particular cultural and religious backgrounds who reported the exposure of private and/or intimate material. This type of content often did not meet the legal threshold for



**19%**  
Reported  
criminal  
content

**Some felt  
dismissed  
by the  
Police**



**INCONSISTENT**  
industry responses

intimate material and was thus was not covered under intimate image abuse laws. Nevertheless, because of the cultural context, clients perceived this material to be just as private and experienced its disclosure as extremely distressing and violating. These types of cases raise questions regarding how online harms are defined and whether those definitions can be universally applied. The findings of this report thus support the recent recognition that more nuanced definitions of private/intimate material need to be developed (Law Commission, 2018).

Finally, one of the most significant issues to be identified was the widespread impact of online harms on mental health; 32% of RHC clients reported negative mental health impacts as a result of viewing or being the victim of harmful content online, with 13% reporting suicidal ideation. These findings clearly support the need for these issues to be understood in their complexity and for mental health online to remain at the top of the agenda for those in positions to address it.

**32%**  
of clients reported  
negative mental  
health impacts



with **13%**  
reporting suicidal  
ideation.

**“Overall, this report  
displays the necessity  
and value of the RHC  
service.”**

Overall, this report displays the necessity and value of the RHC service. Alongside mediating between social media users and industry to facilitate the successful actioning of harmful content, RHC practitioners were able to assist clients by providing them with information and/or clarification on the nature of their issue or redirecting them to reporting links on other (non-industry) sites. The service also offered vital emotional support, alongside signposting to other agencies and services, either for additional emotional support or practical assistance. Qualitative feedback from clients (included in the report) documented the service's speed and efficacy, alongside the quality of support offered by practitioners. This report identifies multiple ways in which the RHC service can be developed so as to respond to the growth and diversification of the types of reports received. All of these areas will enhance the scope and quality of the RHC service and further contribute towards empowering everyone to report harmful content online.

# Introduction and background to service

## What is Report Harmful Content?

**Report Harmful Content (RHC) is a national reporting centre that has been designed to assist everyone in reporting harmful content online.**

RHC is provided by UK Safer Internet Centre and operated by South West Grid for Learning (SWGfL). The service grew out of SWGfL's previous experience running the Professionals Online Safety Helpline and the Revenge Porn Helpline. Whilst these services offer essential support to members of the children's workforce and adults who have had their intimate images shared, respectively, certain elements of online safety provision were identified, with which neither of these helplines could assist.

RHC was designed to fill that gap. It empowers anyone who has come across harmful, but not necessarily criminal, content online to report it by providing up to date information on community standards and direct links to the correct reporting facilities across multiple platforms. The service also provides further support to clients based in the UK, over the age of 13, who have already submitted a report to industry and would like outcomes reviewed. RHC is able to act in this mediatory role with a number of social media platforms, with whom it has a trusted flagger partnership. These platforms include: Facebook, Instagram, Snapchat, Twitter, Roblox, TikTok, Microsoft (which includes LinkedIn, Bing, Xbox, Skype and

Minecraft) and Google (which includes YouTube, YouTube Kids, Google Search and Blogger). All types of support offered by RHC are provided to clients via email contact.

The term 'harmful content' can be very subjective. In order to remove ambiguity, specialist online safety practitioners studied the community guidelines of several different social media platforms. They found that 8 areas of content are likely to violate platform terms: abuse, bullying and harassment, threats, impersonation, unwanted sexual advances, violent content, self-harm/suicide content and pornographic content. RHC practitioners review reports associated with these eight types of online harm. They also offer advice on further issues faced online and signpost to support services and the police when necessary.

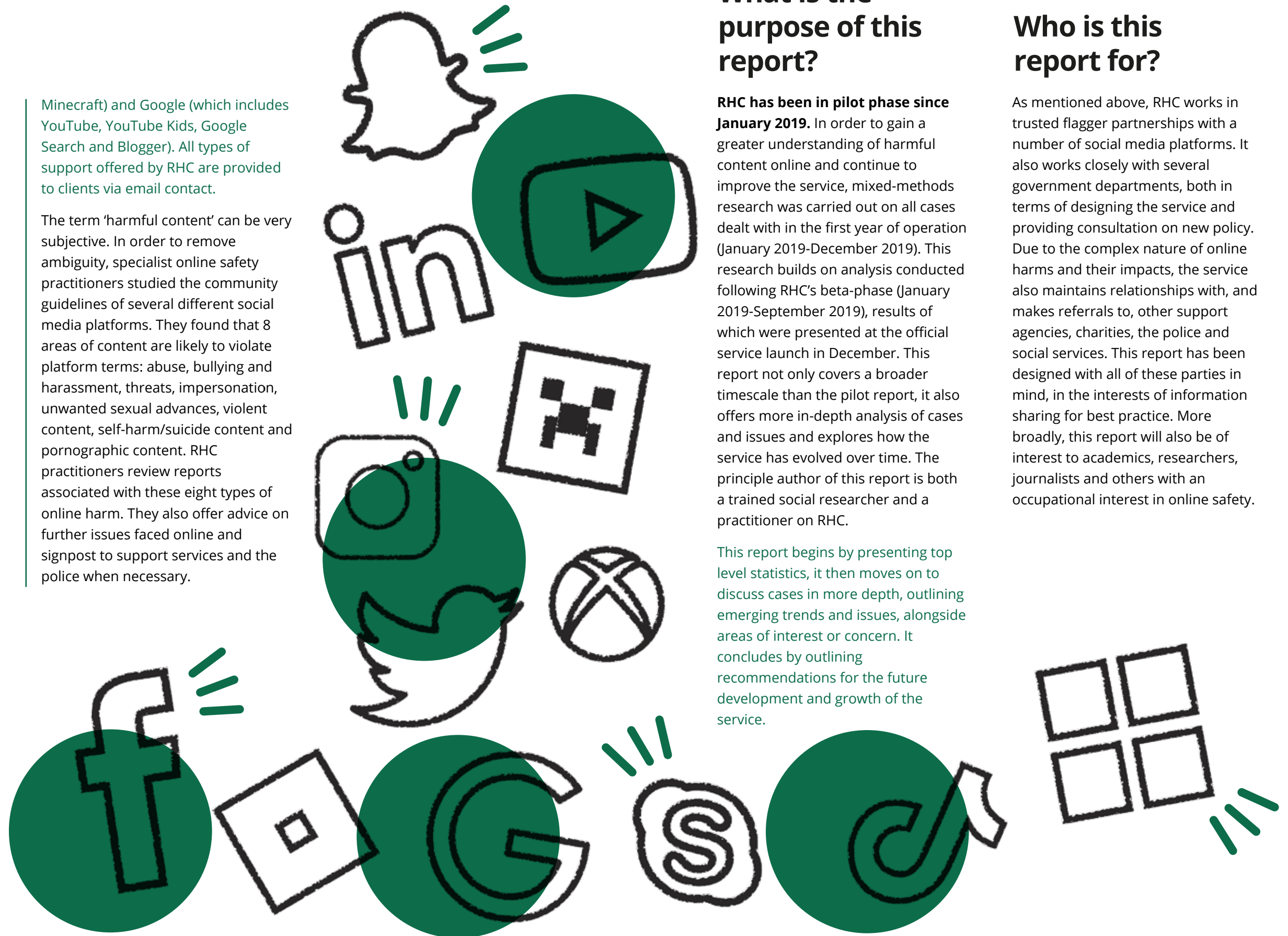
## What is the purpose of this report?

**RHC has been in pilot phase since January 2019.** In order to gain a greater understanding of harmful content online and continue to improve the service, mixed-methods research was carried out on all cases dealt with in the first year of operation (January 2019-December 2019). This research builds on analysis conducted following RHC's beta-phase (January 2019-September 2019), results of which were presented at the official service launch in December. This report not only covers a broader timescale than the pilot report, it also offers more in-depth analysis of cases and issues and explores how the service has evolved over time. The principle author of this report is both a trained social researcher and a practitioner on RHC.

This report begins by presenting top level statistics, it then moves on to discuss cases in more depth, outlining emerging trends and issues, alongside areas of interest or concern. It concludes by outlining recommendations for the future development and growth of the service.

## Who is this report for?

As mentioned above, RHC works in trusted flagger partnerships with a number of social media platforms. It also works closely with several government departments, both in terms of designing the service and providing consultation on new policy. Due to the complex nature of online harms and their impacts, the service also maintains relationships with, and makes referrals to, other support agencies, charities, the police and social services. This report has been designed with all of these parties in mind, in the interests of information sharing for best practice. More broadly, this report will also be of interest to academics, researchers, journalists and others with an occupational interest in online safety.



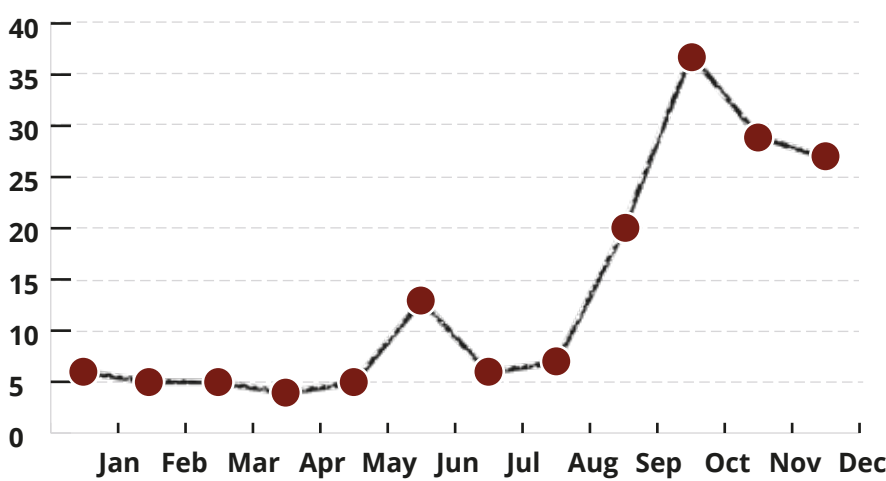


# Top level statistics

## Helpline use and growth

In the year analysed, the RHC website received 9,282 visitors and practitioners dealt with 164 unique cases. Figure 1 shows the volume of cases per month. Evidently, the service is increasing in popularity as awareness grows. Popularity rapidly advances in September, coinciding with the transition from beta to alpha phase, the launch of the new RHC website and the build-up to the service launch.

Figure 1: Volume of cases per month



## Service response

**RHC practitioners respond to cases in one of four ways.** Responses are dependent upon various factors including nature of harm, location of content, age of client, whether the client is based in the UK, potential criminality of content and previous reporting channels pursued.

**1. Escalated to industry:** Where content is (1) deemed to fall under the definition of an online harm, (2) located on one of the social media platforms with which RHC works in partnership and (3) has already been reported unsuccessfully to the platform by the client, practitioners escalate reports for review. This is done directly to social media contacts via trusted

flagger routes. The client must be over the age of 13 and UK based for this route to be pursued.

**2. Redirected to industry reporting links:** Where all of the above conditions are met, with the exception that the client has not already made a report to social media, practitioners direct clients to the correct reporting links and encourage them to re-report to RHC should industry reports be unsuccessful.

**3. Outside of project remit:** Some cases fall outside the remit of the RHC project, including clients located outside of the UK, content hosted on

non-partnership platforms, reports where no clear harm is evident and reports concerning offline, rather than online, issues. In these instances, practitioners provide clients with clarification as to the correct nature of their issue and direct them to more appropriate sources of support.

**4. Criminal/legal:** Where the issue reported is clearly criminal in nature, practitioners direct clients to the appropriate law enforcement bodies (e.g. the police, Child Exploitation and Online Protection command (CEOP) and the Internet Watch Foundation (IWF)).

As part of all four responses, practitioners also provide signposting to additional sources of support (practical and emotional) where necessary.

**Figure 2 shows the breakdown in service response. Whilst this chart displays a relatively even split in responses, in the majority of instances (35%) content was escalated to industry.**

Because the two categories 'escalated to industry' and 'redirected to industry reporting link' fall under the explicit aims and objectives of the RHC service, service responses can be further broken down into clients who were assisted internally and directly (51%) and those who presented with issues outside of the project remit and were thus directed to external sources of support (49%).

## Following initial service response

**Of the 35% of content escalated to industry, 92% was successfully actioned by industry (e.g. removed/restricted/ regained access to).** Of the 92% of content successfully actioned, 62% was done so within 72 hours of practitioner requests.

Of the 16% of clients redirected to industry reporting links, only 9% got back in touch with RHC, either seeking clarification or re-reporting an unsuccessful industry report.

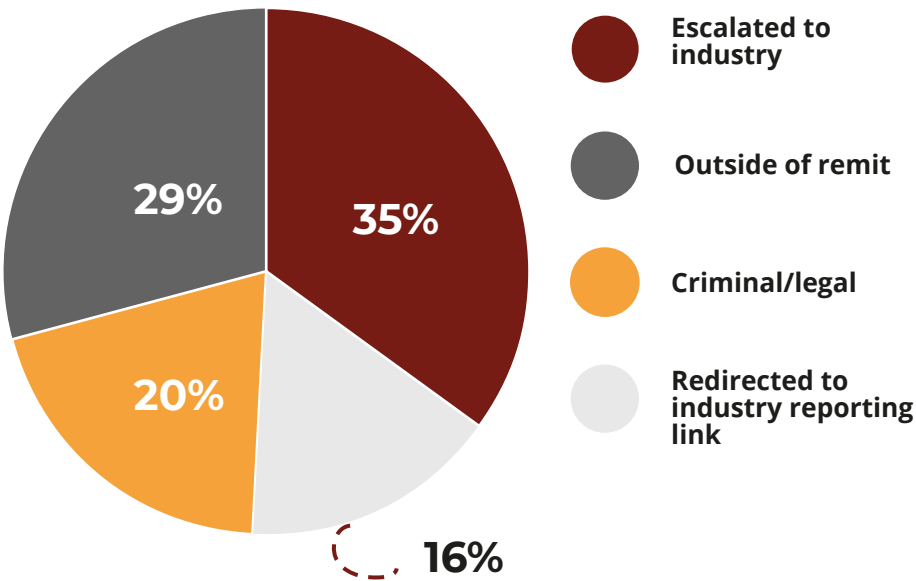


Figure 2: Proportion of service response by type

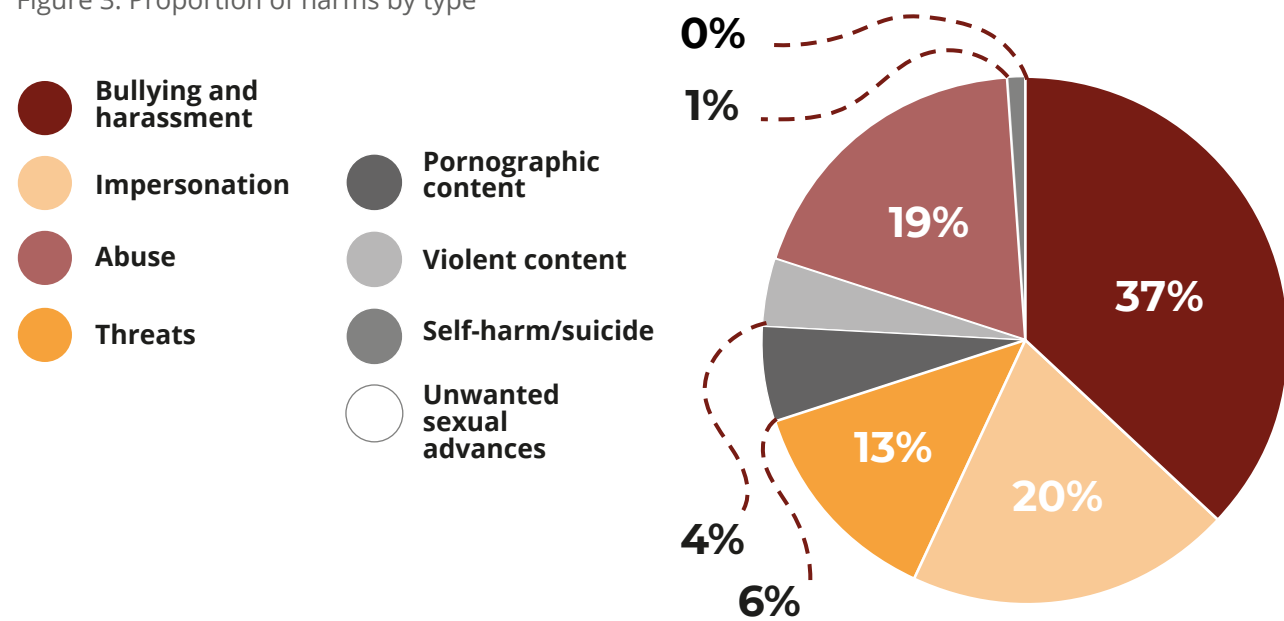
Of the 30% of clients reporting content outside of project remit who were re-directed to more appropriate advice, only 5% got back in touch with RHC.

Of the 19% of clients who reported criminal/legal issues, 47% got back in touch with RHC. At times, this was under the guidance of law enforcement who were investigating the issue and seeking further help with the online aspects of it. In other instances, however, clients returned to RHC having had an unsatisfactory experience with law enforcement (e.g. police displayed insufficient understanding of online crimes and the issue was dismissed).

## Nature of cases

**Out of the main eight online harms, cases involving bullying and harassment were most common (79 cases).** This was followed by impersonation (43 cases), abuse (40 cases), threats (27 cases), pornographic content (13 cases), violent content (8 cases), self-harm/suicide content (2 cases) and unwanted sexual advances (1 case). Many cases involved more than one harm, which is why the total harms recorded is greater than the total number of cases. Figure 3 shows the proportion of each type of harm against the overall harms recorded.

Figure 3: Proportion of harms by type

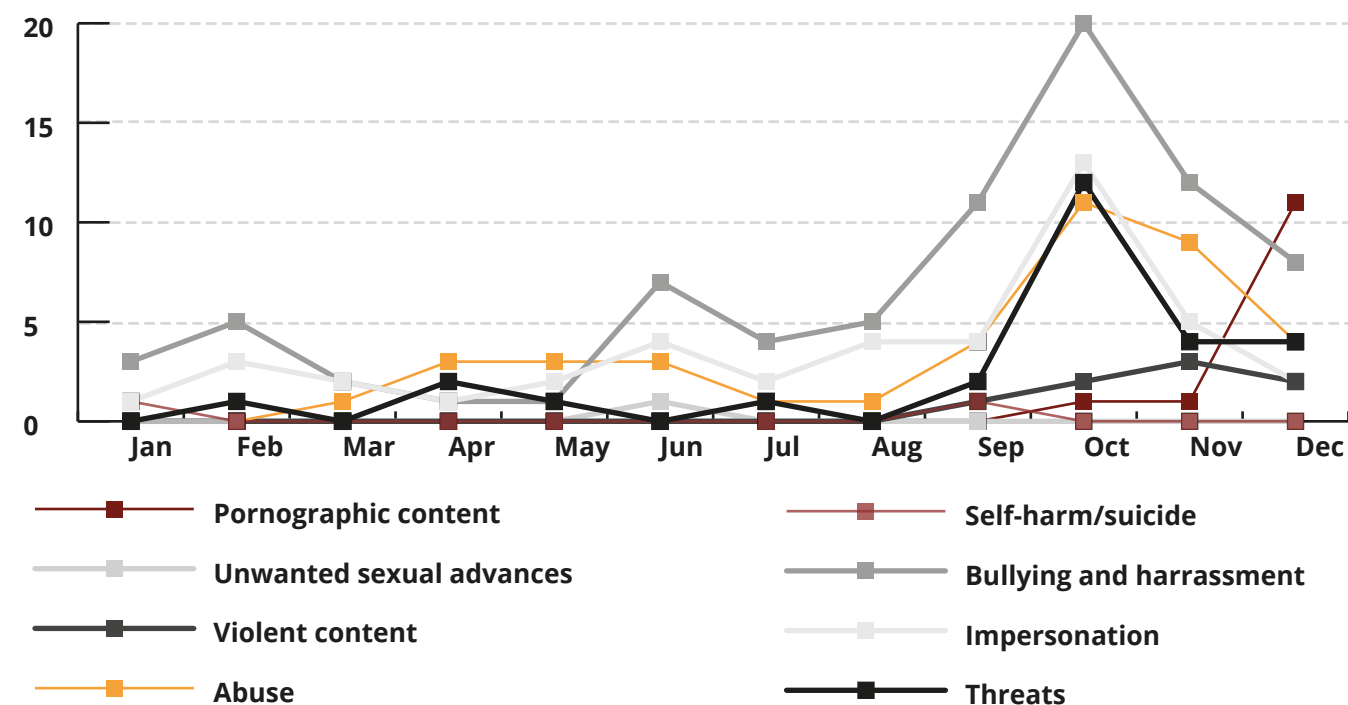


Alongside the main eight harms, the service received reports concerning a number of other online issues. The most common of these was privacy violation (e.g. personal data shared without consent, including photos, videos, ID documents, home address and telephone number, 92 cases), followed by hacked accounts (20 cases),

intimate image abuse (20 cases), hate-speech (18 cases), account access issues (7 cases), Child Sexual Exploitation Material (7 cases), intellectual property concerns (6 cases), blackmail/fraud (5 cases) and grooming (3 cases). Figure 5 shows the proportion of each type of issue against the overall 'other online issues' recorded.

The types and proportion of harms being reported was fairly consistent throughout the year, with the exception of September-December, which saw a large spike in reports concerning pornographic content and a smaller, although still significant, increase in reports concerning violent content. Figure 4 shows type of harms reported by month.

Figure 4: Frequency of harms by type per month



Oftentimes, reports made to the service not only concerned online harms, but associated offline issues and contexts. The most common associated offline issue/context was negative impact on client's mental health (54 cases), followed by bullying and harassment (25 cases), criminal case/investigation (22 cases), domestic abuse (17 cases), business/workplace dispute (14 cases), school involvement (11 cases), political campaign (10 cases), honour based abuse (6 cases), social service involvement (5 cases), stalking (2 cases), involvement of a public figure/celebrity (2 cases) and trafficking (1 case). Figure 6 shows the frequency of associated offline issues and contexts by type.

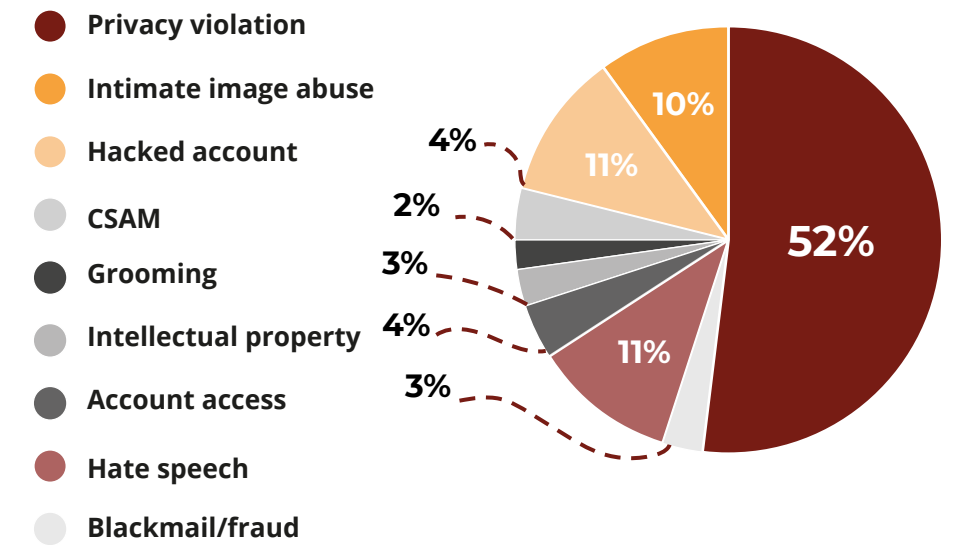


Figure 5: Proportion of other online issues by type

## Referral routes

**11% of clients described having been referred to Report Harmful Content by a service or agency.** Of this 11%, 7% were referred by the police, followed by the Professionals Online Safety Helpline (2%) and the Revenge Porn Helpline (2%). These figures might not represent the total proportion of referrals, however, as practitioners do not routinely ask for this information.

Practitioners commonly signposted clients to other agencies and support services. Figure 7 shows signposting breakdown by agency or service.

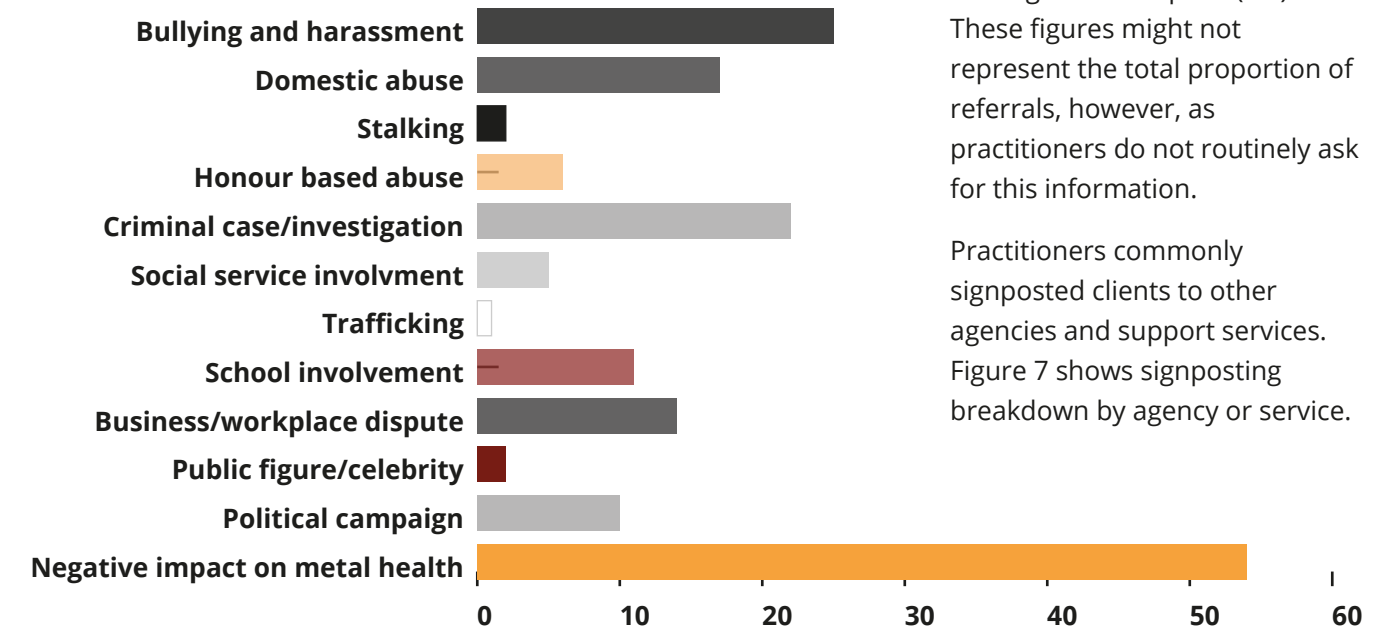
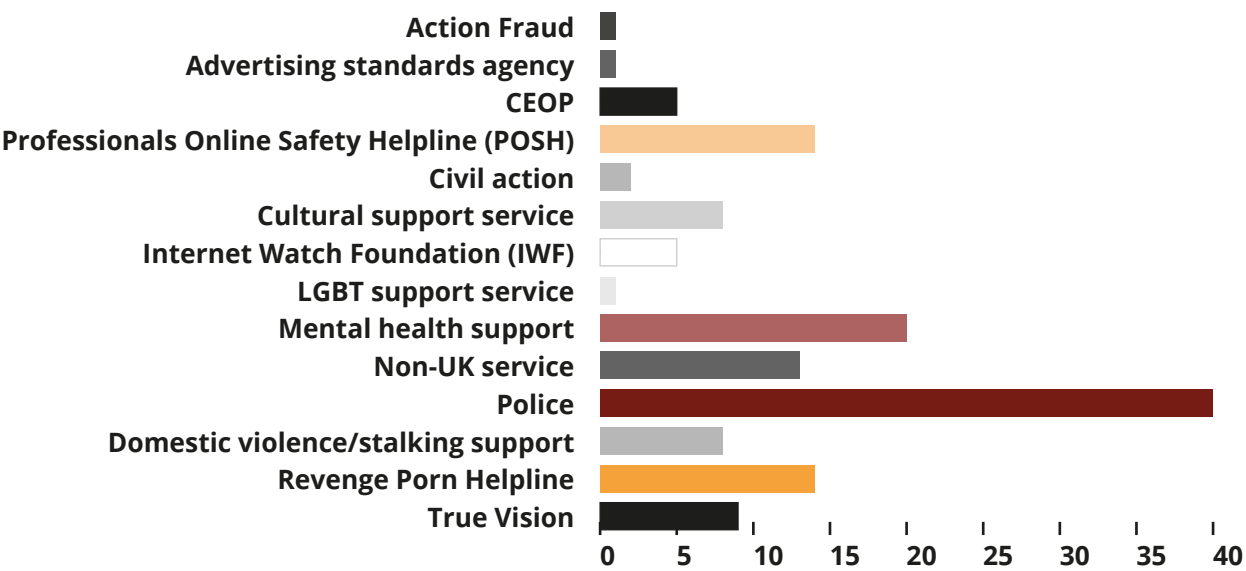


Figure 6: Frequency of associated offline issues and contexts by type.



Figure 7: Frequency of signposting according to service/agency



How were clients helped?

In addition to escalating content to industry and redirecting to correct industry reporting links (as discussed above under ‘service response’), practitioners were able to assist clients by providing them with information and/or clarification on the nature of their issue or redirecting them to reporting links on other (non-industry) sites.

They also offered emotional support, alongside signposting clients to other agencies and services, for either emotional or practical support (as discussed above under ‘referral routes’). Figure 8 shows a frequency breakdown for the way in which clients were helped. It was common for clients to receive assistance in more than one

why the total recorded assistance channels is greater than the total number of cases.

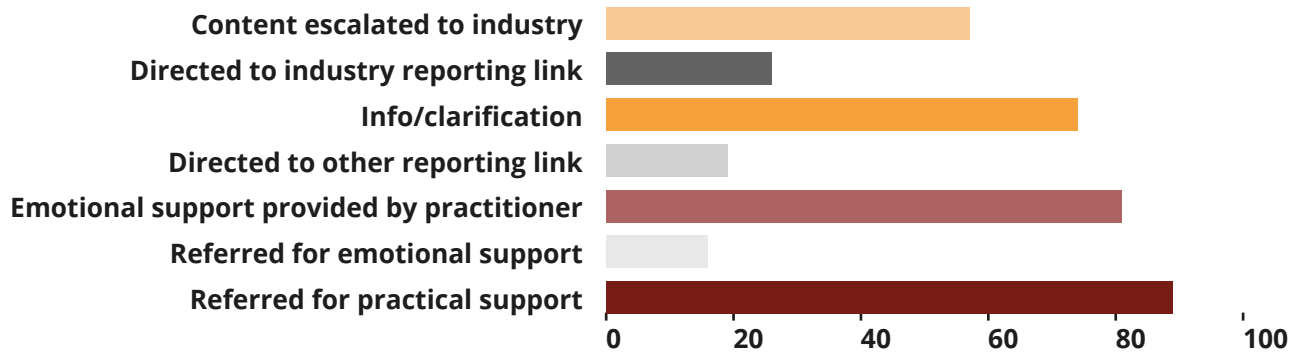


Figure 8: Frequency of assistance channels

Qualitative data also offers insight into the way in which clients were helped. Client testimonials (either communicated directly to practitioners or through follow up surveys) revealed the positive impact of RHC’s assistance. Clients remarked on the speed of the service, making comments such as:

“The speediness of the response with clear advice on what to do next ensured that the situation was resolved quickly and effectively.”

“The staff are very proactive and the results were fast.”

Clients also remarked on the quality of support given by practitioners making comments such as:

“The advisor got the job done, she was amazing, she was fabulous. Great service!”

“I found the ladies extremely helpful, really helpful and respectful. They didn’t judge me, they knew my family needed help and support and they gave their time and commitment to ensure the videos were removed promptly. I will be forever grateful to them for assisting me through these tough times.”

Finally, the overall impact that RHC had on addressing the online harm was commented upon. Clients wrote:

“I was relieved that finally I could be supported directly. Without Report Harmful Content, I think I would be in a very different place today. This service allowed me to take back control and protect my children.”

“It is very easy to feel alone, especially with other agencies’ lethargy, but the tremendous role you have played and your ongoing assistance have been a godsend.”

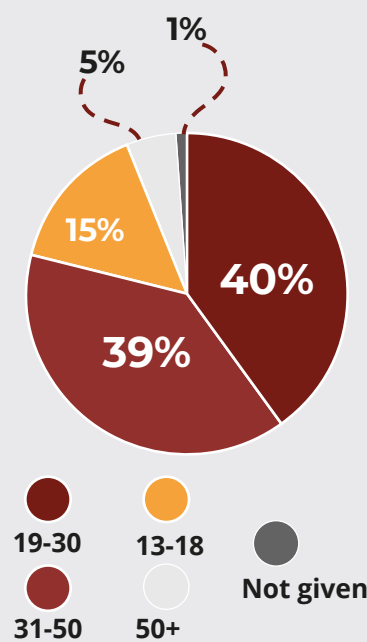


Figure 9: Age of clients

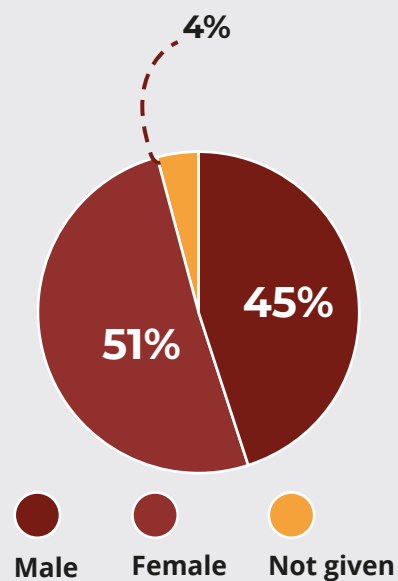


Figure 10: Gender of clients

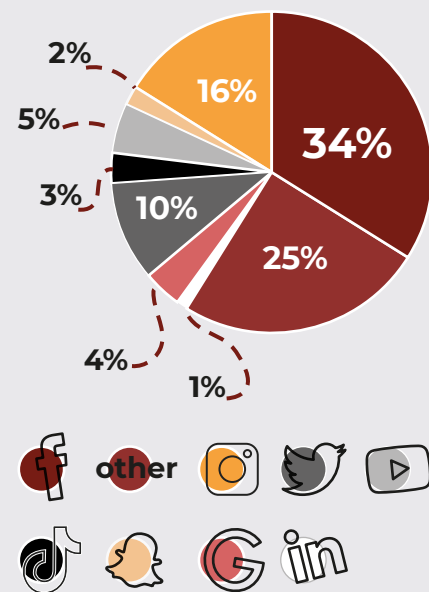


Figure 11: Location of content

## Client demographics

RHC collects basic demographic information from clients (age and gender), alongside recording the location of the harmful content. The age group most likely to report to RHC was 19-30 (65 cases), closely followed by 31-50 (64 cases). Less likely to make reports were those aged 13-18 (25 cases) and 50+ (9 cases) (1 case did not give their age). Figure 9 shows age group of clients, represented as a proportion of total cases.

The gender of RHC clients was relatively evenly split, with females only marginally more likely to use the service than males (83 cases as compared to 73). Figure 10 shows gender of clients, represented as a proportion of total cases.

Of the large social media platforms, harmful content was most likely to be located on Facebook (62 pieces of content),

Twitter (18), YouTube (10), Google (7), TikTok (6), Snapchat (4) and LinkedIn (1). Forty-six pieces of content were located on sites other than the social media sites with which RHC work in partnership. Clients occasionally made reports about multiple pieces of harmful content, located on a range of platforms, which is why the total pieces of harmful content is greater than the total number of cases. Figure 11 shows location of harmful content, represented as a proportion of total pieces of harmful content.

Client demographics and location of content were cross-tabulated to reveal further trends. Figure 12 shows proportions of content location by age group. As this figure indicates, the younger age group (13-18) were more likely to report content on platforms such as Instagram, Snapchat and TikTok, with the older age groups more likely to report content on Facebook, Twitter and Other sites. No one in the 50+ age group reported content on Snapchat, YouTube, TikTok, Google or LinkedIn.

Figure 12: Proportions of content location by age group

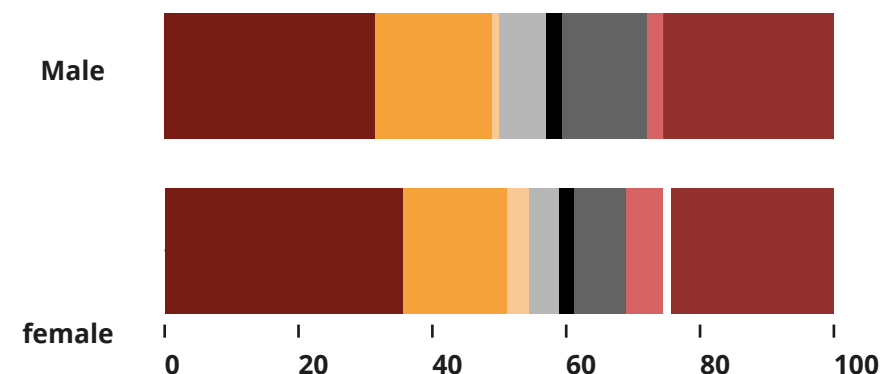
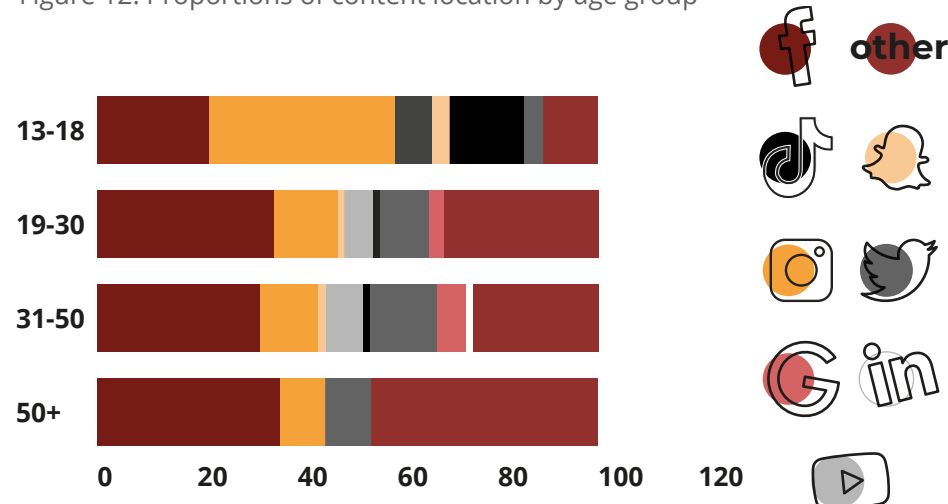


Figure 13: Proportion of content location by gender

Shifts in the location of harmful content could be noted throughout the year. In particular, the proportion of reports concerning harmful content on 'Other' sites grew steadily as the year progressed. Figure 14 shows proportions of content location by month.

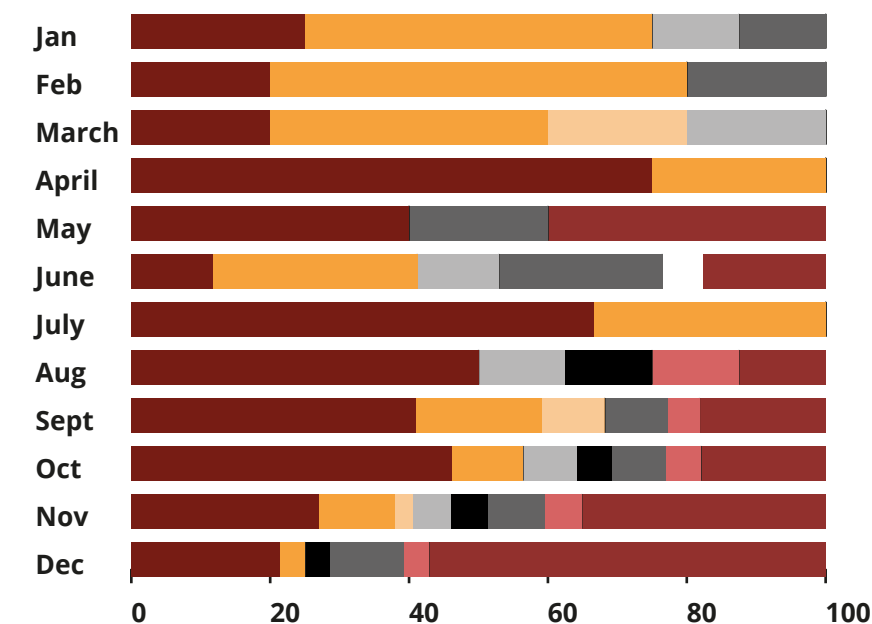


Figure 14: Proportion of content location by month

# Exploring cases in more depth

## Emerging trends

### Trend one: impersonation + bullying and harassment + privacy violation

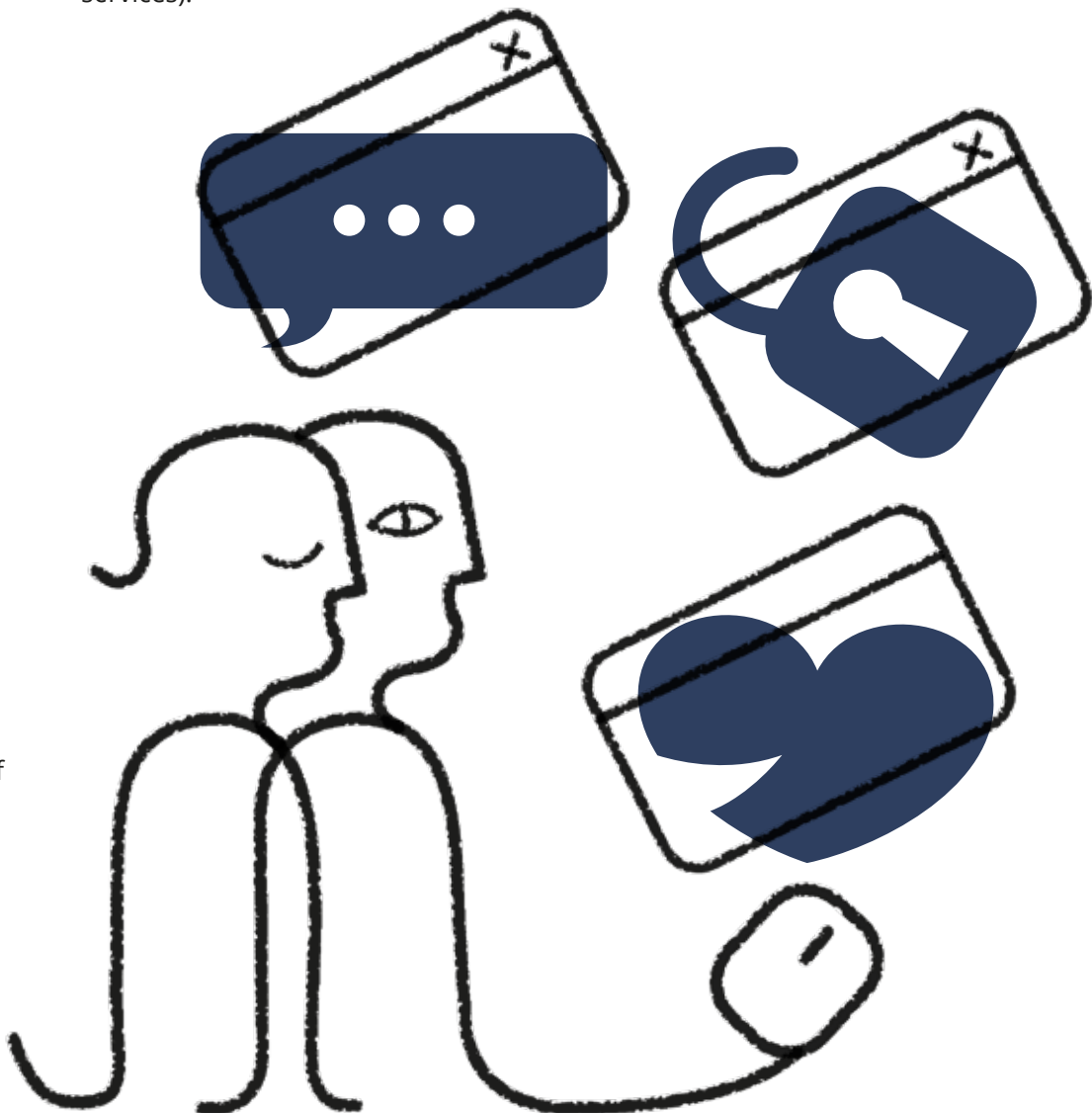
It was rare for a client to report just one online ‘harm’; harms frequently overlapped and intersected with other issues, both on and offline. The first trend identified on RHC contained a cluster of three types of issues: impersonation, bullying and harassment, and privacy violation. The way these three issues intersected was largely through someone creating a fake social media profile, masquerading as the client by using their videos, images and other personal data. This profile was used over a period of time to bully the client, by spreading lies about them, ‘outing’ them or harassing them with abusive language and/or humiliating images and videos. 20% of RHC cases contained all three of these issues and a striking 44% contained two out of the three (in various combinations).

This trend most commonly occurred on Facebook (32%), closely followed by Instagram (29%). Women were much more likely to be the victim of this type of harm (60%), particularly those in the 31-50 age group (females aged 31-50 comprised 32% of this cluster). Whilst men were less likely to report this trend

than women (39%), when they did, they were in a younger age bracket (26% age 19-30 vs only 7% males aged 31-50). Usually (in 69% of cases) the perpetrator of this harm was known to the client (e.g. ex-partner, friend, family member or colleague). 25% of clients also described this cluster of online harms as being connected to offline bullying and harassment, with an additional 19% describing ongoing and serious domestic violence. Finally, 43% of clients impacted by this trend described it as negatively affecting their mental health.

Because of the impersonation aspect of this type of abuse, it often appeared subtle to the outside observer, or went undetected by social media platforms. The role of RHC practitioners was thus to communicate the true nature and broader context surrounding this trend when escalating to industry. In many of these instances, RHC practitioners were able to action the harmful content, resulting in impersonation accounts being removed and private information being deleted. Practitioners did, however, acknowledge that actioning

content only deals with a small aspect of this trend. This was evident through the way in which they made onward referrals to the police (in 26% of these cases), domestic and sexual violence services (11%) and cultural support services (11%) (e.g. honour based violence services).



### Trend two: abuse + threats + hate speech

The second RHC trend also contained a cluster of three issues, namely abuse, threats and hate speech. Whilst only 6% of RHC cases contained all three of these issues, 16% contained hate speech with either abuse or threats. This latter figure, alongside the distinct nature of this type of harm, was considered significant enough for this cluster of issues to be identified as an emerging trend. Figure 15 shows the proportion of hate speech by type.

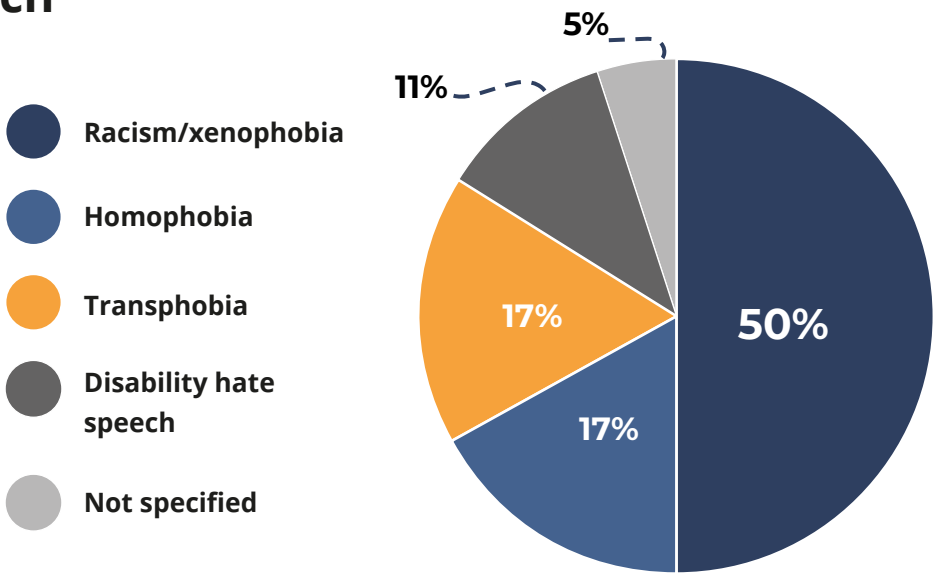


Figure 15: proportion of hate speech by type

This trend was most common on Facebook (59%), closely followed by Other platforms (26%) and Twitter (19%). This trend seemed to be split more equally between the genders, with women making 48% of reports, men 37% and the remaining 15% not giving their gender. This trend was also reported fairly evenly across the age groups: 22% of these reports were made by 13-18 year olds, 33% by 19-30 year olds and 37% by 31-50 year olds (the 50+ age group, making only 7% of these reports, being the exception). 1% did not disclose their age.

Unlike trend one, clients reporting trend two were overwhelmingly bystanders (92%) as opposed to victims (8%). Furthermore, perpetrators were usually unknown individuals on social media (67%), as opposed

to family members, friends or partners. This type of abuse was reported as an isolated event, as opposed to an ongoing pattern. In addition to this, trend two did not significantly intersect with offline abuse, harassment and/or domestic violence (as was the case with trend one). Despite these differences, and even though they were only bystanders, 22% of clients described this type of online harm as negatively impacting their mental health. Regarding this trend, RHC practitioners were able to action the harmful content, resulting in it being removed and/or perpetrator accounts suspended. Beyond this, practitioners also made onward referrals to the police (44% of cases) and True Vision (19% of cases).



## Trend three: perpetrator-less harms

Alongside clusters of harms, a third type of trend was identified in RHC cases. As discussed above, in trend one and two, a perpetrator was evident (whether that be the family member, friend or partner carrying out harassment via impersonation, or the unknown individual engaging in hate speech). Contrary to this, the third RHC trend involved clients inadvertently viewing harmful content rather than being victim of or witnessing targeted, harmful behaviour. The types of harmful content viewed included graphic/violent content, self-harm/suicide content or pornography, often overlapping (e.g. violent pornography, sexualised self-harm). These reports can be grouped together as a trend due to their 'perpetrator-less' nature.

These types of reports comprised 13% of RHC cases and were most common on Other sites (50%), followed by Facebook (18%). Men were most likely to report this type of content (64%), with men in the 19-30 age group reporting a significant proportion (39%). This can be explained by the fact that the majority of this content was hosted on pornography sites, with young men being more likely to visit these platforms.

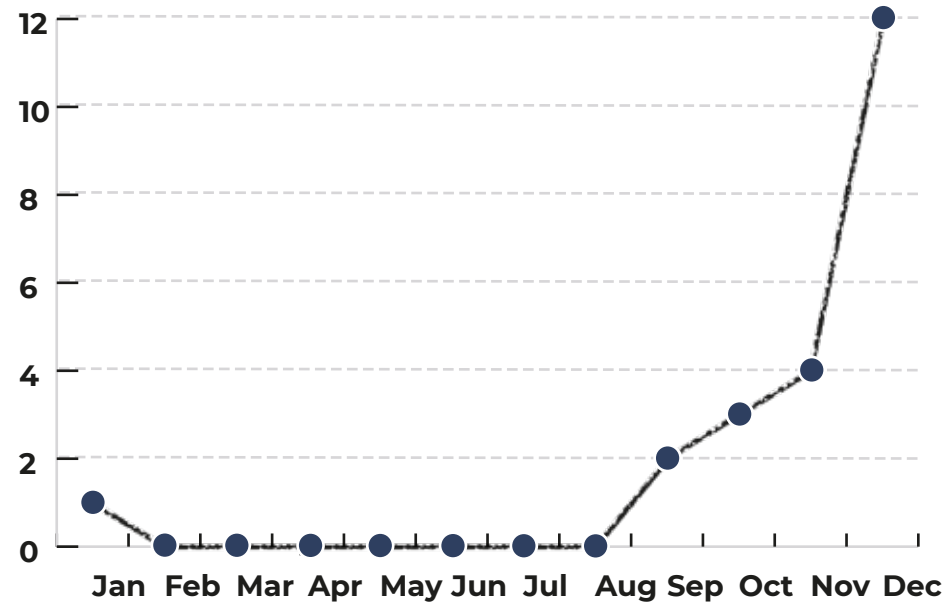


Figure 16: Number of trend three cases by month

32% of clients described viewing this content as having a negative effect on their mental health. Due to the high proportion of this content located on Other sites, RHC practitioners were often unable to action it. They did, however, offer information and clarification and signpost to other support services, most commonly the IWF (18%). As touched upon in previous sections, reports of this nature began to increase towards the end of the year, peaking in December. Figure 16 displays this increase.

## Emerging concerns and challenges

### Mental health

**The high proportion of clients experiencing negative mental health impacts as a result of witnessing harmful content online is concerning.** As already discussed 32% of total clients reported negative mental health impacts. This figure rose to 43% for clients affected by trend one.

Of that 32%, 13% of clients described feeling suicidal. For example, one client was being repeatedly harassed by a relative over social media. She had tried to report her issue to the police, with no success. When she made

a report to RHC she was desperate. She told practitioners:

**'I have (already) tried to commit suicide with an overdose but she is still carrying on I don't know what to do anymore other than another overdose'.**

Aside from suicidal ideation, other reported mental health impacts included distress (70%), anxiety (52%), a decline in social functioning (36%), depression (27%), agoraphobia (5%) and post-traumatic stress disorder (4%). 18% of clients experiencing negative mental health impacts had sought medical treatment (e.g. medication or therapy). Figure 17 shows the frequencies of these different types of impacts.

In addition to causing new mental health problems, harmful online content was described as exacerbating existing mental health issues. For example, one client had recently left an abusive relationship. Her ex-partner created numerous fake social media profiles in her name, with the aim of continuing his harassment of her. She told practitioners:

**'I had PTSD because of him and this had settled with a lot of therapy, but has recurred since all this online abuse started again'.**

Often, social media had been a positive coping mechanism for clients who were already mentally unwell. Being targeted online threatened this coping mechanism. One client, who was being harassed over social media, told practitioners:

**'I (was) already on medication for my depression and suicide attempts...I don't go online to be abused. As someone with agoraphobia...it is my only way to interact with friends and the wider world. I can feel this slipping away right now'.**

Finally, mental health impacts went beyond just the 'victim' and could also be seen to affect family and friends who reported on their behalf. One friend, acting as an advocate, told practitioners:

**'We are worried for her well-being. She has a history of self-harm and attempted suicide. Unless her ex can be stopped and/or forced to remove the videos I fear for her well-being, let alone my own mental state. I'm currently signed off with depression and anxiety because of this...I'm at my wits end and close to a full emotional breakdown'.**

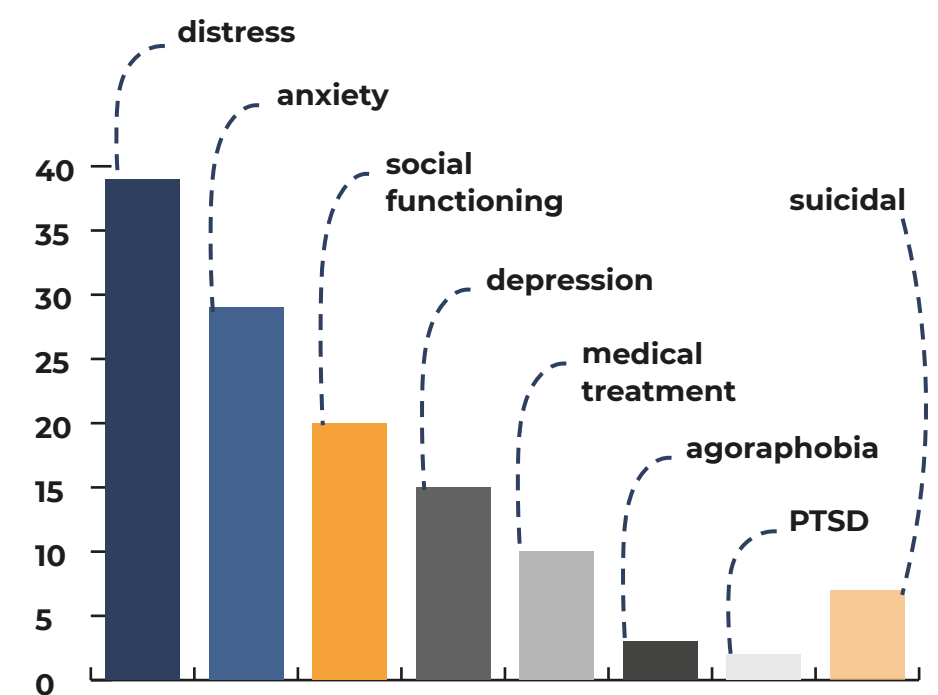
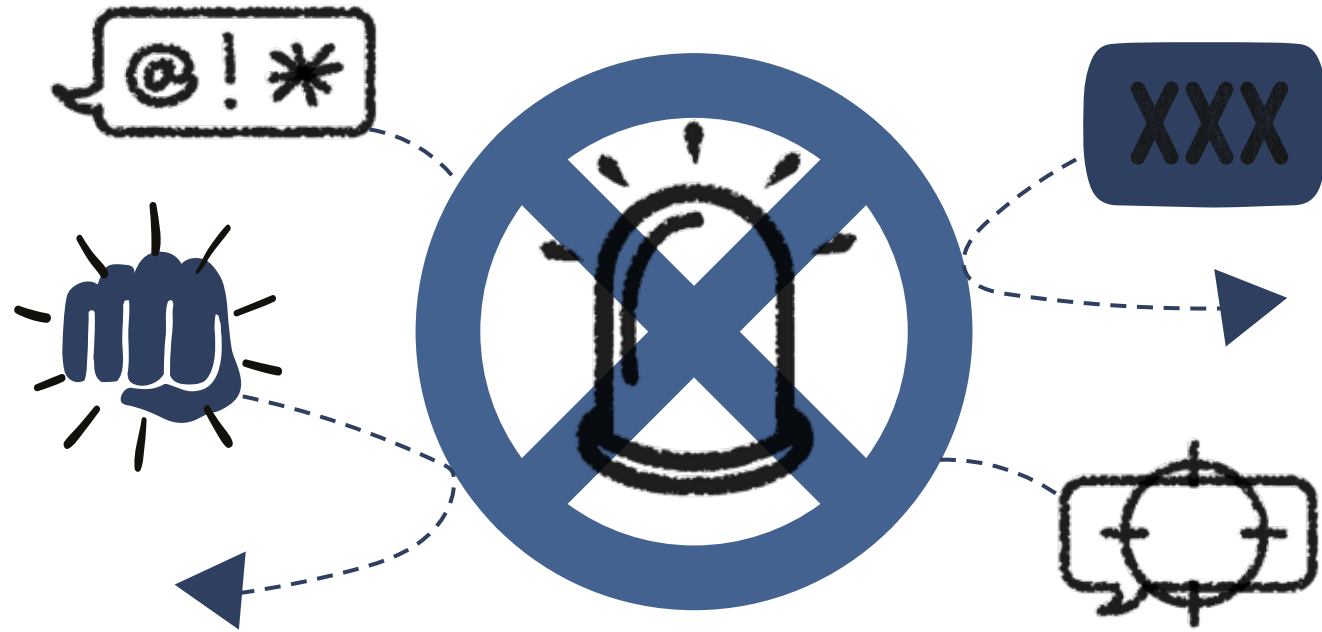


Figure 17: Frequency of mental health impacts by type



## Criminal/legal issues

As discussed in the top level statistics, there were instances where practitioners worked closely back and forth with law enforcement: 7% of clients were referred to RHC by the police and 19% were signposted onwards to law enforcement. Practitioners made the decision to refer onwards to law enforcement in a number of situations; firstly, where content was deemed to be outright criminal, secondly, in instances of ongoing harassment with an offline element, wherein RHC had reached the limit of what they could do to help and thirdly, when practitioners felt that escalating content for removal would interfere with an ongoing criminal investigation. Practitioners have a full understanding of the law. They work closely with this understanding to ensure they are striking the right balance between not making unnecessary referrals of

non-criminal content to the police and not taking on issues which would be better dealt with by law enforcement.

However, as also discussed in the top level statistics, after making referrals to the police, an extremely high proportion (47%) of clients got back in touch with RHC. At times, this was positive (e.g. police had assessed the case, were taking it seriously and wanted assistance with the online aspects of it). In other instances, however, it was less positive; police commonly displayed insufficient understanding of online crimes and dismissed clients, incorrectly informing them that their issue was non-criminal and would be better dealt with by services such as RHC. When this occurred, RHC practitioners became caught in a frustrating loop with clients as they were continually referred back and forth between the service and the police.

## Cultural and religious issues

A significant proportion of clients reporting cluster one harms (impersonation + bullying and harassment + privacy violation) were Muslim and outlined additional cultural and/or religious issues which complicated their cases. In these instances, perpetrators (usually ex-partners or family members) disclosed images and/or videos which severely threatened the privacy of clients. This type of material included images of female clients with their hair uncovered or displaying parts of their bodies not prohibited by religious conventions (e.g. upper arms and shoulders). It also included images of clients kissing or hugging non-marital partners and videos of female-only religious ceremonies.

This type of material did not meet the legal threshold for private, intimate material and thus was not covered under intimate image abuse laws.

Nevertheless, because of the cultural context, clients perceived this material to be just as private and experienced its disclosure as extremely distressing and violating. One client told a RHC practitioner:

**‘I understand this video doesn’t contain sexual stuff but this is a private video which shouldn’t have been shared to the outside world...it was not meant to be seen by anyone and it is bringing shame on me to have so many people see it’.**

As a result of this type of material being disclosed, clients risked being ostracised from their families and wider communities. One client told practitioners:

**‘This is making me feel very threatened and upset...Even though the picture isn’t explicit, it’s on there without my permission. I would really like this to be removed as it could cause serious issues between me and some of my family members’.**

Furthermore, the disclosure of this type of material was often carried out with the explicit intention of inciting honour based violence. In one

particularly severe incident, a client had a ‘bikini photo’ of her published online. She had already experienced violence and shaming as a result of this and the ongoing risk to her safety, and that of her family, was high.

These types of cases were concerning for practitioners, primarily because of the level of risk posed to clients and the safeguarding implications. Practitioners have the option of making referrals to social services where safeguarding concerns are high although, in reality, this is often unfeasible due to the anonymous nature of the RHC service. As such, in order to address risk, it was common for practitioners to signpost these types of cases to the police and cultural support services. A further issue in dealing with these types of cases

These types of cases were concerning for practitioners, primarily because of the level of risk posed to clients and the safeguarding implications. Practitioners have the option of making referrals to social services where safeguarding concerns are high although, in reality, this is often unfeasible due to the anonymous nature of the RHC service. As such, in order to address risk, it was common for practitioners to signpost these types of cases to the police and cultural support services. A further issue in dealing with these types of cases was the difficulty in

communicating their harmful nature to industry partners. As indicated, the type of content disclosed in these cases often didn’t breach the law or, indeed, social media community guidelines. Whilst practitioners outlined the full context to industry partners, it was still common for a considerable length of time and much back and forth to take place before content was actioned.

## Clash of protected characteristics

The Equality Act sets out nine ‘protected characteristics’, on the basis of which it is against the law to discriminate. These include age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation (Equality Act, 2010). One final emerging challenge for RHC practitioners could be found in cases where there was a clash of protected characteristics.

In one instance, a client ran a LGBTQ+ group on social media which was receiving homophobic reviews, using religion as a justification. The type of language used in these reviews was obviously harmful. Whilst the platform in question took a considerable length of time to respond to this report (given the clash of religious belief and sexual orientation), ultimately,



the content was deemed to breach hate speech and hateful conduct guidance, resulting in its removal.

Although this case was relatively easy to resolve, other cases concerning two protected characteristics were not, in particular where gender reassignment and sex clashed. RHC practitioners dealt with a number of cases where clients had reported content as transphobic. Oftentimes, this content was simply advocating for the protection of women's sex based rights and did not contain any explicit transphobia or, indeed, mention trans people at all. In these instances, the difficulty for practitioners lay in communicating this to clients, some of whom had misguided definitions of what constituted hate speech. This difficulty was compounded both by the lack of platform guidance regarding this issue and the inconsistency in responses to this type of content across different platforms.

## Relationships with platforms

**As was indicated in the top level statistics, RHC practitioners have effective working relationships with industry partners:** content escalated to them was, in the majority, successfully and rapidly actioned. Nevertheless, as touched upon elsewhere in this report, other issues with industry partners were found to exist. These issues can be categorised into three areas. Firstly, there was often an inconsistency in both response rate and type of content successfully actioned, with a lack of explanation as to why. At times, content was removed by industry partners without question. In other instances, however, extremely similar types of content were not actioned or actioning took much longer; RHC practitioners would have to go back to industry reiterating the harms, explaining context in great depth and drawing explicit attention to the specific breach of community guidelines.

The second platform issue related to the disparity in volume of reports across different industry partner platforms. As previously indicated, RHC received a high volume of reports from Facebook (62) and Instagram (30), a lower volume from Twitter (18), YouTube (10), Google (7), TikTok (6), Snapchat (4) and LinkedIn (1) and none from other industry partners including Roblox, many Microsoft services (e.g. Bing, Xbox and Minecraft) and a number of Google services (e.g. YouTube Kids and Blogger).

The final platform issue concerned the increasing volume of reports received from platforms with which RHC do not have relationships (the 'Other' category). This included content hosted on a variety of platforms including pornography sites, independently owned and moderated online communities, blogging sites and e-commerce sites. In these instances, practitioners offer clients tailored advice, signposting and support. Nevertheless, the lack of ability to take more substantial action was frustrating, for practitioners and clients alike.

# Discussion and Recommendations

## The value of the service

**Report Harmful Content is clearly meeting its objective of helping everyone to report harmful content online.** It deals with reports from a range of demographics, across a number of platforms. As is evident from this report, RHC practitioners deal with a wide variety of online harms, the majority of which overlap with other harms and issues, both on and offline. The value of the service lies in the way in which it addresses online harms, not in isolation, but holistically. This is evident through the way in which practitioners draw upon a range of escalation options, support services and referral routes in order to offer support that is uniquely tailored to individual cases.

Not only is RHC effective at tackling the complexity of online harm, it is also efficient. The high percentage of content which was successfully actioned by industry, alongside the rapid response rate of industry to practitioners clearly demonstrates this. Furthermore, the low percentages of clients who got back in touch with RHC after being offered advice and/or signposting can be taken as evidence that practitioners are providing precise instructions to clients to deal with a range of

online harms and issues. The high level of referrals to RHC from the police, alongside the openness for police to work on cases in conjunction with practitioners, demonstrates the way in which RHC is becoming a trusted service to be used in conjunction with official criminal procedures.

Finally, the steady growth in reports as the year progressed evidences the clear and increasing demand for this service. The diversification in reports towards the end of the year also evidences the spread of demand across a broader range of issues. RHC practitioners are keen for the service to expand and develop, however, they are currently working at full capacity. To this end, an increase in funding is desperately needed to meet existing demand and to equip practitioners to deal with the widening range of cases.

## Responding to emerging trends

**As discussed, three emerging trends could be identified on RHC.** Each trend has its own implications regarding both the type of support RHC and other services should be providing to clients and, more broadly, in terms of policy and legislation.



The first trend was a cluster of three online issues: impersonation, bullying and harassment and privacy violation. This cluster also intersected with offline abuse and was much more likely to affect women, in particular those aged 31-50. In order to tackle this cluster of issues, RHC practitioners should continue to work holistically with individuals in order to address both their online and offline needs. In practice, this means continuing to direct clients to additional support services and law enforcement channels, alongside monitoring cases over a longer period of time, keeping channels of communication open should the issue resurface. It is recommended that all support services coming into contact with this trend acknowledge the complex ways online and offline abuse intersect, thus recognising that a range of tools and a multi-agency approach is necessary if this type of abuse is to be tackled.



- **Significantly, trend one can also be seen to confirm existing knowledge surrounding the gendered nature of intimate/domestic abuse and harassment**

(Dobash & Dobash, 2004; ONS, 2019; Women's Aid, 2019), alongside the growing trend for this type of violence to be perpetrated through technology (Drouin, Ross & Tobin, 2015; Refuge, 2020; Thacker, 2017). This confirmation supports the broader need for law and policy to be revised so as to address the technological aspects of intimate/domestic abuse, alongside recognising its severe psychological impacts (Home Office, 2019).

- The second trend identified was also a cluster of harms and issues: abuse, threats and hate speech. This trend was reported fairly evenly across genders and age groups, was reported by bystanders as opposed to victims and did not appear to intersect with offline abuse. Identification of this trend also has implications in terms of policy and legislation in that it reflects recent research documenting the growth of online hate speech, in particular racism, extremism and far right ideology (Williams & Mishcon de Reya, 2019). Worryingly, research also indicates the close link between online hate speech and the perpetration of offline hate crimes (Hatzipanagos, 2018; Williams, 2019), meaning that this trend is a cause for concern.

- Significantly, unlike trend one, wherein the abuse can appear subtle or undetected, the hate speech contained in reports to RHC was much less ambiguous (e.g. direct slurs, negative stereotypes and threats inciting violence). This is worrying as it indicates a lack of action taken on the behalf of social media platforms who should, in theory, be removing this type of content proactively or, at the very least, positively responding to user reports. These findings thus support the recognition that social media companies need to develop more consistent and rigorous policies for identifying and responding to hateful conduct (Williams & Mishcon de Reya, 2019). It also supports recommendations which outline the need for an independent regulator to enforce this duty of care (Department for Digital, Culture, Media and Sport, 2019; Williams & Mishcon de Reya, 2019).

- The third trend identified was the spike in reports towards the end of the year involving 'perpetrator-less harms' e.g. clients inadvertently viewing graphic/violent content and/or pornography. The growing increase in these types of reports is encouraging, given that it demonstrates the broadening reach of the RHC service. Nevertheless, due to the high proportion of this type of content being located on independently owned and moderated sites, RHC



to directly action its removal. If reports of this nature continue to grow, it might become necessary for the RHC service to introduce additional tools to deal with this demand. The service might want to look towards building relationships with a broader range of platforms or, alternatively, creating a resource pack outlining comprehensive advice for dealing with this type of content.

- In addition to analysing the way in which existing trends can be used to inform policy and practice, the identification of absent harms can also be used towards this end. As discussed in the introduction, when RHC was initially developed, it outlined eight types of online harm which could be dealt with by the service. These eight harms were defined in relation to SWGfL's experience in operating the Revenge Porn and Professionals Online Safety Helplines, alongside close analysis of community guidelines, across a number of social media platforms. Prior to the launch of the service, practitioners had an expectation that reports concerning each of these eight harms would be relatively evenly split. As it stands, RHC dealt with a significantly lower proportion of reports concerning unwanted sexual advances (1 case) and self-harm/suicide content (2 cases). RHC practitioners are aware that both of these harms are significant online issues; as such, the low proportion of

reports concerning unwanted sexual advances and self-harm/suicide content is an area of interest. It is useful to speculate on the reasons for these absences, so as to inform service development and policy.

Regarding the absence of reports concerning unwanted sexual advances, research suggests that this type of gender-based abuse is drastically normalised and minimised by women, meaning that victims often do not identify it as an issue (Amnesty International, 2018; Fawcett Society, 2017). Even when this abuse is identified, there is evidence to suggest that poor or inconsistent responses from social media platforms lead women to feel that there is little point in reporting it (Amnesty International 2018). Further to this, significant costs can be associated with drawing attention to unwanted sexual advances, as evidenced by various high profile cases (Elgot, 2015; Webber, 2017).

When it comes to self-harm/suicide content, encouragingly figures suggest that social media platforms are becoming more adept at proactively removing this type of content (Facebook Transparency, 2019; Hicks & Gasca, 2019, Mosseri, 2019). Nevertheless, evidence also indicates that this type of content isn't immediately obvious on public social media platforms but is, instead, hidden under secret and

ever evolving hashtags (Moreno et al, 2016; Scherr et al, 2019). These hashtags are actively sought out in private groups by those who suffer with these issues. As a result, this type of content is much less likely to be viewed as a problem or reported for removal by those coming into contact with it.

*All of the above might go some way towards explaining the low instance of reports regarding unwanted sexual advances and self-harm/suicide content to RHC. Social media platforms, policy makers and support services alike all have a responsibility to address unwanted sexual advances and self-harm/suicide content online. Towards this end, RHC might look towards developing awareness-raising campaigns and resource packs to draw attention to these issues in order to better support victims.*

## Tackling emerging issues

- **The statistics concerning the detrimental mental health impacts of viewing harmful content online, alongside the broad range of mental health impacts experienced, are worrying.** The correlation between poor mental health and harmful online content is one of the largest areas of concern, not only for RHC, but for platforms, policy makers, other support agencies and researchers (Shakya & Christakis, 2017; Department for Digital, Culture, Media and Sport, 2019). Mental health online is not an issue which a service like RHC can tackle in isolation. In fact, it is often the case that, once RHC is contacted, significant damage has already been done. Nevertheless, RHC does have a role to play, not only in providing emotional support and making onward referrals, but in reporting trends and highlighting impacts, such as has been done in this report. Reports such as these ensure that these issues are understood in their complexity. They also ensure that mental health online stays at the top of the agenda for those in positions to address it.
- The data regarding the back and forth referrals between the service and the police is also concerning, not only because of the amount of time expended on this process by practitioners but, crucially, because of the impact

on clients' wellbeing. Unfortunately, this data reflects previous research which indicates a lack of police understanding surrounding online crime and/or an unwillingness to take action, particular in instances of interpersonal abuse (Bond & Tyrrell, 2018; Smith, 2016; Sharratt, 2019). Research also points to the detrimental impact of this on victims' mental health and safety (McGlynn, Rackley & Houghton, 2017; Sharratt, 2019). These findings thus support broader recommendations regarding the need for better training of law enforcement on issues of online crime and abuse (Bond & Tyrrell, 2018; Home Office, 2018). In order to have a more direct impact, the RHC service might look towards developing a resource pack, designed to make police aware of their responsibilities and duties as law enforcement vs the boundaries of the RHC service.

- The additional cultural and religious issues experienced by a proportion of RHC clients is a further area needing address. Looking towards the future, the RHC service might benefit from engaging in training swaps with cultural support services, so as to build stronger working partnerships and enhance client support. These types of cases also raise questions regarding

how online harms are defined and whether those definitions can be universally applied. The findings of this report thus support the recent recognition that more nuanced definitions of private/intimate material need to be developed (Law Commission, 2018). Encouragingly, the Law Commission is currently undertaking such a review, upon which RHC practitioners have consulted.

- The potential for protected characteristics to clash in online content is ever present and is a difficult area to navigate. Individuals are always going to hold contrasting views and the challenge lies in finding a way to accommodate a range of opinions in a way that is harm-free and respectful. The growth in reports concerning a clash of sex and gender reassignment has, undoubtedly, been compounded by the recent proposed reforms to the Gender Recognition Act (Government Equalities Office, 2018). Debates concerning how these issues can best be addressed so as to safeguard everyone are still unresolved and attract heated debate. It is the hope of RHC practitioners that, as policy regarding gender recognition is clarified, this will trickle down to platform guidance and public understanding.

- Finally, improving relationships with platforms will, in all likelihood, need a multi-pronged approach, given the range of issues identified. The issue regarding inconsistency of industry partners when responding to harmful content has, unfortunately, been identified elsewhere, both by users and moderators (Gillespie, 2018; Matsakis, 2018; Notopoulos, 2019). In these instances, it is easy to point the finger of blame at social media companies, however, as those working within online safety know, platforms are tasked with moderating a huge volume of content. Alongside this, they must balance their users' need for safety vs their right to privacy, a task which will always be subject to error. This is not, however, to say that platforms have nothing to learn from the identification of this issue. Industry content moderators are encouraged to keep abreast of the emerging trends and issues reported by services such as RHC so as to enhance the ways in which they deal with user reports. Alongside this, policy and safety teams are advised to educate themselves on issues specific to their platforms so to ensure their community guidelines are clear, consistent and fit for purpose.

**Regarding the disparity in volume of reports across different platforms:** to some extent this can be seen as proportional to the amount of users across various platforms. This explanation doesn't, however, fully account for such a stark difference and thus the minimal reports received from platforms such as TikTok, Snapchat and many of the Microsoft and Google services is an area warranting further enquiry. It may be the case that these platforms are simply better at responding to reports of harmful content without the need for mediation. Less optimistically, it may be that positive cultures of reporting, such as those that exist on Facebook and Instagram, are yet to develop on these less established platforms. For example, users of services such as TikTok, Roblox and Snapchat are generally younger (Khoros, 2020; LSE, 2018) and there is evidence to support the fact that this age group view harmful content online as normal and inevitable (Lavis, 2016; Marchant, Hawton, Stewart, Montgomery, & Singaravelu, 2018). The solution here thus lies in research, education and greater awareness raising.



When it comes to the increasing volume of reports received from platforms with which RHC do not currently have relationships, the solution is more straightforward. RHC should continue to track the data on these types of reports, with the view to developing relationships with a broader range of platforms. Where relationship development is not possible (e.g. due to a lack of will on the behalf of platforms), it is recommended that the RHC service design a tailored information and advice pack to provide to clients.



# Conclusion

**This report has presented results from mixed-methods research carried out on all cases dealt with in the first year of RHC's operation (January 2019-December 2019).**

It has discussed top level statistics concerning service response, nature of cases and client demographics. It has also identified a number of emerging trends, issues and challenges experienced by the service.

The trends and emerging issues identified in this report were discussed, both in terms of service evaluation and, more broadly, in terms of policy and legislation. Encouragingly, policy makers and industry platforms

appear to be abreast of some of these trends. However, there is an evident need for more work to be done to ensure that online harms and their implications are tackled and that individuals, particularly those that are vulnerable or belonging to a minority group, are protected.

There are some limitations to this report. Whilst it discussed age and gender and, to a degree, analysed cases where clients self-identified as Muslim, it was unable to analyse the way in which other structural oppressions, such as race or sexuality, might intersect with various online harms. Future reports might look to include this type of analysis, however, this might not be feasible, given that this data is not routinely collected on RHC reporting

forms. Towards this end, readers of this, and future RHC reports, are encouraged to bear in mind that the range of identities experienced by Internet users might not be fully represented.

What has emerged throughout this report is the increasing popularity of the RHC service, alongside the multitude of areas where the service can be developed so as to respond to this growth. To this end, there is an evident need to secure additional partnerships, greater government and industry support and, crucially, further sources of funding in order to meet expanding service demands.



Amnesty International. (2018). #ToxicTwitter: Violence and abuse against women online. Amnesty International. Retrieved from <https://www.amnesty.org/en/documents/act30/8070/2018/en/>

Department for Digital, Culture, Media and Sport (DCMS). (2019). Online harms white paper. Department for Digital, Culture, Media and Sport. Retrieved from <https://www.gov.uk/government/consultations/online-harms-white-paper>

Dobash, R. & Dobash, R. (2004). Women's violence to men in intimate relationships. Working on a Puzzle. British Journal of Criminology, 44(3), 324-349.

Drouin, M., Ross, J. & Tobin, E. (2015). Sexting: A new, digital vehicle for intimate partner aggression? Computers in Human Behaviour, 50, 197-204.

Elgot, J. (2015, September 10). Law firm partner hits out at female lawyer after LinkedIn sexism row. Guardian UK. Retrieved from <https://www.theguardian.com/technology/2015/sep/10/law-firm-partner-says-no-more-briefs-for-linkedin-sexism-row-barrister>

Government Equalities Office (2018). Reform of the gender Recognition Act 2004. Government Equalities Office. Retrieved from <https://www.gov.uk/government/consultations/reform-of-the-gender-recognition-act-2004>

The Equality Act. (2010). Retrieved from <http://www.legislation.gov.uk/ukpga/2010/15/contents>

Facebook Transparency. (2019). Community standards enforcement report. Facebook Transparency. Retrieved from <https://transparency.facebook.com/community-standards-enforcement#suicide-and-self-injury>

Fawcett Society. (2017). Twitter is 'failing women' experiencing online threats and harassment. Fawcett Society. Retrieved from <https://www.fawcettsociety.org.uk/News/twitter-failing-women-experiencing-online-threats-harassment>

Hatzipanagos, R. (2018). How online hate turns into real life violence. Washington Post. Retrieved from <https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/>

Hicks, D. & Gasca, D. (2019). A healthier Twitter: One year in. Twitter. Retrieved from [https://blog.twitter.com/en\\_us/topics/company/2018/healthupdate.html](https://blog.twitter.com/en_us/topics/company/2018/healthupdate.html)

Home Office. (2019). Domestic abuse bill 2019: Factsheets. Home Office. Retrieved from: <https://www.gov.uk/government/publications/domestic-abuse-bill-2019-factsheets>

Khoros. (2020). The 2020 social media demographics guide. Khoros resources. Retrieved from <https://khoros.com/resources/social-media-demographics-guide>

Lavis, A. (2016). Social media and anorexia: A qualitative analysis of #proana. Education and Health, 34(2), 57-62.

Law Commission. (2018). Abusive and offensive online communications: A scoping report. Law Commission. Retrieved from <https://www.lawcom.gov.uk/abusive-and-offensive-online-communications/>



LSE (London School of Economics and Political Science). (2018). Social Media Platforms and Demographics. LSE Digital Communications Team. Retrieved from <https://info.lse.ac.uk/staff/divisions/communications-division/digital-communications-team/assets/documents/guides/A-Guide-To-Social-Media-Platforms-and-Demographics.pdf>

Marchant, A., Hawton, K., Stewart, A., Montgomery, P. & Singaravelu, V. (2018). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good the bad and the unknown. PLOS ONE, 13(3), e0193937.

Moreno, M., Ton, A., Selkie, E. & Evans, Y. (2016). Secret society 123: Understanding the language of self-harm on Instagram. Journal of Adolescent Health. 58, 78-84.

Moserri, A. (2019). Changes we're making to do more to support and protect the most vulnerable people who use Instagram. Instagram Blog. Retrieved from <https://about.instagram.com/blog/announcements/supporting-and-protecting-vulnerable-people-on-instagram>

Office for National Statistics (ONS). (2019). Domestic abuse prevalence and trends, England and Wales, year ending March 2019. Office for National Statistics. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseprevalenceandtrendsenglandandwales/yearendingmarch2019>

Refuge. (2020, January 9). 72% of Refuge service users identify experiencing tech abuse. Refuge. Retrieved from <https://www.refuge.org.uk/72-of-refuge-service-users-identify-experiencing-tech-abuse/>

Scherr, S., Arendt, F., Frissen, T., & Oramas, J. (2019). Detecting intentional self-harm on Instagram: Development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. Social Science Computer Review.

Shakya, H. & Christakis, N. (2017). Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. American Journal of Epidemiology, 185(3), 203-211.

Thacker, F. (2017). Understanding cyberviolence in intimate partner relationships: A qualitative study. Thesis, Georgia State University. Retrieved from [https://scholarworks.gsu.edu/iph\\_theses/563/](https://scholarworks.gsu.edu/iph_theses/563/)

Webber, J. (2017, October 16). Anita Sarkeesian: 'It's frustrating to be known as the woman who survived #Gamergate'. Guardian UK. Retrieved from <https://www.theguardian.com/lifeandstyle/2017/oct/16/anita-sarkeesian-its-frustrating-to-be-known-as-the-woman-who-survived-gamergate>

Williams, M. (2019). The connection between online hate speech and real world hate crime. Oxford University Press blog. Retrieved from <https://blog.oup.com/2019/10/connection-between-online-hate-speech-real-world-hate-crime/>

Williams, M. & Mishcon de Reya. (2019). Hatred behind the screens: A report on the rise of online hate speech. Mishcon Academy. Retrieved from <https://www.cardiff.ac.uk/news/view/1729524-online-hate-speech-report-launched>

Women's Aid. (2019). Domestic abuse is a gendered crime. Women's Aid. Retrieved from <https://www.womensaid.org.uk/information-support/what-is-domestic-abuse/domestic-abuse-is-a-gendered-crime/>



**REPORT  
HARMFUL  
CONTENT**

[reportharmfulcontent.com](https://reportharmfulcontent.com)



Co-financed by the Connecting Europe Facility of the European Union