How people are harmed online: testing a model from a user perspective

Ofcom

About Revealing Reality

Revealing Reality is an independent social research agency, working with regulators, government and charities to provide insight into people's online behaviours and experiences.

Studying how the digital world is shaping people's behaviours is something we do every day. We regularly conduct detailed qualitative behavioural research, observing how people really use digital products, services and technology. This includes exploring how digital design shapes behaviour – across technology, gambling, financial products, the health service, and more.

Contents

About Revealing Reality 2
Ofcom foreword 4
Executive summary 6
Context for the research 10
Introduction 10
Why Ofcom is developing a model for how harm manifests 10
Background to the research 11
How research was used to develop and validate the model from a user perspective 11
Project objectives 11
Methodology 11
Overview of the model 13
Introducing the model 13
Overview of the model 13
Research findings 16
How online experiences led to harm 16
People experienced different routes to harm 16
Overview of the routes to harm 16
What harms were experienced 23
People experienced a range of types and severities of harm 23
Why people experienced harm 27
The likelihood of harm was affected by a range of risk factors 27
Risk factors associated with the content / contact 28
Risk factors associated with the characteristics of the distributor 32
Risk factors associated with the fixed characteristics of the user 35
Risk factors associated with the circumstances of the user 37
Risk factors associated with what else the user is exposed to or does online 40
Risk factors associated with wider societal context 42
Risk factors associated with the characteristics of the platform 43
Risk factors associated with the actions of other users 48
Conclusions 50
Considerations for the model based on this research 50
The model focuses on psychological and physical harm, and could do more 50
The model works equally well for legal and illegal content online 50
The model works well for isolated routes to harm, and can also account for cumulative routes 50
The model does not account for indirect routes to harm 50
The model does not imply which risk factors are most significant 51
The model does not account for the positive outcomes that occur because of online experiences 51
Research gaps and future opportunities 52
Annex: Case studies 53
Annex: Participant overviews 81
Annex: Methodology 88

Ofcom foreword

Ofcom has a statutory duty to promote and research media literacy. A key way we seek to fulfil this duty is through our *Making Sense of Media* programme, which aims to help improve the online skills, knowledge and understanding of children and adults in the UK. Ofcom was also given powers in autumn 2020 to regulate UK-established video-sharing platforms (VSPs). And in December 2020, the Government confirmed its intention to nominate Ofcom as the regulator for online safety in the UK, under the Online Safety Bill, which is currently in Parliament.

As referenced in our Roadmap to Online Safety Regulation, this report is one in a series of research studies into online safety that will inform our preparations for implementing the new online safety laws. As part of these preparations, we are building a robust evidence base, bringing together internal and external data, collected using different methods, from a variety of different sources.

In this context, this programme of research further develops our understanding of online harms and how we can help to promote a safer user experience. The findings should not be considered a reflection of any policy position that Ofcom may adopt when we take up our role as the online safety regulator.

How Harms Manifest

The Online Safety Bill, as currently drafted, will require Ofcom to assess, and publish its findings about the risks of harm arising from content that users may encounter on in-scope services, and will require in-scope services to assess the risks of harm to their users from such content, and to have systems and processes for protecting individuals from harm.

Online users can face a range of risks online, and the harms they may experience are wide-ranging, complex and nuanced. In addition, the *impact* of the same harms can vary between users. In light of this complexity, we need to understand the mechanisms by which online content and conduct may give rise to harm, and use that insight to inform our work, including our guidance to regulated services about how they might comply with their duties.

This report sets out a generic model for understanding how online harms manifest. This research aimed to test a framework, developed by Ofcom, with real-life user experiences. We wanted to explore if there were common risks and user experiences that could provide a single framework through which different harms could be analysed. There are a couple of important considerations when reading this report:

- The research goes beyond platforms' safety systems and processes to help shed broader light on what people are experiencing online. It therefore touches on issues that are beyond the scope of the proposed online safety regime.
- The research reflects people's views and experiences of their online world: it is based on people self-identifying as having experienced 'significant harm', whether caused directly or indirectly, or 'illegal content'. Participants' definitions of harmful and illegal content may differ and do not necessarily align with how the Online Safety Bill, Ofcom or others may define them.

This research represents a deep dive into the factors in the model which are visible and known to users. Having demonstrated the model is effective in capturing these, we also need to understand the factors that are not visible or knowable to a user, many of which are indicated in this report as areas for further study.

Trigger warning

This report contains content or experiences, or allusions to topics that some people may find distressing, including:

- Abusive and threatening comments and posts
- Gore content
- Cyberflashing / unsolicited sharing of nudes
- Intimate image abuse
- Sexual violence
- Physical violence
- Misogynistic / sexist content
- Homophobic and transphobic content
- Eating disorders
- Financial scams
- Suicide and self-harm content
- Abusive content related to ethnicity or religion

Throughout the report the warning sign (left) is used to serve as a trigger warning.

Executive summary

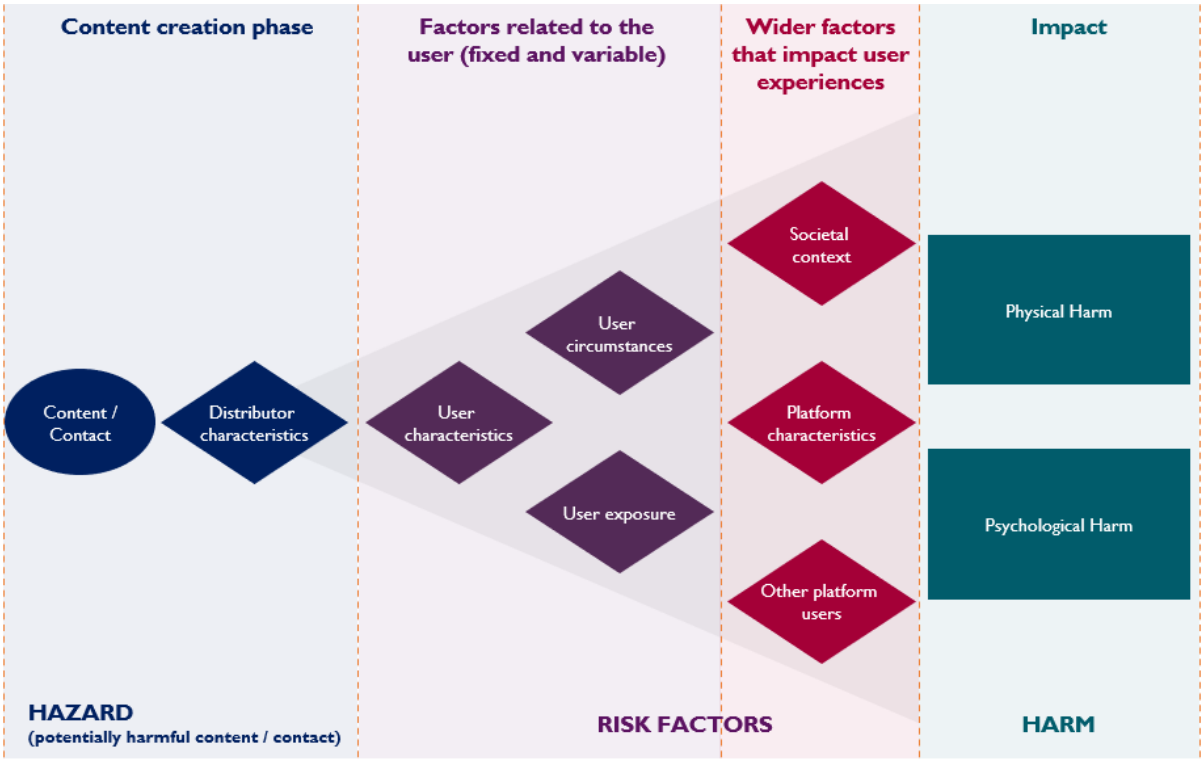
Ofcom has created a generic model to seek to explain how and why harm can originate from online experiences and content:

Online harm is a complex and nuanced issue. Ofcom is in the process of carrying out research and developing its evidence base to understand, amongst other things, the mechanisms by which online content and conduct may give rise to harm, and the impact of online harm on users. One of the ways Ofcom has approached this has been to work up a general framework that differentiates cause, effect and risk factors:

- Hazards (potentially harmful content / contact): The stimuli that are a potential source of harm – e.g. images, videos, messages;
- Risk factors: Things that change the likelihood that a hazard will cause harm to an individual;
- Harm: The negative consequence for someone resulting from a hazard combined with risk factors.

Not all potentially harmful content / contact will go on to cause harm, and the role played by risk factors is crucial to understanding how and why online harms do or do not occur. Ofcom has developed a model that seeks to categorise and describe these risk factors and determine their role in why harm does or does not occur.

Figure 1: Ofcom's model for how harms manifest



Ofcom commissioned this research to test, iterate and validate this model. This was done by exploring the experiences of real people to test whether the model was sufficient to capture that experience and the way in which people come to harm online. It also aimed to map the corresponding risk factors that played a role in harm occurring.

A qualitative research approach, employing in-depth individual interviews, was chosen for this project to enable an unhurried, discursive conversation with each participant in a private environment. (Owing to Covid-19 restrictions at the start of the research process, all interviews were conducted by video-call to protect participants and researchers from any risk of infection.)

The total number of participants was 55 adults, the majority of whom were recruited because they self-identified as having experienced harm as a result of online experiences. A smaller control group was recruited because they reported similar online experiences (i.e. exposure to hazards), but did not feel they had been harmed.

Researchers explored their experiences, mapping what had happened, what they had been exposed to and how, and all the contextual and situational factors involved. The goal was to build up a detailed and nuanced picture for why, in each specific situation, harm had (or had not) occurred.

There are some limitations to this approach. Some people may not know that they have been harmed – e.g. someone whose actions have been shaped by disinformation that they did not identify as such. Others may know they have been harmed but might not be aware of all the reasons why – e.g. an algorithm shaped what they saw, or their own low media literacy affected their experience – and so some risk factors are less likely to emerge in this type of research.

Summary of key findings:

- There are multiple routes to experiencing harm as a result of online experiences. In some cases, an isolated experience of a hazard (e.g. a fraudulent advert) can lead to harm straight away or potentially after a delay. In others, cumulative exposure to multiple hazards over time can cause harm – and this can either be through passive exposure (constantly seeing a high concentration of a particular body type or content that makes certain unrealistic body shapes aspirational), or active, self-reinforcing engagement with hazards (e.g. engaging actively in pro-anorexia communities online). In some cases, harm was seen to occur to people indirectly as a result of *others* being exposed to hazards relating to them (e.g. nude images of someone being shared without their consent).
- Harms resulting from an *isolated* incident that has an *immediately* harmful impact appear to be much easier for respondents to recognise. There is therefore a risk that single, discrete pieces or genres of content receive disproportionate focus when thinking about what leads to harm. However, the research identified that the cumulative route to harm appeared to cause the most severe harm experienced by respondents.
- Across these routes, a wide range of risk factors were documented that appeared to make harm more or less likely to occur. Some risk factors did so by making it more likely that someone would be **exposed to a hazard**. Others made it more likely for exposure to a hazard to **go on to cause harm**. Some did both.

All the risk factors identified were able to be mapped against a component of the model, although there are significant overlaps and inter-dependencies between them in each individual's experience. Each risk factor is usually exacerbated by others, and none can be assumed to always, in all cases, increase the risk of harm.

Media literacy also has an important role to play in online safety, and elements of this are apparent throughout the model, although not explicitly identified by the users themselves.

Summary of risk factors identified:

Risk factors associated with the content / contact

Attributes of the harmful content / contact itself, such as, the content of a message, the nature of an image, or the way in which it was experienced by a user. For example this includes if the content / contact is geographically local or how expected the contact/content is.

Risk factors associated with the characteristics of the distributor

Characteristics of the person or account who distributed the content or was the source of the interaction. This includes, for example, if the distributor was known to the user.

Risk factors associated with the fixed characteristics of the user

Fixed characteristics are those that the user can't change, including some protected characteristics.

Risk factors associated with the circumstances of the user

User circumstances, compared to characteristics, could theoretically change or be changed over time. This includes, for example, the mental health of the user or their social support network.

Risk factors associated with what else the user is exposed to or does online

User exposure can be split into two categories: what other content / contact users are exposed to, and what they do in response to seeing content.

Risk factors associated with wider societal context

These factors were observed to make it more likely that people would be exposed to content / contact which may lead to harm, such as trending topics.

Risk factors associated with the characteristics of the platform

The design and set-up of the platform's contribution to risk, including popularity indicators.

Risk factors associated with the presence and actions of other users

The factors that were observed to increase the potential that people would be exposed to content, and/or experience harm, if other risk factors were also present.

The research confirmed the applicability of the model and identified a number of considerations for its use:

1. The model focuses on **psychological** and **physical** harm as these are explicitly referenced in the definitions of 'harm' included in the current version of the Online Safety Bill. However, other types of harm outside the scope of the duties in the current Bill including economic (e.g. financial loss through scams, loss of productivity) appeared in the research and could also be accounted for by the model. Societal harm is also likely to have occurred but is harder for research participants to self-report – although this could conceivably be accounted for in the model if required.
2. The research included people who had experienced harm from being exposed to **legal** content, as well as content the research team considered likely to be **illegal**. The model appears to account for the impact of both legal and illegal content.
3. The model works well for **isolated** routes to harm by describing the nature of the isolated hazard/s in the content / contact category. The model can also account for **cumulative** routes to harm through the "user exposure" category of risk factors (e.g. acknowledging the other experiences and actions of the user on the platform) although it is potentially less intuitively structured for this route. Given that the cumulative route was often seen to lead to more severe harm than isolated experiences, this is worth Ofcom taking into consideration when using the model.
4. The model sought to explain how exposure to hazard/s can lead to an individual experiencing harm, and therefore does not account for the **indirect** route to harm.

5. The model does not seek to imply which factors are most impactful or more prevalent and only includes factors seen across different experiences of harm. However, this evidence base could be built in the future using the model as a framework. It is important to note that different factors will be relevant for the different routes to harm, and potentially for different content genres as well.
6. The model was created to show how 'significant harm' occurs. It does not account for the potential positive outcomes that occur because of online experiences.

Context for the research

Introduction

Why Ofcom is developing a model for how harm manifests

Online harm is a complex and nuanced issue. Ofcom is in the process of carrying out research and developing its evidence base to understand, amongst other things, the mechanisms by which online content and conduct may give rise to harm, and the impact of online harm on users. One of the ways Ofcom has approached this has been to work up a general framework that differentiates cause, effect and risk factors:

- Hazards (potential harmful content / contact): The stimuli that are a potential source of harm – e.g. images, videos, messages;
- Risk factors: Things that change the likelihood that a hazard will cause harm to an individual;
- Harm: The negative consequence for someone resulting from a hazard combined with risk factors.

Not all potentially harmful content / contact will go on to cause harm, and the role played by risk factors is crucial to understanding how and why online harms do or do not occur. Ofcom has developed a model that seeks to categorise and describe these risk factors and determine their role in why harm does or does not occur.

Ofcom set out with the ambition to explore whether it was possible to create a generic model that can categorise and describe these different components – hazards, risk factors, and harms – to understand how and why harm manifests.

What must the model be able to do?

Ofcom is seeking to use this model to help understand the mechanisms by which online content and conduct may give rise to harm, and use that insight to inform its work, including its guidance to services in scope of the Online Safety Bill about how they might comply with their duties. It wanted to discover if there were common risks and user experiences that could provide a single framework through which different harms could be analysed. To be useful for understanding how harms manifest online, the model must:

- Make clear what needs to be taken into consideration when thinking about the risk of harm occurring on a platform.
- Be clear and intuitive so that it can be easily picked up and used as a tool for different purposes.
- Be robust and work for most harms (not necessarily all harms), allowing for flexibility to be applied to different types of online harm where possible.
- Make clear what can be measured.

What does the model not do?

The focus of the model is on how *significant harm* manifests online. It therefore covers the range of different types of and routes to harm that people experience online, which are detailed in this report. Because of its focus on how harm manifests, the model does not cover the potential corresponding **benefits** of online content or contact, or **mitigations** that could be used to reduce the risk of harm.

Research was needed to test, iterate and validate the model

The model was developed based on existing knowledge and hypotheses of the factors that might be important to explain online harm. These were ordered by a) the distinctions of potentially harmful content or contact/risks/harms and b) intuitive groupings and categories. To test this structure and map what factors might be important to include within it, Ofcom commissioned research to explore the real-world experience of online harm from a user perspective.

Background to the research

How research was used to develop and validate the model from a user perspective

Project objectives

Ofcom has had media literacy duties since Ofcom was created in 2003 and is currently the regulator for video-sharing platforms (VSPs) established in the UK. Ofcom is also preparing for new responsibilities as the regulator for online safety in the UK.

In this research Revealing Reality were asked to explore online harms at a broad level, as a complement to Ofcom's suite of research on online harms, which includes the Online Experiences Tracker. The aim was to understand whether it is possible to have a generic model, which captures common risk factors which work for different types of online harm and for different routes to harm.

Specific project objectives for this research were to:

- Explore the diverse experiences of people who report experiencing harm because of online experiences, and those who have been exposed to content / contact but who do not report experiencing harm. This includes exploring the types of content / contact they interacted with and the risk factors at play when the harm occurred.
- Test Ofcom's draft model of how harm manifests against these experiences.
- Provide recommendations on how the model can better reflect all experiences of online harm.

Methodology

A qualitative research approach was chosen as the most suitable method for meeting the objectives of this project. The aim of qualitative research is to hear and/or observe the experiences, attitudes and behaviours of carefully-selected individuals in more depth than can be achieved by putting a set of fixed questions to large numbers of people in quantitative surveys.

For this project, individual interviews were employed to enable a discursive conversation with each participant in a private environment[1]. A sample of 55 adults was selected, the majority recruited on the basis that they believed they had experienced severe or significant harm because of online experiences within the last six months. The experience of harm is subjective, and this work has taken the individual's perspective as a starting point, interviewing people who self-define as being significantly harmed by something online. Equally, demonstrating a clear causal link between online experiences and harm is difficult, and this research relies on people's perceptions of what caused the harm they experienced.

The sample was therefore intentionally focussed on people who had had bad experiences online and was not reflective of the general population. The requirement for the exposure to be recent was included to ensure that participants could still recall the experience in detail—especially as, to minimise the risk of further harm, participants were not asked to show us the content / contact during the interview, meaning they had to rely on recall.

Of the 55 individuals:

[1] Owing to Covid-19 restrictions at the start of the research process, all interviews were conducted by video-call to protect participants and researchers from any risk of infection.

- 36 people reported they had experienced significant harm from legal online content / contact.
- 10 people reported they had experienced significant harm from online content / contact believed by the research team to be illegal.
- 9 people reported they had experienced exposure to similar content / contact but were unharmed – this group was recruited as a control group.

Recruitment was focused on ensuring that participants across the sample had experienced a broad range of content / contact, including:

- Content / contact attacking protected characteristics.
- Content / contact of a sexual nature.
- Content / contact selling or promoting illegal behaviour/objects e.g. weapons, drugs.
- Eating disorder content.
- Glamourised lifestyle content.
- Gore and violent content.
- Online abuse and harassment.
- Online fraud and scams.
- Suicide content[2].

Similarly, recruitment sought to include a range of types of harms. This included both physical and psychological harms which were reported as 'significant' by participants.

Other factors considered during recruitment to ensure a varied sample included participants' online activity (e.g. types of online activity, devices used, time spent online) and demographic criteria (e.g. household income/social grade, age, gender, ethnicity, accessibility needs, sexuality, religion, and region within the UK). These factors were of secondary importance to ensuring a range of online experiences and harms were accounted for in the sample.

All participants took part in an in-depth, remote interview lasting approximately 1-2.5 hours. Interviews took a semi-structured approach, covering the following key areas of discussion:

- Participant background.
- Online habits and device use.
- Overview of the content / contact experienced.
- Mapping the experience through each component of the model.

Note that because the recruitment criteria required people to be aware they had been harmed (except in the control group), people who might have been subject to mis/disinformation were not able to be included in the sample. A different study design will be required to understand how harms manifest for this audience, given they may not self-identify as being harmed.

Throughout the recruitment and interview process, thorough ethical and safeguarding procedures were followed.

For a more detailed breakdown of the project methodology, please refer to the annex.

This research complements other ethnographic research that Ofcom is carrying out on the risk factors that may lead children to experiencing harm online. That project collected data from 40+ young people aged 8-17 about their experiences online. Some of the data from the research with children[3] has been referenced in this report to illustrate the effect that some harms can have across age groups, and to demonstrate that the model also works for the experiences of children.

---

[2] Note: experiences of terror-related content and child sexual abuse material (CSAM) was not represented in the sample. People in the sample who experienced exposure to illegal content / contact included death threats, content assisting suicide, illegal harassment, racist abuse, homophobic abuse, online scams, cyberstalking and intimate image abuse.

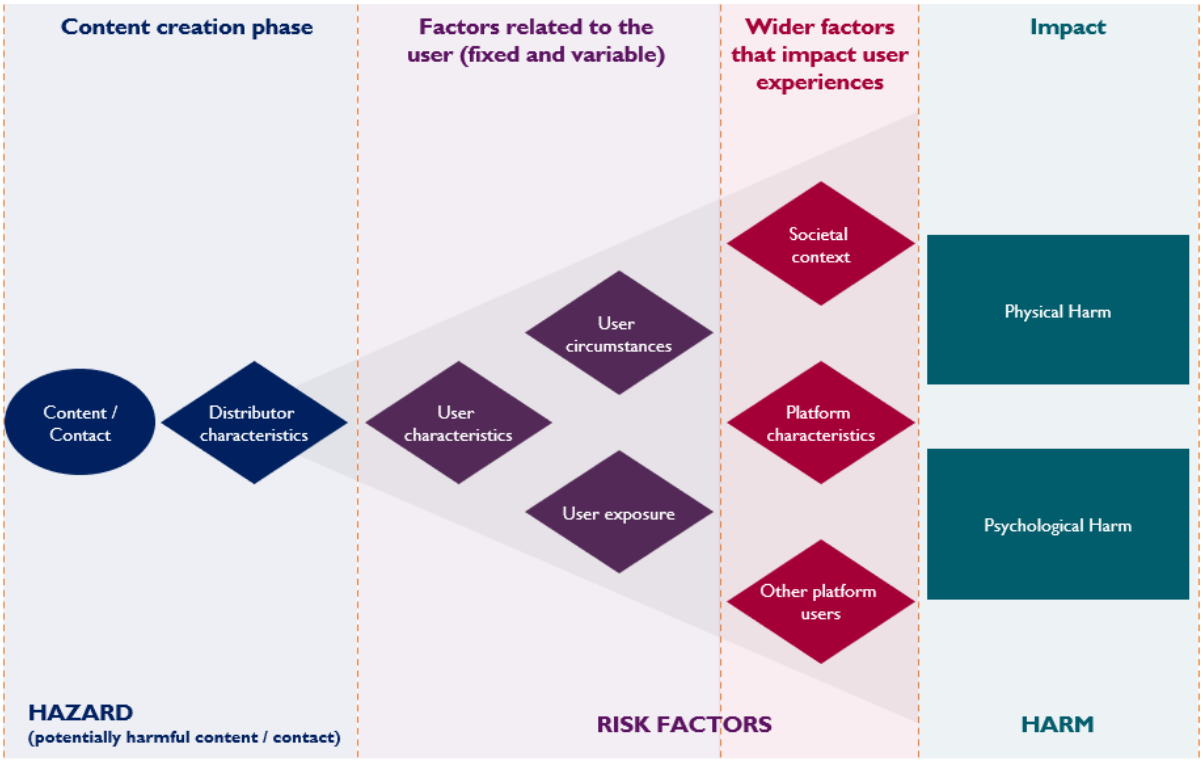
[3] Findings from this research will be published later on in 2022.

Overview of the model

Introducing the model

This chapter gives an overview of the different components of the model. The model tested in the research was originally developed by Ofcom. and was further tested and developed throughout this research. This research focusses on the factors that users are able to self-report. Additional components are noted which Ofcom suggested may be relevant to explore in this research but which this research could not evidence due to the known limitations of the methodology.

Overview of the model



The model is a simplified representation of how harm manifests as a result of online experiences

This model was originally created by Ofcom to explore the possibility of whether a common framework for understanding how harms manifest online harm could be developed. The intention was therefore that individual harms should fit within the model, recognising there will be nuances by type of harm. The model does not attempt to place a weight on the different factors, and the relative importance of each is likely to vary depending on context and the interaction between one risk factor and another- which is not the same in all cases.

The research included in depth interviews with those who had reported experiencing harm from being exposed to legal content, as well as content the research team considered likely to be illegal. Whilst the requirements on platforms may be different for legal and illegal content, it is worth noting that in terms of harm experienced, the model works equally well within both the legal but harmful and the illegal categories.

The model aims to work across a range of content types. This report has focused primarily on harm which has occurred through an adult's interactions with **content** and **contact,** which is why the model explicitly refers to these. However, examples of **conduct** and **contract** which led to harm were also seen across the sample and successfully fitted within the model. Cases of conduct were seen in stories from people who had been led

to harm by following the actions of people around them, for example, after spending significant amounts of time in online forums or communities. Cases of contract were seen in stories from people who had fallen victim to scams.

This distinction is taken from the CO:RE (Children Online: Research and Evidence) framework[4] which categorises the hazards children can face online, but which may also be helpful to understand adult experiences:

- **Content** (child as recipient) – where a child engages with or is exposed to potentially harmful content. This can be violent, gory content, hateful or extremist content, as well as pornographic or sexualised content that may be illegal or harmful, including by being age inappropriate.
- **Contact** (child as participant) – where a child experiences or is targeted by contact in a potentially harmful adult-initiated interaction, and the adult may be known to the child or not. This can be related to harassment (including sexual), stalking, hateful behaviour, sexual grooming, sextortion or the generation or sharing of child sexual abuse material.
- **Conduct** (child as actor) – where a child witnesses, participates in, or is a victim of potentially harmful conduct such as bullying, hateful peer activity, trolling, sexual messages, pressures or harassment, or is exposed to potentially harmful user communities (e.g. self-harm or eating disorders). Typically conduct risks arise from interactions among peers, although not necessarily of equal status.
- **Contract** (child as consumer) – where a child is party to and/or exploited by potentially harmful contract or commercial interests (gambling, exploitative or age-inappropriate marketing, etc.). This can be mediated by the automated (algorithmic) processing of data. This includes risks linked to ill-designed or insecure digital services that leave the child open to identity theft, fraud or scams. It also includes contracts made between other parties involving a child (trafficking, streaming child sexual abuse).

There are many risk factors which appear to increase the likelihood of harm occurring

The model sets out how the potential for harm can be affected by an array of different risk factors. Some of these risk factors increase the likelihood that a user will be **exposed** to content / contact (e.g. the distributor is known to the user, therefore the user is more likely to see what they post) - either for the first time, or repeatedly. Some risk factors increase the likelihood for content / contact to **lead to harm** (e.g. the content feels personally targeted or relevant). Some risk factors do both.

The presence of one of these factors in isolation does not indicate that harm will occur, but a combination of risk factors can work together to make the potential for harm more likely. For most of the factors, we found people in the counterfactual (control) group for whom the factor was also present but harm did not occur.

Additional components of the risk factors are likely to be important, but not apparent to users.

It is likely that there are risk factors that could be important but that are difficult to evidence using this research. Some are factors that research participants will inherently find difficult to self-report, e.g. the impact of an algorithm on what they are served, or their own media literacy. There are others that appeared relevant in individual cases and may impact many other people but are very context dependent, e.g. other protected characteristics of the user such as their ethnicity or sexuality. These appeared as relevant to specific research participants in particular cases where the experience related directly to this characteristic – e.g. racist or homophobic abuse. These are mentioned in the 'user characteristics' section of the 'Why people experienced harm' chapter.

[4] 4 Cs of online risk: short report & blog on updating the typology of online risks to include content, contact, conduct, contract risks – CO:RE Knowledge Base (core-evidence.eu)

The model is not a linear representation of how harm occurs and there are likely to be many feedback loops within it

The model is not intended to be a linear flow-chart onto which a user's experience can directly be mapped. Instead, it is meant as a tool to map out all the factors that explain why harm manifests online for use within a policy context.

The different categories within the model are not isolated and there are overlaps and relationships between them. There may be some feedback loops:

- In some scenarios the harmed individual may go on to create further content, starting a vicious cycle.
- Psychological harm can lead to physical harm (e.g. believing hateful content may lead to offline violence) and vice versa (e.g. financial scam can lead to lost trust).
- There may be feedback loops between the platform and the user which can potentially amplify the harm (encouraging users to go down a 'rabbit hole' of further content).
- User characteristics are fixed. They may be 'learned' by the platform and in turn this may impact what the user is exposed to (e.g. by showing content that 'people like you' engage with).
- User circumstances and user exposure can, and potentially will, change in response to the content the user is served. Equally, the platform may change what content it serves in response to the user circumstances and exposure it identifies.
- Similarly, user circumstances might change user exposure, and user exposure might change user circumstances.
- Finally, the actions of other platform users may be input into the algorithms which in turn impact what the user sees. They might also impact directly on how a user chooses to act (e.g. encouraging them to act a certain way).

Research findings

How online experiences led to harm

People experienced different routes to harm

The in-depth interviews informing this research began with building an understanding of the harm participants had experienced (except among the control group). It became clear there were three main 'routes' via which harm occurred.

The evidence shows that people believed the harm they experienced was rarely caused solely – or even primarily – by individual pieces or genres of content. It was often content that seemed on the surface less harmful – even benign – that had a cumulative impact on people resulting in more significant and long-lasting harm. Some of the key components of the model – outlined in the next chapter – were refined to reflect this.

Overview of the routes to harm

The participants had different experiences of engaging with content / contact online, and the harm they experienced varied. However, the route through which they experienced harm appeared to group into three types.

These three routes were less related to the specific *risk factors* users had experienced – detailed later – and more about the overall pathway that led them to harm.

The three routes to harm are:

1. **Isolated** exposure to hazards leading to harm
   *Harm could manifest **immediately** or after a **delay***
2. **Cumulative** exposure to hazards leading to harm
   *Engagement with hazards could be **passive** or **active***
3. **Indirect** exposure to hazards leading to harm
   *Harm occurs due to hazards being exposed to **other people***

1. **Harm resulting from an isolated experience of being exposed to potentially harmful content / contact**

Within route 1 there are two subsets:

a) An isolated experience that has an **immediate** harmful impact
b) An isolated experience that has a **delayed** harmful impact

2. **Cumulative harm – caused by repeated exposure to multiple pieces of content / contact over time**

Within route 2, the way in which people engage with the content appeared to differ:

a) **Passive engagement** – the build-up of harm as a result of repeated exposure to potentially harmful content / contact in the online environment. An example in the research is exposure to high quantities of body-image related imagery in a social media feed.
b) **Active, self-reinforcing engagement** – this is when people actively engage with content that they may find appealing, but that over time encourages negative, self-reinforcing behaviours. Things like eating disorders or self-harm were often the areas of harm found from this route.

3. **Indirect harm -** caused by other people being exposed to content / contact that relates to someone, such as content that is spreading rumours about them or shaping others' perceptions of them. They don't have to be exposed to this content for it to cause them harm.

These different routes to harm highlighted some of the key points the model needed to be able to accommodate. For instance, **route 2 (cumulative harm) was prevalent across the sample, and more severe harm tended to result from it**. It was therefore important for the model to work for this route to harm. The model does not as effectively capture harm that occurs from the indirect route because it sought to capture harm as a result of someone being *exposed to content or contact*.

Some people described experiencing harm immediately from engaging with isolated pieces of content

The first route identified was:

1a) Harm resulting from an *isolated* experience of exposure to potentially harmful content / contact that has an *immediate* impact.

There were several instances of people coming across a single piece or type of content / contact and being negatively impacted by it immediately.

For example:

⚠ **Gore content**

Simon was shocked and revolted upon seeing a piece of extreme content he was not expecting.

Simon (26-30) used to browse niche online discussion board threads about Game of Thrones and certain anime. Sometimes, content was labelled to indicate it was, for example, a spoiler from a recent episode and you had to click through to actually see the content.

One day, he clicked on some content he later realised was intentionally mislabelled. He thought he was clicking on a piece of content about Game of Thrones (which had a spoiler alert tag on it). When he clicked through, it went to a page with gore content on it – a picture of dismembered bodies. Simon had no reason to believe that these bodies weren't real and found the experience extremely shocking.

The effect on Simon was instant: "The immediate effect was absolute revulsion…It sent me down, it really did". Simon said the content made him feel anxious, revolted and even guilty. He was worried that he'd seen something he wasn't allowed to see.

This happened to Simon relatively recently, and whilst he felt an immediate sense of disgust and shock about what he saw, soon afterwards these feelings started to fade.

⚠ **Homophobic and transphobic content, abusive content related to ethnicity or religion**

Eden is a streamer. They and their moderators were subjected to a 'hate raid' on their birthday stream.

Eden (21-25) identifies as non-binary. They are a professional streamer who shares livestreams of their life five days a week. They also use a social media platform to promote their streams and have a messaging app linked to their stream which is regulated by moderators (online friends who Eden has appointed).

On their birthday, Eden shared a celebratory livestream. In a one-off instance, their chat became flooded with 'bots' and 'raiders' whose usernames and profile pictures contained racist slurs and imagery. The 'raiders' flooded Eden's messaging app with violent and gory imagery.

While the 'raiders" profiles didn't seem to be anti-LGBTQ+ (their photos and usernames contained racist slurs and imagery) Eden was still shocked and upset by the content they were sent. They felt targeted and questioned whether

they were targeted for 'raiding' due to their LGBTQ+ status. They also felt guilt for their moderators, who had to engage and remove each piece of content the 'raiders' posted. Eden felt responsible for subjecting the moderators to this. For several weeks afterwards, they suffered from paranoia and anxiety about the security of their streams. Eden continues to stream.

Abusive content related to ethnicity or religion

Jayesh was affected by an Islamophobic comment left under a video he posted to a video-sharing platform

Jayesh (26-30) is a British Muslim man living with his partner. He has a passion for travel and keeps a video blog on his video-sharing platform, documenting the different adventures he has been on.

Jayesh is used to receiving mild criticism from his viewers. Sometimes his videos receive questions which are personal or ill-intentioned. Once, when he posted a video of him visiting Iraq, however, someone left an Islamophobic comment.

The unexpected and targeted nature of the content made Jayesh feel distressed and angry. He was disheartened about the comment and felt confused and upset for several days. He recovered relatively quickly and continues to do his vlog.

Generally, people tended to easily be able to self-identify as harmed and recall instances of harm that happened in this way. As in the cases above, while the immediate impact was significant, it quickly lessened.

One of the key risk factors within this route is the fact that the content is usually unexpected. Often, as in Simon's case, this comes in the form of a mislabelled or ambiguous link being clicked on and leading to extreme content. Those who experienced this route to harm tended to experience a short-term impact as a result.

Because route 1a is easier for people to recognise and recall, there is a risk that single, discrete pieces or genres of content receive disproportionate focus when thinking about what leads to harm.

Some people described experiencing delayed harm from engaging with specific pieces of content / contact

This route is: **1b) Harm resulting from an isolated experience of exposure to potentially harmful content / contact that has a *delayed* impact**.

This route describes instances whereby a user is exposed to a single piece of content / contact but isn't impacted straight away. It is only several months or even years later that they feel a negative impact.

This route was not seen within the adult sample because the recruitment criteria required the harm to have happened recently, but is evidenced in Ofcom's forthcoming report on *Risk factors that may lead children to harm online*.

Lucy, from this project, is an example of someone who experienced harm via this route.

Cyber-flashing/unsolicited comments and posts

Lucy (16) received unsolicited nude images when she was 13 and only later reflected on the harm this contact had on her

Lucy lives in London with her family. When she was 13, she received a few unsolicited nude pictures from someone she knew from school. In the moment, Lucy admitted that she didn't "pick up on how bad it was", and due to being younger admitted, "those kinds of things feel like validation".

Instead, for Lucy, there was a delayed effect. She said she now, age 16 realises how "bad" the situation was. It has negatively shaped her perceptions of men and has changed her behaviour on social media by shaping what she posts. She now wants to make sure that things she posts won't attract any unwanted attention from people who might try to do the same thing. She said: "I just don't want that kind of response. I post less now to avoid that stress".

She felt these decisions were a small and indirect consequence of the content she saw when she was 13. However, the effect of the unsolicited nudes shared with Lucy shaped her behaviour online over the next few years.

Lucy's experience shows that this route exists but, as it was not common even in the children's sample, further research is required to investigate the prevalence of this type of experience.

Like route 1a, the impact people cited resulting from this route tended not to be as severe as that detailed by others. What is less clear is the duration of harm. Further research may be needed to explore this in more detail.

Some people described experiencing cumulative harm due to repeated exposure to similar content / contact in their environment

The second route to harm is: **2a)** The build-up of harm as a result of **repeated exposure to potentially harmful content / contact in the online environment.**

Many people reported their online environment (on any given platform) being dominated by certain kinds of content that eventually caused them to have negative perceptions about themselves and/or the world.

This was the case for the people below:

Glamourised lifestyle content

Tala saw a high concentration of content that glamourised luxurious and 'successful' lifestyles in her social media feeds

Tala (21-25) spends a lot of time on social media. She is currently between jobs and enjoys scrolling to keep herself entertained. Her feed generates content which showcases success and wealth, especially among young people a similar age to her.

Tala enjoys engaging with content creators who showcase aspirational lifestyle content. She engages with content of this genre in high concentrations, meaning there is a limited number of other genres appearing on her feed. As a result, Tala's online environment depicts a world where seemingly all young people have wealth and success.

Meanwhile, Tala has recently moved back in with her parents in a rural place away from her friends and feels quite isolated. She has struggled to find a job after university and in the past has experienced periods of poor mental health related to her body image. She finds it easy to compare herself and her life to those she follows, and finds it lowers her self-esteem and makes her question her own success.

The build-up of glamourised lifestyle content has created an environment online which accumulated and caused Tala a degree of psychological harm – it shaped her perceptions of what the world is like and what she should value, profoundly affecting her self-esteem in a negative way.

Abusive content related to ethnicity or religion

Mahalah's opinion of her Jewish Israeli identity has changed over time in response to anti-Israel stances and antisemitism she perceives in the news and world around her.

Mahalah (26-30) was born in Israel and is Jewish.

After recent events in the Israel-Palestine conflict caused new waves of social media engagement with individuals often vocalising anti-Israel or anti-Palestine sentiment, Mahalah felt targeted and troubled by pro-Palestine stances taken by big brands and other social media users. She believed that this presented a simplified, one-sided view of the complicated conflict.

Over the course of several months, as the topic circulated on social media, Mahalah began to feel increasingly alienated from wider society. She believes she was made to feel ashamed of her heritage, a process she found very stress-inducing. In the long-term, she is still worried about revealing she is Israeli, especially to her colleagues at her new job.

Abusive and threatening comments and posts

Lydia sees a lot of content which she believes has an ageist slant which has changed the way she sees herself in society and impacted her mental health.

Lydia (66-70) lives alone. She has multiple sclerosis and struggles to leave the house. Being unable to socialise in person, Lydia spends a lot of time online using social media and reading news websites.

Lydia believes that old people are subtly ostracised in society, in the news and on social media. She used to have a preferred social media platform, but she withdrew from it during the Brexit vote when she felt it became overwhelmed with ageist content. During Covid she frequently saw old people referred to as 'bed-blockers', 'bed-wetters' and 'senile'. Now, Lydia uses other platforms and websites, but she still sees ageist remarks in the comments on a daily basis.

For Lydia, comments like this are indicative of a society which does not have time for older people. This has lowered her self-worth and she feels that people's negative perception of her as an older person in society has added to her depression.

Some people described developing harmful self-reinforcing behaviours and perceptions from repeatedly engaging with similar content / contact over time

This route to harm is: **2b)** engagement with potentially harmful content / contact that feels initially positive, but which over time cause **negative, harmful, self-reinforcing behaviours and perceptions.**

The people below reported experiences in line with this route.

Suicide and self-harm content, eating disorders

Aria wanted to find connection with people who'd had similar experiences, so she started engaging with content that resulted in her developing extremely negative and self-reinforcing behaviours

Aria (18-20) is a student.

She has a history of mental health issues stemming from her parents' divorce and the three-year development of an undiagnosed chronic disease which led to severe physical and psychological symptoms. From age 13-17, Aria described going through a "sad" phase where she was an "emo". She started an anonymous 'sad' account on a social media platform which she used to post and follow pages related to themes like depression and self-harm to try to find and connect with people who were going through similar things.

Through this page, she was added to an anonymous messaging app group purporting to be 'supporting' people with eating disorders. It was called "girls ED support group" – and turned out to be an anonymous group of users who were 'coached' in eating disorder tips. The group leaders asked the girls to submit their weight on a weekly basis and, if a girl had gained weight, they would have to send visual evidence of them self-harming as punishment.

Aria was in the group for 1.5 years and became very thin. She left the group when a friend stole her phone and read the messages. Shortly after, she attempted to overdose and was admitted into full-time psychiatric care and

remained in care for approximately one year. Aria still has twice weekly therapy sessions and takes medication for her depression.

Physical violence, sexual violence

Nora engages with high volumes of true crime content which she believes has reinforced feelings of mistrust and insecurity when mixing with other people

Nora (21-25), identifies as queer. She has been diagnosed with anxiety, depression, borderline personality disorder and avoidant personality disorder. She also describes herself as being "hyper-vigilant"' around other people and does not enjoy socialising, especially with people she does not know.

During the evenings, Nora likes to engage with true crime content. She reads stories for several hours before going to bed on an online discussion board which outline unsolved murder and sexual assault cases from around the world. A lot of the content has violent, gory, and sexual themes.

Nora reflected that her desire to engage with true crime content has exacerbated her distrust of strangers and her need to be "hyper-vigilant". For Nora, the content reminds her of people's violent capabilities. Consequently, she continues to remain relatively isolated.

The most severe instances of harm throughout the sample were similar to Aria's experience. The highest impact and longest lasting harm occurred when engagement with certain kinds of content / contact (often for seemingly benign or positive reasons to begin with) meant someone was repeatedly exposed to things that not only shaped their perceptions, but encouraged negative, self-reinforcing behaviours as well.

Some described experiencing harm as a result of content / contact relating to them being exposed to other people

The third route to harm is: **3) Harm caused by other people being exposed to content / contact that relates to a person.**

Whilst most of the harm reported was the result of the victim of harm being exposed to content or contact and experiencing harm as a result, there were examples of those who hadn't been exposed to anything, but who still experienced harm. This included harm such as social ostracism due to *other people* seeing content that related to them (e.g. rumours, content portraying them in a negative light, etc.).

This content / contact may go on to shape others' perceptions of or behaviour towards them, so they don't have to see this content / contact themselves to experience harm.

Abusive and threatening comments and posts

Negative rumours were spread about Connie to her friends and acquaintances without her knowledge

Connie (26-30) is an aspiring musician. Connie supplements her income working in various hospitality venues; spaces where she often casually socialises with local friends.

After falling out with a friend, with whom Connie explained she had a "toxic" friendship, Connie's friend created a fake social media profile under the name 'Joe King'. The profile was used to directly message Connie's local friends and acquaintances to spread untrue rumours about Connie.

Connie only learned about the fake profile when visiting her friends in a local pub one evening. One of her friends approached Connie and aggressively questioned 'her' actions. Connie learned that the fake account had been used to tell this friend that Connie had been spreading rumours containing sensitive information about the friend and their partner around the local area.

Over the next few weeks, several similar incidents occurred. At the time of the interview, Connie still didn't know how many people the fake account contacted, or what had been said. She felt socially ostracised, fearing to leave the house and socialise in the local area. She now prefers to stay at home with her boyfriend. Connie suffered these impacts because untrue content about her was exposed to other people.

Abusive and threatening comments and posts

Noah found out that people had been making fun of him in a group chat he was no longer part of

Noah (12-14) lives with his mum and two older brothers.

Recently, Noah reported experiencing bullying by some other boys at school whom he had previously been friends with. During the period of bullying, the boys had logged into one of Noah's gaming accounts and had spent his virtual currency. They had also shared content about Noah on his gaming account which he did not know about. This included an image of himself, originally taken on a social media platform, with an embarrassing filter. On another occasion, the boys had been making jokes about Noah's dad, who had passed away several years earlier, on a group chat.

Noah got sent screenshots of the conversation about his dad. He also found out about the boys' actions on the game. He was distressed and upset, especially about the comments made about his dad.

"He said he'd put my dad in a spliff or something like that… I'd left the group chat, so my friend had to take screenshots and stuff and send it to me."

What harms were experienced

People experienced a range of types and severities of harm

The model covers different types of harm

The model was designed to encompass different types of harm ranging from, for example, distorted worldview to depression, sleep deprivation, or financial harm. In some cases, these harms overlap or occur simultaneously. Following discussion between the Ofcom and Revealing Reality teams, these have been placed into either physical and psychological harm and were used as examples in the screening process. This is not an exhaustive list of harms but serves to illustrate that a range of types of harm may be experienced as a result of online content / contact.

Physical	**Psychological**
Harms that were observed / explored in this research: • Diminished relationships offline • Eating disorder • Financial harm • Not following health advice • Obesity and sleep deprivation • Physical harm (e.g. impact on health, broken limbs, hospitalisation) • Poor performance at work/school • Suicide attempts • Self-harm Harms that were not observed / explored in this research: • Legal liability for content (e.g. when someone receives child sexual abuse material they risk breaking the law – particularly if they share it, even if they were sharing for 'good' purposes) • Participation in crime or disorder • Risk taking (e.g. stunts, things that put you at physical risk)	Harms that were observed / explored in this research: • Anger • Anxiety, depression, stress, PTSD • Compulsion to view more • Encouraged unhealthy behaviours e.g. gambling, drinking excessively, smoking • Being groomed to participate in harmful behaviours • Harassment • Normalised 'negativity' • Self-blame, shame, low self-esteem, and guilt • Social comparison and pressure • Withdrawal/isolation Harms that were not observed / explored in this research: • Mistrust, including cynicism and scepticism of mainstream ideas • Radicalisation • Reinforce existing beliefs about distorted worldview

There are other types of harm which the model can accommodate

The model has two principal types of harm: **physical** and **psychological** – each with multiple examples of harm underneath. The reason for this is that it maintains the current framing of the Online Safety Bill, where 'harm" is defined as physical or psychological harm.[5]

Beyond direct physical and psychological harm there are other types of harm that exist and came out in the research. Whilst these aren't explicitly referenced, the model with its constituent parts can also support an understanding of how they manifest.

Examples of two additional types of harm that exist are **'economic harm'** and **'societal harm'**, although, as noted above, the research only engaged people who were aware they had been harmed, and consequently were unlikely to identify societal harm directly as this is difficult to observe at an individual level.

Economic harm

The model also describes the factors relevant to the risk of economic harm. Economic harm includes financial harm caused by scams and fraud, but can also include loss of time and productivity in the form of opportunity costs.

Opportunity cost

Opportunity cost was a harm that came up several times throughout the research. Opportunity cost can be defined as the forgone benefit that would have been derived from an option not chosen. For example, a harm can be experienced when the amount of time someone spends doing something (in this case, being online) means they are missing out on other things that may benefit them. Equally, an opportunity cost could come from choosing to remove oneself from a platform to avoid some of the content / contact there, even when there are benefits to being on that platform (such as being able to connect with friends, make money etc). One adult from this research, Ellie, was able to reflect on this after she spent an unusual amount of time on a social media platform:

Ellie's screen time soared during the BLM protests as she engaged heavily with the content being shared. She believes her work suffered as a result.

Ellie (26-30) is a part-time illustrator and uses social media heavily to promote her artwork. Ellie is mixed race. During the BLM protests, she claimed she felt an obligation to educate her followers and be an active ally.

"During BLM, I felt a lot of responsibility to share information and trying to be a good ally. I felt like there was no space for me to rest because I was a non-black person of colour. It was a very real pressure to feel like we were doing all we could, but there was no real 'upper limit'. I was spending 32 hours a week on [a social media platform] whilst I had a full-time job. Of course, my work suffered! I was so sad and anxious because I felt like my greater duty was educating people and posting about [BLM] rather than work."

Ellie believed that her online interactions during the BLM protests came at the cost of her progress at work and her mental health. She eventually decided to take a step back from social media and now rarely promotes her artwork or political views on the social media platform.

Harms like that experienced by Ellie come about when online activities mean losing out on other things in wider life. Some people told us that the amount of time spent online may prevent some people from taking opportunities in their offline lives.

The parallel research into children's online experiences in Ofcom's forthcoming report on *Risk factors that may lead children to harm online,* found that several parents expressed concern about the potential opportunity cost their children experienced when spending more time online, rather than offline or with friends in 'real-life'. As

[5] Online Safety Bill (Clause 190): https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf

with adults, the amount of time they spend online can potentially have directly harmful consequences in terms of what they are missing out on.

Financial harm

Harm caused by scams and fraud may eventually lead to psychological and even physical harm in the long term. But it also exists as a harm in itself. There were a few examples within the sample of people who had suffered financial harm, some of which are mentioned above. Again, factors outlined in the model help predict that these scenarios are likely to lead to harm.

Financial scams

Neil fell for a romance scam online and lost £3,000 over the course of the 'relationship'

Neil (61-65) has settled in the Southwest of England after living in France in his youth. He works doing odd-jobs and flexible gig-work, but his finances are tight. At the time of interview, he was saving to visit one of his children who lives abroad.

Neil uses social media to stay in contact with his children but admits he does not fully understand the different platforms or how to use security settings. Several months earlier, Neil had been contacted by a woman who claimed to live in a French town that he had lived in when he was younger.

They began messaging. Over three months, the relationship developed and the woman had asked Neil for different amounts of money to 'help her get established' in her new life in France. As her stories grew more outlandish, Neil began to grow suspicious and stopped sending money to her. By this point, however, Neil had already sent £3,000 in different payments.

At the time of interview, Neil still did not know if the person he was speaking to was real and questioned the elaborate lies she wove (including introducing him to her family members and employers over video call). He had not told his children about losing the money because he felt ashamed and upset that he would have to wait longer to visit them. He was forced to work extra hours to recuperate the money he lost.

Financial scams

Josh was encouraged by an influencer to invest in a new crypto-currency and lost £5,000 when the currency failed

Josh (26-30) works for a financial institution. He used to work with someone who grew to have a large social media presence as a 'crypto-currency expert'. Josh decided to follow them and enjoyed the glamourous lifestyle content the former colleague shared.

Josh is a social person who enjoys networking. He reached out to the colleague and began to learn about crypto from him. Josh was eventually convinced to invest in a new crypto currency. The currency failed.

Josh lost £5,000 through the currency failing. This triggered a bout of depression and extreme weight gain.

Abusive and threatening comments and posts

Eden, a professional Twitch streamer, was the victim of a 'HateRaid' on a streaming platform. They also suffered financial harm through the event.

'Hate-raiders' flooded Eden's stream with violent and gory imagery on their birthday livestream.

As well as the psychological impact the raid had on them, the status of Eden's Twitch account suffered when it was followed and raided by bots and accounts which shared hateful content. Their follower account dwindled and the reputational damage of the raid meant sponsors were less likely to engage with Eden's streams. This impacted the revenue that Eden could make from streaming.

Societal harm

Some online content / contact can also lead to societal harm. For example, content / contact aiming to radicalise or misinform could result in physical or psychological harm to the individual, but they could also constitute societal harm independently of whether physical or psychological harm occur. These are harms that may have more of a social impact, or an impact on society. If someone comes to believe something that is not true, it may be eventually harmful to them or others, but also has wider implications for society by increasing polarisation and fragmentation.

The research team believe the model *can* accommodate this type of harm, although it was not directly observed due to the recruitment approach which required people to be aware they had experienced harm. The model includes the 'user exposure' component to emphasise the prevalence and severity of cumulative harm. Although there is no evidence of radicalisation in the sample for this research, it's reasonable to hypothesise that it results from repeated exposure to a number of different pieces of content / contact (promoting a certain narrative) in high concentration over time. The model and the key accompanying risk factors allow this form of harm to be considered alongside physical and psychological harm.

Why people experienced harm

The likelihood of harm was affected by a range of risk factors

Across the sample, the researchers studied the factors which influenced how harm occurred for each participant. There were several factors which seemed to increase the potential that content or contact would lead to harm, which are described here as **risk factors**. Although it was not an explicit focus of this research, risk factors may also increase the severity of the harm experienced.

Some of the factors appeared to increase the likelihood that a user would be **exposed to potentially harmful content or contact** – either for the first time, or repeatedly. Others appeared to increase the potential of the person **experiencing harm** as a result of exposure. This is indicated where relevant.

None of the risk factors can be assumed to always, in all cases, increase the risk of harm and **some risk factors were present for respondents in this research who had not experienced harm.** The presence of one risk factor alone did not mean harm would occur. Participants in the control group had experiences with content or contact where risk factors were present, but where significant harm did not occur. This was due to the presence of other factors – for example, not feeling that something was personally aimed at them or having a supportive social network.

In some cases, the presence of some risk factors might not be apparent to the user being interviewed, for example, the design of an algorithm that affects the content being served to users or the source of the content they saw. As a result, case studies reflect what people were able to report but cannot necessarily tell the full story.

Media literacy also has an important role to play in online safety, and elements of this are apparent throughout the model, although not explicitly identified by the users themselves.

There were some components of the model about which the research could not draw clear conclusions. A factor or characteristic may have been observed but there was not sufficient evidence to indicate that it was a significant risk factor. These have been listed under the relevant component as factors which were considered and may warrant further investigation in the future.

How to use this section of the report

This section outlines the risk factors which were observed. Each factor is introduced by a summary of the key features that relate to it:

Key features:

- Whether it appears to increase risk of **exposure** to content
- Whether it appears to increase risk of **harm** as a result of exposure to content
- Which routes to harm (**isolated or cumulative**) and which type of engagement (**active or passive**) it most commonly appeared alongside
- Whether there were any other risk factors it commonly interacted with or was exacerbated by.

The description of each risk factor includes a link to a case study in the annex for further detail on how that factor played out in a research participant's experience of online harm.

Risk factors associated with the content / contact

This section describes risk factors associated with potentially harmful content / contact, e.g. the content of a message, the nature of an image, or the way in which it was experienced by a user. It is important to note that people in the control group saw similar content or contact and said they were not harmed - for example, one participant intentionally sought out and enjoyed looking at gore content relating to real-life injuries (appeals to a desire and perceived to be real). In this case, the fact that they sought out the content seemed to mean that they did not feel harmed by it.

1. Content / contact that appeals to a desire or feels like a 'quick fix'

Key features:

- Appeared to increase risk of **harm**
- Appeared most commonly in **cumulative** routes to harm with **active engagement**
- Often exacerbated by:
    - **User characteristics**: Users under the age of 18 from Ofcom's forthcoming report *Risk factors that may lead children to harm online* had an increased likelihood of harm occurring as they were more vulnerable and impressionable to the content / contact they experienced
    - **User circumstances:** The presence of a mental health condition, especially an eating disorder

For a case study related to this risk factor, see 'Case study 1: Scarlett' in the annex.

The research documented the experiences of users who had sought out diet content or porn that made them feel good in the short term. This appeared to increase the likelihood that a person would experience **harm** because the content gave the impression that it was harmless or beneficial to begin with. In some cases, this meant the user felt no need to stop engaging with similar content. Often users did not recognise the detrimental impact of the content until later.

2. Content / contact that is perceived as aspirational or holds social status

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared most commonly in **cumulative** routes to harm with **passive** or **active engagement**
- Often exacerbated by:
    - **User exposure**: Users who saw high quantities and concentrations of this content appeared more likely to experience harm as a result
    - **User circumstances:** Users with low self-esteem or pre-existing mental health conditions seemed more susceptible to this kind of content

For a case study related to this risk factor, see 'Case study 2: Laura' in the annex.

People sometimes described content or contact they had experienced as aspirational, cool or popular. Some seemed to feel that they may gain status from emulating or engaging with it.

Some reported experiences when they had seen content that portrayed dangerous eating habits positively, showed unattainable glamourous lifestyles, or where people were part of communities that rewarded certain body shapes or behaviours. These appeared to increase the likelihood that a person would experience **harm** because users felt encouraged to emulate the behaviour even when it was not healthy or achievable.

This factor also appeared to increase **exposure** to content because it was widely circulated, with people sometimes describing it as 'viral' or something that many people re-shared.

3. Content / contact that is geographically local

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared across all routes to harm – **isolated**, **cumulative** and **indirect**

For a case study related to this risk factor, see 'Case study 3: Maya' in the annex.

Some people had experienced harm from content that was 'geographically local', in that it related to or originated from their local area. This content might depict, be related to, or be tagged to a location close to the user, or a location which the user regularly visited.

For example, some people reported having seen videos of violence taking place in their local area. This experience appeared to increase the likelihood that a person would experience **harm** because it affected how they felt about places they went or people they interacted with every day.

This factor might also increase **exposure** to potentially harmful content / contact because the content is geographically close to the user and may therefore be more likely to be served to them by the platform or shared or posted by people they know.

4. Content / contact that is perceived as real / genuine

Key features:

- Appeared to increase risk of **harm**
- Appeared across both **isolated** and **cumulative** exposure to potentially harmful content
- Often exacerbated by:
    - **Content / contact factors:** When content was unexpected, shocking or graphic
    - **Content / contact factors:** When content was aspirational

For a case study related to this risk factor, see 'Case study 4: Simon' in the annex.

In this report, the 'perceived realness of content / contact' is thought by respondents to be a true representation of events, or content that hasn't been faked, edited or adapted.

For example, the researchers heard about experiences when people had seen violent content that they thought portrayed real events or people. This led to **harm** as it made the images harder to dismiss and meant they stayed in people's minds more, increasing anxiety.

Others had seen content relating to glamourous lifestyles or unrealistic body shapes that they thought were real. When people didn't dismiss this as faked or adapted, they felt that these lifestyles or body shapes were more achievable, lowering their self-esteem and increasing the pressure people felt to conform.

The researchers also heard about experiences when people had received fraudulent messages which they believed to be real. This led to them engaging in fake relationships or activities leading to financial harm.

5. Content / contact that is relatable or personal

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared across both **isolated** and **cumulative** exposure to potentially harmful content / contact
- Often exacerbated by:
    - **User exposure**: Users who saw high quantities and concentrations of this content appeared more likely to experience harm as a result
    - **User circumstances:** Users with low self-esteem or pre-existing mental health conditions seemed more susceptible to this kind of content

For a case study related to this risk factor, see 'Case study 5: Tala' in the annex.

'Personal or relatable' content / contact often features or is from a person of a similar age or background, or who shares similar experiences to the user. In other cases it is personal because it attacks or targets an individual or characteristics that the individual identifies with.

Some people reported seeing content showing people similar to them living glamourous lifestyles. This appeared to increase the potential that a person would experience **harm** because they could see people they thought were like them living lives they felt were unobtainable. This made them negatively reflect on their own achievements and lives. Some did not identify that this type of content was harmful until later down the line.

This factor might also increase **exposure** to content because it is related to the individual and may therefore be more likely to be served to them by the platform, due to the algorithmic functionality of platforms.

This risk factor also includes content / contact that personally attacks the user. For example, some people reported harassment relating to their sexuality from a person known to them, or harassment about their gender or appearance. Sometimes it was clear that the distributor had malicious intent. This appeared to increase the likelihood that a person would experience **harm** because it made them feel personally targeted, self-conscious or vulnerable.

6. Content / contact that is unexpected

Key features:

- Appears to increase risk of **exposure** to content
- Appears to increase risk of **harm**
- Appeared most commonly in **isolated** routes to harm
- Often exacerbated by:
    - **User circumstances:** Users with pre-existing mental health conditions or trauma seemed more susceptible to this kind of content
    - **Distributor characteristics:** Users who knew or lived geographically close to the sender seemed more susceptible to this kind of content

For a case study related to this risk factor, see 'Case study 6: Luke' in the annex.

In this context 'unexpected' means content / contact that comes out of the blue, or equally which people do not anticipate appearing in their feed, search results or direct messages. It is incongruous with the other content / contact they come across online.

For example, people reported experiences where they had unexpectedly seen videos depicting violent scenes on their social media feeds, had received unexpected sexual videos from unknown distributors via direct message or had received abuse about their appearance out of the blue.

The likelihood that a person will be unexpectedly **exposed** to content / contact increases when a user only expects certain types of content in their feed. They may unquestioningly click, follow, or engage with content without assessing any initial information which might convey that it is potentially harmful.

This also appears to increase the potential for a person to experience **harm** because this content is a sudden change from what they normally see online and is not something they are prepared for. For example, people in the sample reported that unexpected content made them feel shocked and violated.

Risk factors associated with the characteristics of the distributor

Distributor characteristics are characteristics of the person or account who distributed the content or was the source of the interaction.

Some of these factors are similar to those in the 'content' section, (e.g. geographically local content and geographically local distributor) – but they merit separate mentions as they do not always overlap, e.g. something may be shared by a user who is geographically local but the content is not related to a nearby location.

7. Distributor who is known to the user

Key features:

- Appears to increase risk of **exposure** to content / contact
- Appears to increase risk of **harm**
- Appeared most commonly in **isolated** and **indirect** routes to harm
- Often exacerbated by:
    - **User circumstances:** Users with low self-esteem or pre-existing mental health conditions seemed more susceptible to this kind of content

For a case study related to this risk factor, see 'Case study 7: Adil' in the annex.

Some people in the research reported experiences where users were harassed by people they knew. This could include people they knew in real life such as (former) partners, (former) friends or people in the local area, and people who were known to the user online, including a user's followers and people who they had formed connections with online but had not met offline.

This appeared to increase the likelihood that a person would experience **harm**. People reported it felt like a more personal attack that could not be discounted as easily as an interaction with an anonymous distributor. Others mentioned they felt there was a greater risk the distributor knew sensitive information about the user.

People also reported greater engagement with content / contact because the distributor was known to them. This appeared to increase the likelihood of the person being **exposed** to content / contact because they trusted the distributor and were therefore less cautious.

This factor might also increase **exposure** because content / contact shared by people known to the user may be prioritised by platform algorithms, especially if the user engages frequently with other content shared by that person.

8. Distributor who is geographically close

Key features:

- Appears to increase risk of **exposure** to content
- Appears to increase risk of **harm**
- Appeared most commonly in **isolated** routes to harm
- Often exacerbated by:
    - **Platform characteristics:** Users who could be easily contacted or contact others who were also geographically close.

For a case study related to this risk factor, see 'Case study 8: Shruti' in the annex.

'Distributor who is geographically close', refers to a distributor who appears to be in the local area, either through being tagged as being nearby, stating that they are nearby, or are known by the user to be nearby.

People in the sample reported experiences when users had altercations online with others who lived nearby, or where users had received violent threats from others nearby in relation to their sexuality or disability.

This appeared to increase the likelihood of **harm** occurring because it made users feel vulnerable and at risk of physical harm or abuse when at home or out in their local area. Some felt they were not in control of who may be able to approach them—and that harassment could continue in real life as well as online.

There was also an example when a participant had connected to other platform users who were geographically close using a platform function which suggests profiles to add who are nearby. These users had later shared content relating to dealing drugs in the local area. Although in this case the participant was not harmed, this factor increased the likelihood of **exposure** to harmful content when in conjunction with platform characteristics.

9. Distributor who holds status or a reputation

Key features:

- Appears to increase risk of **exposure** to content / contact
- Appears to increase risk of **harm**
- Appeared across both **isolated** and **cumulative** routes to harm and is also likely to appear in **indirect** exposure

For a case study related to this risk factor, see 'Case study 9: Humza' in the annex.

'Distributor who holds status or a reputation', refers to distributors who seem to be liked or engaged with by lots of other users, or who hold a position that the user sees as desirable or aspirational.

For example, there were several instances of people with an eating disorder engaging with eating disorder recovery pages where the distributor had high numbers of followers and their posts garnered high engagement. Similar experiences were reported by people engaging with aspirational lifestyle content and advice online distributed by people with lots of followers, and was reported once by someone who had lost money in a financial crypto-investment scam.

In the group of people who were unharmed, people who had been sent scam messages promoting modelling and brand representation opportunities but had not fallen for them, commented that the image-sharing platform profiles messaging them had few followers. This indicated the sender had a less significant profile and was therefore less trustworthy.

This risk factor appeared to increase the likelihood of **harm** occurring because the high status of the distributor meant people sought to engage with the content being shared and were more likely to adhere to their advice or alter their behaviour in response to their content.

It also increased the likelihood of **exposure** to content / contact because the high engagement on their posts increased the likelihood that people would see the content.

People also reported experiences of threats being sent to a user by a distributor known to hold status in the local area. This appeared to increase the experience of **harm** as it made the user feel more at risk.

Factors for further consideration / research related to distributor characteristics

- **Malicious intent by the distributor** and **persistence of the distributor**. Malicious intent may increase the likelihood that a distributor will create content in the first place, but it was not observed as a factor in this research that increased the potential for harm. However, if a feature of the content indicated malicious intent (e.g. a user is threatened), it has been included under the *personal/relatable* risk factor. Persistence of the distributor is linked to volume of content received which is also covered under 'Exposure'.
- **Sophistication (e.g. troll farms).** This was not noted by any of the respondents, although it may not be visible to users.

It will be difficult to identify these factors using primary research with users. Instead, experts in the field may be able to explain how these factors could interact with those outlined, to assess the value of adding them into the model.

Risk factors associated with the fixed characteristics of the user

Fixed characteristics are those that the user can't change, including some protected characteristics. The research analysed whether these factors were seen to make it more likely that participants had been exposed to content or contact and/or whether they were more likely to experience harm.

The only fixed characteristic that was seen among the participants to consistently increase likelihood was the user's age, although there are observations about the role of protected characteristics below too.

10. The age of the user

Key features:

- When users are children, this appears to increase the risk of **harm**
- Appeared across all routes to harm – **isolated**, **cumulative** and **indirect**

For a case study related to this risk factor, see 'Case study 10: Ahmed' in the annex.

This study did not include any participants under the age of 18 so this research also drew upon some of the stories and knowledge from Ofcom's forthcoming report *Risk factors that may lead children to harm online* project, where the children's characteristics, among other factors, were considered as part of their journey to potentially experiencing harm online.

However, some participants in this adult study also referred to harms they experienced online when they were under the age of 18.

There were examples of people who said they had been groomed to participate in harmful behaviours; exposed to self-harm, eating disorder and suicide content; or had been harassed by other online users. Being under 18 influenced the likelihood of an individual experiencing **harm** because younger age groups were more likely to be vulnerable or impressionable to content / contact they experienced.

Factors for further consideration / research relating to user characteristics

- **Holding 'targeted' opinions or characteristics**. This links with the 'Acceptability of posting that content' within societal context.
- **Public role.** The sample did not specifically include people who held a public role. People in public roles might be more likely to be exposed to potentially harmful content / contact, such as harassment.

These factors could be explored in more depth through targeted research with people in these groups, comparing the results with the experiences of those who are not in these groups. It is possible that they are relevant to some types of harm but may not belong in a generic model which seeks to work across all harms.

Protected characteristics

Based on this research project alone, it cannot be concluded that the possession of any other user characteristic increased the likelihood of exposure to or harm from a hazard. The sample was deliberately designed to include a range of people and so it is not possible to identify if harm was more prevalent in certain

groups. Relevant user characteristics are likely to vary by harm, as demonstrated in Ofcom's online experience tracker findings[6].

There were people across the sample who did experience exposure to and harm from content / contact targeting an identity or **protected characteristic** they possessed. For example:

- Brian felt **harmed** by anti-Irish hatred aimed at him. This was shared in messaging app groups he was added to by members of an opposing football team after Brian's team had beaten them.
- Mahalah is from Israel and is Jewish. She said she was **harmed** by antisemitic content, individuals and companies taking anti-Israel stances on social media.
- Adil said he was **harmed** by homophobic abuse from friends and acquaintances in his 'Indian International Students' student groups, after he came out as bisexual on his image-sharing platform story.
- Jemma, who was part of the unharmed control group, was **exposed** to abuse following arguments with anti-trans campaigners on a social media platform.
- Jayesh is a British Muslim. On one video he posted of himself in Iraq, someone left an Islamophobic comment that he said **harmed** him by leaving him feeling distressed and angry.
- Vera uses her social media to share images of her work, which is important for her networking. She frequently receives misogynistic abuse, which she finds **harmful**.

Whilst protected characteristics were influential in the experience of harm in these cases, the nuance of each story means it is not possible to draw conclusions about other protected characteristics which may make exposure to or harm from content / contact more likely to occur. Having protected characteristics could be a risk factor but this warrants further research.

[6] Ofcom urges tech firms to keep women safer online - Ofcom

Risk factors associated with the circumstances of the user

User circumstances differed from user characteristics because they were the result of individual experience or they could theoretically change or be changed over time (although they might not be in practice).

The user circumstances observed to increase the likelihood of exposure to or risk of harm included the cost to the user of withdrawing from the online space, isolation or lack of social support, the user's mental health or their past experiences.

11. The cost to the user of withdrawing from the online space

Key features:

- Appears to increase risk of **exposure** to content / contact
- Appeared most commonly in **isolated** routes to harm

For case studies related to this risk factor, see 'Case study 11: Lucy' and 'Case study 12: Michelle' in the annex.

In this context, 'cost of withdrawal' means that the user will face significant negative repercussions from reducing their engagement with online activities.

For example, there were experiences where people relied on their use of social media and streaming for income. Other users spoke about how they engaged with potentially harmful content as a result of being on platforms as a source of social information; platforms where they reported seeing more potentially harmful content.

This factor appeared to increase the likelihood that a person would be **exposed** to further content / contact across both the harmed and unharmed people in the sample. For example, one participant, who reported receiving unsolicited nude images and sexual messages but was unharmed by this, also used the platform to communicate with her family who lived abroad. There would be a high social cost if the user were to withdraw from the platform, so they continued to use it, increasing their likelihood of being **exposed** to further content / contact.

This was also seen in Ofcom's forthcoming report *Risk factors that may lead children to harm online* project. The fear of being left out or losing social interaction meant some young people were inclined to stay in online communities, even though some reported disliking the content.

12. Isolation or lack of social support

Key features:

- Appears to increase risk of **exposure** to content
- Appears to increase risk of **harm**
- Appeared in both **isolated** and **cumulative** routes to harm, and alongside **active** and **passive** engagement
- Often exacerbated by:
    - **User circumstances:** Users with low self-esteem or pre-existing mental health conditions.

For case studies related to this risk factor, see 'Case study 13: Jackie' and 'Case study 14: Neil' in the annex.

Across the sample, there were several people who were isolated or lacked social support during time spent abroad away from family and friends, through being single, or through having few offline friendships. This appeared to increase the likelihood of a participant being **exposed** to potentially harmful content / contact as they often spent higher amounts of time online, increasing the volume of content / contact they engaged with.

One participant recognised that when she was feeling lonely and isolated on her exchange year during university, her screen time rose to 10 hours per day.

Across the sample, it appeared that there was also a higher likelihood of content / contact leading to **harm** when a user was isolated or lacked social support. For example, participants who had few people around them to step in or support them when harmful behaviours started to occur or after they had been exposed to potentially harmful content / contact.

13. The mental health of the user

Key features:

- Appears to increase risk of **exposure** to content
- Appears to increase risk of **harm**
- Appeared in both **isolated** and **cumulative** routes to harm and will likely also appear in indirect routes. Routes to harm appeared alongside **active** and **passive** engagement
- Often exacerbated by:
    - **Content / contact:** that is relatable or personal, or unexpected, local, aspirational
    - **Distributor characteristics:** known to the user

For a case study related to this risk factor, see 'Case study 15: Kevin' in the annex.

There were people in the sample with mental health conditions, poor mental health and low self-esteem who increased their likelihood of **exposure** to hazards by seeking out content / contact relating to their condition. This included content / contact about or encouraging eating disorders, suicide, self-harm, and unobtainable glamorous lifestyles.

In these cases, people in the sample with mental health conditions, poor mental health and low self-esteem had an increased likelihood of **harm** occurring from the content they were exposed to, as they were more susceptible to following advice which might worsen their condition or cause them further physical or psychological harm.

A user with a mental health condition, poor mental health, or low self-esteem may also be more likely to experience **harm** from content unrelated to their condition but which could be considered upsetting or distressing. For example, there were experiences where factors related to content / contact increased the likelihood of harm occurring, including when it was unexpected, geographically local, aspirational, or content where the distributor was known to the user.

14. Past trauma or experiences of the user

Key features:

- Appeared to increase risk of **harm**
- Appeared most commonly in **isolated** routes to harm, but did also appear in **cumulative** routes to harm. Appeared with both **active** and **passive** engagement.
- Often exacerbated by:
    - **Content / contact**: whether the content / contact was unexpected

For case studies related to this risk factor, see 'Case study 16: Kylie' and 'Case study 17: Katya' in the annex.

There were people in the sample who had experiences of in-patient psychiatric care, suffering with a mental health condition, falling for an online scam, receiving online abuse and harassment, and being a victim of physical and/or sexual assault and abuse. All of these experiences influenced how the users interpreted the content they engaged with and the **harm** they reported experiencing.

Not all people who had experienced past trauma experienced subsequent online harm. Within the sample who were unharmed, one participant had a previous traumatic experience but was not 'triggered' by content related to her experience which she believed was because she had received adequate therapy and support after the experience.

Factors for further consideration / research into user circumstances

- **Offline friends: confirmation bias or desensitisation**. The attitudes and behaviours of a user's offline friends are hypothesised to increase the likelihood of their being exposed to harm.
- **Mistrust mainstream narrative.** This could be hypothesised to be a factor, but it was not evidenced within our current sample which focused on people who know they have been harmed.
- **Using a shared device.** Most people reported using their own devices (e.g. personal smartphone). There were no examples of people encountering content / contact on other people's devices. However, this may be more likely when users are children and more likely to be sharing devices with others (e.g. parents, siblings). This was observed in the research project *Risk factors that may lead children to harm online*.
- **Environment when consuming content (time, surroundings)**. This may not be a standalone risk factor but may overlap with time spent online (e.g. if viewing late at night), which could be hypothesised to be a factor, but was not evidenced within this research.
- **Addiction / over-use of platforms.** There was some evidence of people using online platforms very frequently, and this would play a role when considering opportunity cost as a type of harm. There were some instances where time spent online increased the likelihood of exposure to content / contact, but this wasn't observed across enough of the sample to elevate it to a key risk factor, and people didn't attribute their experience to time spent online. It could be hypothesised that if a user was spending more time online and engaging with higher volumes of content, it would be more likely that they would be exposed to more content / contact. This is also discussed in the *Isolation/ lack of social support* risk factor.
- **Financial stability.** In one or two cases of fraud/scams, people appeared to be more likely to fall victim to the scam when they were less financially secure, especially when they were influenced by other people's glamourous lifestyles. This did not seem to be a risk factor in other cases but may warrant being elevated to a risk factor in the future or for specific harms.

These factors could be explored in greater depth through targeted research with people in these groups, comparing the results with the experiences of those who are not in these groups. It is possible that they are relevant to some types of harm, but may not belong in a generic model which seeks to work across all harms.

Risk factors associated with what else the user is exposed to or does online

User exposure can be split into two categories: what other content / contact users are exposed to, and what they do in response to seeing content.

15. Not seeing counter narratives

Key features:

- Appeared to increase risk of **harm**
- Appeared most commonly within **cumulative** routes to harm with **active** engagement
- Often exacerbated by:
    - **User circumstances:** the mental health of the user
    - **User characteristics:** Users under the age of 18 from Ofcom's research *Risk factors that may lead children to harm online* had an increased likelihood of harm occurring as they were more vulnerable and impressionable to the content they experienced

For a case study related to this risk factor, see 'Case study 18: Josh' in the annex.

Within the sample, there were several people who recognised that being exposed to a high concentration of certain content, without any counter narratives, was a contributing factor to the **harm** they experienced. This was particularly seen with content related to steroid use, self-harm, crypto-currency investment and scams.

People reported that their thinking and beliefs intensified, leading them to take actions they may not have done had they been prompted to question what they were engaging with.

16. Seeing a lot of similar content

Key features:

- Appeared to increase risk of **harm**
- Appeared to increase risk of **exposure**
- Appeared most commonly within **cumulative** routes to harm with **active** or **passive** engagement
- Often exacerbated by:
    - **User circumstances:** isolation or lack of social support; the mental health of the user

For a case study related to this risk factor, visit 'Case study 2: Laura' in the annex.

There were several accounts of people seeing a lot of content relating to diets, weight, and healthy eating. Other types of content people saw repeatedly included true crime content, glamourised lifestyle content, violent pornographic content, abuse and harassment, health or illness related content, and content related to steroids.

People impacted by this risk factor often reported that seeing lots of similar content ended up shaping their expectations, perceptions, and beliefs in a negative way.

This factor may also increase **exposure** to content if the user actively engages with the content they see a lot. Algorithms are likely to play a significant role in presenting more similar content to users, although this was challenging to evidence specifically within this research as people's ability to self-report on this varied.

17. Active user engagement with the content / contact

Key features:

- Appeared to increase risk of **exposure** to content
- Appeared in both **isolated** and **cumulative** routes to harm

For case studies related to this risk factor, visit 'Case study 19: Kathleen,' 'Case study 20: Lana,' 'Case study 21: Katie,' and 'Case study 22: Jemma,' in the annex.

People appeared to have been **exposed** to more content / contact after sharing content on social media, responding to comments, engaging with others in comments sections of social media posts and in live streams, and searching for content online and on social media.

Engaging with content / contact is likely to influence platform algorithms and therefore increase the likelihood of the user being exposed to similar content / contact. However, it is only likely to lead to **harm** if the user is engaging with content / contact when they have other risk factors.

There were many examples of people reporting or blocking content. Whilst blocking did seem to be effective in reducing exposure in some cases, many people felt that reporting content was not effective, and some stopped attempting to report content as a result.

Factors for further consideration / research into user exposure

- **Joining fewer mainstream sites or groups.** There was some evidence of people choosing to join fewer mainstream sites or groups. It could be hypothesised that this may be a risk factor for increased exposure to potentially harmful content but there was not sufficient evidence within this research. There is some connection to the *How easily users can be linked / signposted to other content or platforms* risk factor, where being able to easily link away from one platform to another and come across unexpected content / contact was observed to increase the potential for the potentially harmful content to lead to harm. However, how easily users can be linked / signposted to other content or platforms is distinct in that it would facilitate stumbling across less mainstream content rather than choosing to join a less mainstream group.
- **Belonging to a group typically seeking that content** e.g. if an algorithm identifies that teenage boys appear likely to engage with violent content then if it identifies a new user as a teenage boy, it could serve them violent content too.

These factors could be explored in greater depth through targeted research with people in these groups, comparing the results with the experiences of those who are not in these groups. It is possible that they are relevant to some types of harm, but may not belong in a generic model which seeks to work across all harms.

Risk factors associated with wider societal context

The factors below were observed to make it more likely that people would be exposed to content / contact which may lead to harm.

18. Acceptability of posting content

Key features

- Appeared to increase risk of **exposure** to content / contact
- Appeared most commonly in **isolated** routes to harm with an **immediate** impact
- Often exacerbated by:
    - Platform characteristics: Platform tolerance

For a case study related to this risk factor, visit 'Case study 23: Adam' in the annex.

Content posted by a user deemed not 'acceptable' by other users of the platform – because it challenged the popular beliefs or values of that online space – was more likely to lead the poster of that content to **exposure** to potentially harmful content / contact in the form of abuse and ostracism.

When content posted by a user was not seen as 'acceptable' on the platform, it increased the likelihood of **exposure** to content / contact which could cause harm if a combination of other risk factors were present.

In the sample, one participant increased their **exposure** to potentially harmful contact when they commented on a post deemed unacceptable by a certain group of people on the platform, and got sent hateful messages, even death threats, as a result (see case study 23 in the annex).

What is deemed 'acceptable' will be influenced by the different cultures and guidelines on different platforms.

19. Trending topics

Key features:

- Appears to increase risk of **exposure** to content
- Appeared in both **isolated** and **cumulative** routes to harm across **active and passive** engagement

For a case study related to this risk factor, visit 'Case study 16: Kylie' in the annex.

'Trending topics' means any topic or trend circulating in popular discourse online or offline in the news / media / society, etc.

For example, several people in the sample were exposed to content because it was a popular or trending topic at the time. This included news stories and popular social media or streaming trends (like using a certain filter or participating in hate raids). Several people at the screening phase of the project also noted they experienced distress and were upset by currently trending content, for example the war in Ukraine or the trending animal abuse video of the footballer Kurt Zouma kicking his cat.

Risk factors associated with the characteristics of the platform

Platform characteristics were challenging for people to identify for themselves, and it was harder for the researchers to identify factors solely from the testimony of research participants. This section illustrates some factors that people were able to identify themselves as contributing to harm, or where it was clear from the evidence that the platform characteristics played a role.

Other research[7] breaks down the key design strategies of the major platforms to understand how the functionalities of platforms are designed to shape the behaviour of users. While it cannot be concluded, from this research, that these functionalities increase the likelihood of harm or exposure to content / contact, it may be hypothesised that, for example, the design strategy of increasingly showing users similar content to the content they engage with may increase the risk of route 2 harm (cumulative).

Some platform characteristics were also enablers of other risk factors outlined throughout this section, for example features that allow users to 'like' content may have a role in making content seem more *Aspirational* (as set out in Content factors). Equally, as with the other factors, there were some examples in the control group where the same platform characteristics were present but the participant did not feel they had been harmed.

20. How easily users can be contacted on the platform

Key features:

- Appears to increase risk of **exposure** to content
- Appeared more often in **isolated** routes to harm
- Often exacerbated by:
    - **User circumstances:** isolation or lack of social support; the mental health of the user

For a case study related to this risk factor, see 'Case study 12: Michelle' in the annex.

In instances where users could be easily contacted on the platform, it appeared to be more likely there would be **exposure** to potentially harmful content / contact. This refers to platforms which allowed users to be contacted by individuals unknown to them. For example, there were experiences where people had received abusive messages, sexual images or scam messages from strangers.

People felt that distributors were less likely to censor what they said or sent when they were able to contact users without going through a formal connection process (e.g. having to approve a friend request). This factor links in with the *How easily accounts can be disposed of and remade on the platform* factor below, as being able to contact people easily with little risk of reprisal creates an environment where potentially harmful content / contact can be distributed.

[7] https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

21. How easily accounts can be disposed of and remade on the platform

Key features:

- Appeared to increase risk of **exposure**
- Appeared most commonly in **isolated** and **indirect** routes to harm

For a case study related to this risk factor, see 'Case study 24: Connie' in the annex.

Users may not place much value on the accounts and feel little accountability for what they do when using disposable or "burner" accounts, i.e. in online environments where it is acceptable or easy to set up and delete accounts. If these users do not fear reprisal if they were to be involved with something negative or harmful, other users appear more likely to be at risk of exposure to content or contact.

For example, people reported experiences when they had interacted with content / contact from anonymous accounts, those with incomplete profiles, or from people who were identifiable but using accounts that were quickly deleted. These accounts distributed content / contact such as abusive messages, or unexpected sexual images. People felt that these distributors would not feel comfortable distributing content / contact like this if their accounts were not easily disposable.

Researchers also heard an example where an account had been set up to impersonate a user. This account was used to spread rumours about the user and to try to damage their reputation. The ability to create the account contributed to the distributor being able to impersonate the user.

22. How the platform indicates popularity and status

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared most commonly in **isolated** and **cumulative** routes to harm
- Often exacerbated by:
    - **Content / contact**: Content that is popular holds social status
    - **Distributor characteristics:** The emphasis on gaining status via indicators of popularity increased the potential for users to experience harm

For a case study related to this risk factor, see 'Case study 25: Aria in the annex.

On several platforms, content creators are incentivised to seek high engagement from other platform users. Features such as 'liking', 'sharing' and other indicators of popularity encourage users to try to distribute content that receives lots of likes, shares or re-posts, thereby increasing the status of the distributor. The existence of these features, along with algorithms which elevate popular content, appeared to increase the potential for content / contact to be **exposed** to platform users.

For example, many people commented that others had posted shocking or violent content that the distributor knew would receive high engagement, but that was potentially harmful to other users.

Researchers also saw that users sometimes experienced **harm** if the content they distributed did not receive enough status (e.g. few likes or comments). This was especially true when people were posting photos of themselves. The emphasis on gaining status via indicators of popularity increased the potential for users to experience harm.

23. How easily users can be linked / signposted to other content or platforms

Key features:

- Appeared to increase risk of **exposure**
- Appeared most commonly in **isolated** routes to harm
- Often exacerbated by:
    - **User characteristics**: Age of user. Links that took users to other platforms could put children at risk of exposure to content that was restricted to them under the guidelines of the first platform

For a case study related to this risk factor, see 'Case study 26: Hillary' in the annex.

When platform design permits or encourages the use of links to content / contact outside the platform, it was more likely that users would be **exposed** to more content / contact which may not fall under the community guidelines of the first platform.

This is observed in Ofcom's research into the risk factors that may lead children to harm online (scheduled for publication in October 2022), for example when children clicked on links which took them to sexualised versions of the original content. It was also observed to a lesser degree in the adult sample, for example when posts linked to other more niche posts in areas of a platform that users felt were less likely to be moderated. It can be hypothesised that this factor would be relevant in other situations too, for example when there are links to scam content.

24. Platform tolerance

Key features:

- Appeared to increase risk of **exposure**

For a case study related to this risk factor, see 'Case study 22: Jemma' in the annex.

When platforms had guidelines or a wider culture that made it acceptable or easy to distribute potentially harmful content or contact, it appeared to make **exposure** more likely, as users feel it is permissible to distribute content without repercussions.

Some respondents reported experiences seeing high volumes of similar content / contact online (for example abusive language, or 'pile-ons'), and reported that it is often unclear whether the platforms were doing anything about it.

How these link to platform functionalities listed in the Online Safety Bill

The current version of the Online Safety Bill defines functionalities of online platforms as any feature that enables interactions of any description between users of the service by means of the service, and as specifically including a list of particular features.[8] The list is not exhaustive. These specified functionalities for user-to-user services are listed in the table below along with commentary on relevant risk factors that were identified during this research, where applicable. It is important to note that it cannot be concluded from this research that when these functions are present, they always increase risk. Equally, some or all may have substantial user benefits. The impact is likely to depend on the presence of other risk factors.

[8] Functionalities is defined in Clause 189 in the Online Safety Bill: https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf

Functionalities listed in the Online Safety Bill for user-to-user services, at time of publication[9]	**Risk factor this functionality is an enabler of**
(a) creating a user profile, including an anonymous or pseudonymous profile;	See *21.How easily accounts can be disposed of and remade on the platform* and *20. How easily users can be contacted on the platform* under Platform characteristics
(b) searching within the service for user-generated content or other users of the service	See *17. Active user engagement with the content / contact* under User exposure
(c) forwarding content to, or sharing content with, other users of the service	See *17. Active user engagement with the content / contact* under User exposure
(d) sharing content on other internet services	See *17. Active user engagement with the content / contact* under User exposure
(e) sending direct messages to or speaking to other users of the service, or interacting with them in another way (for example by playing a game)	See *20. How easily users can be contacted on the platform* under Platform characteristics
(f) expressing a view on content, including, for example, by— (i) applying a "like" or "dislike" button or other button of that nature, (ii) applying an emoji or symbol of any kind, (iii) engaging in yes/no voting, or (iv) rating or scoring content in any way (including giving star or numerical ratings);	See *22. How the platform indicates popularity and status* under Platform characteristics and *2. Content that is perceived as aspirational or holds social status* under Content / contact factors
(g) sharing current or historic location information with other users of the service, recording a user's movements, or identifying which other users of the service are nearby	See *3. Content that is geographically local* under Content factors and *8. Distributor who is geographically close* under Distributor characteristics
(h) following or subscribing to particular kinds of content or particular users of the service	See *2. Content that is perceived as aspirational or holds social status* under Content factors and *9. Distributor who holds status or a reputation* under Distributor characteristics and *17. Active user engagement with the content / contact* under User exposure
(i) creating lists, collections, archives or directories of content or users of the service	No evidence from this research related to this functionality
(j) tagging or labelling content present on the service	See *5. Content that is personal or relatable* and *3. Content that is geographically local* under Content factors
(k) uploading content relating to goods or services	No evidence from this research related to this functionality
(l) applying or changing settings on the service which affect the presentation of user-generated content on the service;	No evidence from this research related to this functionality

[9] https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf

(m) accessing other internet services through content present on the service (for example through hyperlinks)	See *23. How easily users can be linked / signposted to other content or platforms* under Platform characteristics

Factors for further consideration / research into platform characteristics

- **Closed groups.** This was not specifically evidenced across our sample. There was evidence of harm occurring in both public and closed groups. This does not necessarily mean these factors are not relevant in individual cases or types of harm. It can be hypothesised that closed groups may lead to a higher likelihood of harm occurring if users do not think others are moderating or watching conversations.
- **Lack of protection measures** (e.g. friction, for example prompts that warn people before posting or forwarding content or following a link)**.** Participants were not always aware of the existence of protection measures, It is hard to observe the absence of something, and this would have been challenging for people to reflect on or be aware of. Other evidence suggests that the lack of friction can contribute to harm.
- **Recommendation systems / algorithms.** It can be hypothesised that recommender algorithms increase the likelihood of **exposure** to potentially harmful content. A user's online actions, or the online actions of other users can influence the behaviour of a recommender system and hence the content that is pushed to users. Some people did mention algorithms potentially playing a role in their experience of harm, but this was challenging to evidence specifically within this research as the ambiguity of how platform algorithms work made it hard for people to report on.

Risk factors associated with the actions of other users

The factors below were observed to increase the potential that people would be exposed to content, and/or experience harm, if other risk factors were also present.

Note: The impact of other users was not directly apparent to the research participants and is broader than a single experience; therefore, there is limited data from this research related to this component.

25. Validation / approval of other content

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared most commonly in **isolated** and **cumulative** routes to harm
- Often exacerbated by:
    - **Distributor characteristics:** The emphasis on gaining status via indicators of popularity increased the potential for users to experience harm

For a case study related to this risk factor, see 'Case study 19: Kathleen' in the annex.

When other platform users seem to validate or approve content, it appeared to make it more likely that a user would be **exposed** to content / contact, or increased the potential for content / contact to lead to **harm**.

For example, if certain types of content are viewed as acceptable by other platform users, this type of content may be more likely to be posted or sent, or less likely to be reported. This may increase the amount of related content on the platform, and therefore the likelihood of a user being exposed to it.

There were examples where other platform users giving weight to content meant it increased the potential for that content to lead to **harm**. For example, when a distributor posted an abusive message, and other platform users 'liked' that message, the user was more affected by that content, because it felt like multiple other users had the same opinion as the original distributor.

26. Peer pressure or recommendations

Key features:

- Appeared to increase risk of **exposure**
- Appeared to increase risk of **harm**
- Appeared most commonly in **cumulative** routes to harm
- Often exacerbated by:
    - **User circumstances:** isolation or lack of social support; the mental health or self-esteem of the user
    - **User characteristics:** age of user

For a case study related to this risk factor, see 'Case study 25: Aria in the annex.

In this context, 'peer pressure' means the pressure users feel to engage with certain things or behave in certain ways online as a result of exposure to content / contact from people within their social network.

In terms of **exposure**, there were experiences when users were encouraged by others to look at content, either explicitly or because users thought it was what other people did or they felt that engaging with content would make them 'cool.'

In other situations, users were explicitly encouraged to carry out unhealthy behaviours, for example engaging in excessively weighing or reducing food intake. It can be hypothesised that peer pressure may also be relevant in situations where users are encouraged to carry out dangerous stunts or illegal activities, although this wasn't something seen in this sample.

The likelihood of **harm** occurring in situations where there is peer pressure could intersect with age. It could be hypothesised that harm is more likely to occur in younger or vulnerable individuals. For example, in Ofcom's forthcoming report on *Risk factors that may lead children to harm online,* it is observed that peer pressure was a factor in young people wanting to have certain apps or games, or be a part of certain online communities because their friends were doing similar things. The fear of missing out meant some young people stayed in, or spent more time in online communities than they otherwise would have liked. Some reported disliking the amount and concentration of potentially harmful content that existed in the online communities they were in, but felt that to *not* be involved in these online communities was to miss out socially.

Factors for further consideration / research into other platform users

- **Norms on the platform.** This connects to 'Platform tolerance for distribution of potentially harmful content'. As with many of these factors, it may be difficult for users to identify this themselves.
- **Homogenous users/no challenge.** This would not necessarily be apparent to the user.
- **Other platform users engaging with potentially harmful content.** Again, this would not be easy for a user to discern, but the impact of other users engaging has the potential to impact recommender systems used by platforms. As such, it is likely to lead to increased exposure for other users.

Further traditional consumer research is unlikely to illuminate the importance of these factors. To explore them further will require a wider suite of innovative research tools. Additionally, stakeholder interviews may provide evidence to support the hypotheses that these factors may be important.

Conclusions

Considerations for the model based on this research

Prior to the research, Ofcom had an existing draft of the model. This research was commissioned to inform the design of that model. Throughout the research, Revealing Reality conducted multiple sessions with Ofcom (detailed in the methodology annex) to go through the findings from the interviews and use them to test and refine the model - to bring it to the current version outlined earlier in this report. This section outlines a number of considerations Revealing Reality thinks are raised by the research for the model that could inform its development and ongoing use.

The model focuses on psychological and physical harm, and could do more

The model focuses on **psychological** and **physical** harm as these are explicitly referenced in the current version of the Online Safety Bill.

However, the research raised the possibility of other types of harm in the experiences of people who took part, including economic. For example, some people suffered financial loss as a result of online scams. Another form of possible economic harm that was raised in the research is the opportunity cost, or lost productivity, caused by excessive time spent online.

Societal harm is also likely to have occurred through the course of many of the stories reported on in this research – e.g. the impact of misinformation or increased distrust. However societal harms are much harder for research participants to self-report, which is why they have not been documented in this report.

While the current iteration of the model does not account for these other types of harm, there is no obvious reason why it could not in the future if required

The model works equally well for legal and illegal content online

The research included people who had experienced harm from being exposed to **legal** content, as well as content the research team considered likely to be **illegal**. In fact it was not always possible for the research team to know for definite whether the content being described would have been illegal. However, even in the clearer cut cases, there appeared to be no obvious difference in terms of the corresponding risk factors surrounding the legal or illegal content / contact.

The model works well for isolated routes to harm, and can also account for cumulative routes

The model works well for **isolated** routes to harm by describing the nature of the isolated hazard/s in the content / contact category.

The model can also account for **cumulative** routes to harm through the "user exposure" category of risk factors (e.g. acknowledging the other experiences and actions of the user on the platform) although it is potentially less intuitively structured for this route. Given that the cumulative route was often seen to lead to more severe harm than isolated experiences, this is worth Ofcom taking into consideration when using the model.

The model does not account for indirect routes to harm

The model sought to explain how exposure to content / contact can lead to an individual experiencing harm. However the research also documented examples where people reported that they had been harmed by

content / contact relating to them being exposed to other people, for example in the case of nude images being shared without consent. The model therefore does not account for the **indirect** route to harm.

The model does not imply which risk factors are most significant

As presented, the model does not suggest any weighting or sense of scale to any one category of risk factor. This research cannot evidence conclusively which categories are likely to be the most significant or prevalent. However, given the cumulative route to harm where an individual is exposed to multiple lots of potentially harmful content over time which can involve them actively engaging with them was observed to often result in the more severe harm, it seems likely that the "user exposure" category will emerge as an important space to explore further.

The evidence base for documenting which risk factors are the most significant and prevalent can be built using this model as a framework.

The model does not account for the positive outcomes that occur because of online experiences

The model was created to show how 'significant harm' manifests. Therefore, it does not account for the potential for positive outcomes to occur as well as negative because of online experiences. This research showed that similar online content / contact can harm one person and be instructive or useful to others (e.g. in the 'unharmed' control group'.

It is important that this potential neglect of positive outcomes is taken into account when using the model.

Research gaps and future opportunities

This research has helped to further understanding of risk factors which contribute to harm manifesting online.

However, this is a complex topic, and not all areas of interest could be fully covered in the scope of this methodology, which focused on people's own awareness of the factors that brought them to harm. For example, this research was designed to identify platform characteristics associated with higher risks of harm, but it was not able to identify platform characteristics that may reduce risks of harm. The research also revealed areas and hypotheses that may need to be explored further in the future.

The model as it stands does not easily accommodate route 3 (content exposed to other people). A separate piece of research may be required to investigate this route further, including potentially exploring risk factors related specifically to this route.

This research also did not identify substantial data related to route 1b: 'isolated, delayed harm'. This is likely due in part because the research design specified that people had been exposed to content / contact within the last six months to ensure people could recall the exposure in detail. It may also be harder for users to identify that they have experienced harm via this route. This may warrant further research in the future to identify if the risk factors identified in this report remain true for this route, or whether there are additional factors to consider.

Because the recruitment criteria required people to be aware they had been harmed (except in the control group) people who might have been subject to mis/disinformation in the sample were not included. A different study design is required to understand how harms manifest for this audience, given they may not self-identify as being harmed.

There were also additional hypotheses around risk factors under many of the components. These include, but are not limited to:

- Whether particular genres of content would make it more likely that a user would experience harm.
- Whether certain protected characteristics beyond age might make someone more at risk, particularly to specific things that relate to that characteristic (e.g. racism, homophobia).
- Whether malicious intent by the distributor and persistence of the distributor would make it more likely that a user would experience harm.
- Whether time spent online would be a factor in likelihood of experiencing harm.

All of the above points are hypotheses that were not evidenced throughout this research, but may warrant further, more targeted research in the future. In particular, it would be useful to explore developing methods to overcome the challenges of relying on self-reported harm, as there are important factors in the routes to harm which people are less able to reflect on.

Additionally, as noted above, this research reflects current experiences of online harm. The online world is constantly changing and as such it is reasonable to assume that online harms will do so too. Therefore, it will be valuable to periodically refresh this research to ensure the model continues to hold true for new and emerging harms, or to change the model as necessary to accommodate them.

Annex: Case studies

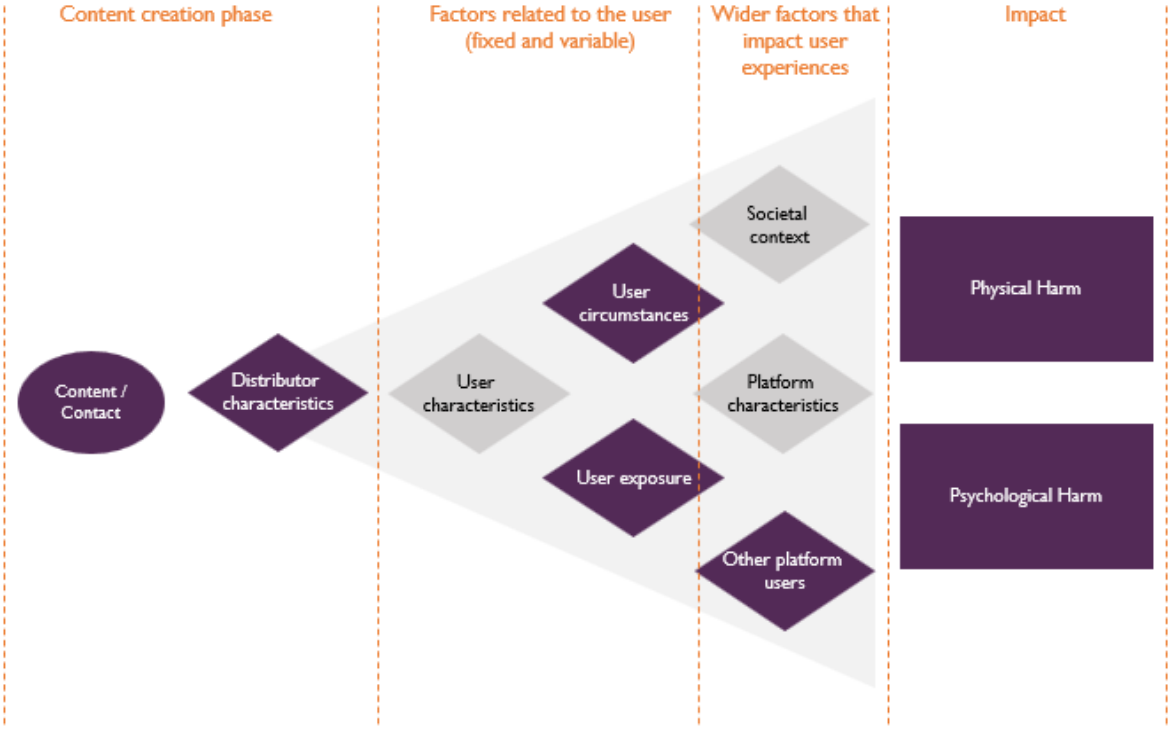
Case study 1: Scarlett (legal hazard, harmed)

Eating disorders

Scarlett (26-30) has suffered with an eating disorder for the past eight years, which flares up when she is feeling isolated or unhappy with other elements of her life. When she is struggling, she is drawn to seek out pages which offer assistance and tips about how to lose weight, reduce calorie intake, and hide her disordered eating from friends and family members. When she first developed her disorder, she sought out pro-anorexia websites. Now, she says they are a lot harder to find and you will be directed to mental health support websites instead. She searches for most of the content on online forums and an image-sharing social media site.

She described this content as "helpful" in the moment but ultimately more damaging for her as they contribute to worsening her eating disorder.

Other factors also exacerbated the harm Scarlett said she experienced. Some of these pages are designed to be fitness accounts where people's weight-loss is celebrated by the content distributor and other users. Other accounts were targeted at people with eating disorders, where restricting eating is condoned by other users in the space. Scarlett also felt isolated and lacked support networks during her most recent flare-up.



Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix' • Perceived as aspirational or holds social status	**Distributor characteristics** • Holds status or a reputation	**User characteristics**	**User circumstances** • Isolation or lack of social support • The mental health of the user • Past trauma or experiences of the user
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users** • Validation / approval of content

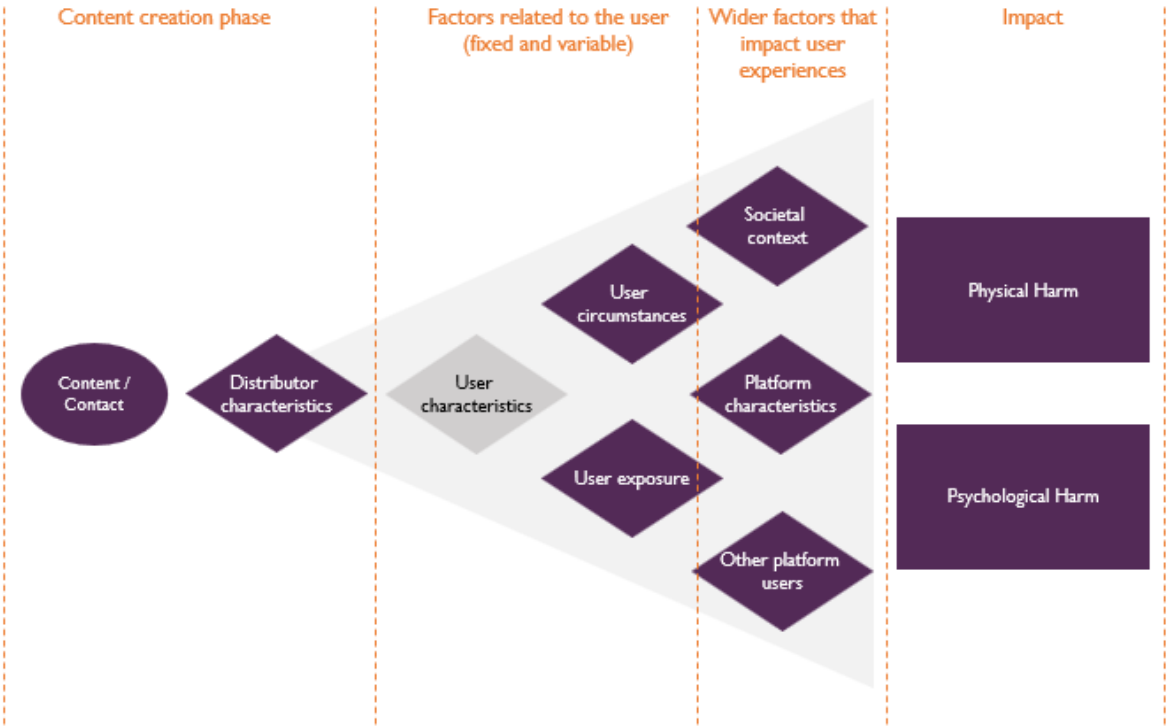
Case study 2: Laura (legal hazard, harmed)

Eating disorders

Laura (18-20) suffered with an eating disorder at the end of her time in college. She found that during her illness, content on her social media included lots of fitness and calorie counting videos (e.g. "What I eat in a day").

Laura found that this positive content harmfully impacted her, encouraging her disordered eating habits by glamourising unhealthy body images.

Other factors exacerbated the harm Laura felt she experienced. Some of the content, like fitness accounts, were condoned by other users and portrayed weight loss positively. Content was sometimes posted by influencers with online status who had thousands of followers, enhancing the aspirational nature of the content.



Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix' • Perceived as aspirational or holds social status	**Distributor characteristics** • Holds status or a reputation	**User characteristics**	**User circumstances** • The mental health of the user
User exposure User experiences: • Seeing lots of similar content User actions: • Active user engagement with the content/contact	**Societal context** • Acceptability of posting content • Trending topics	**Platform characteristics** • How the platform indicates popularity and status	**Other platform users** • Validation / approval of content

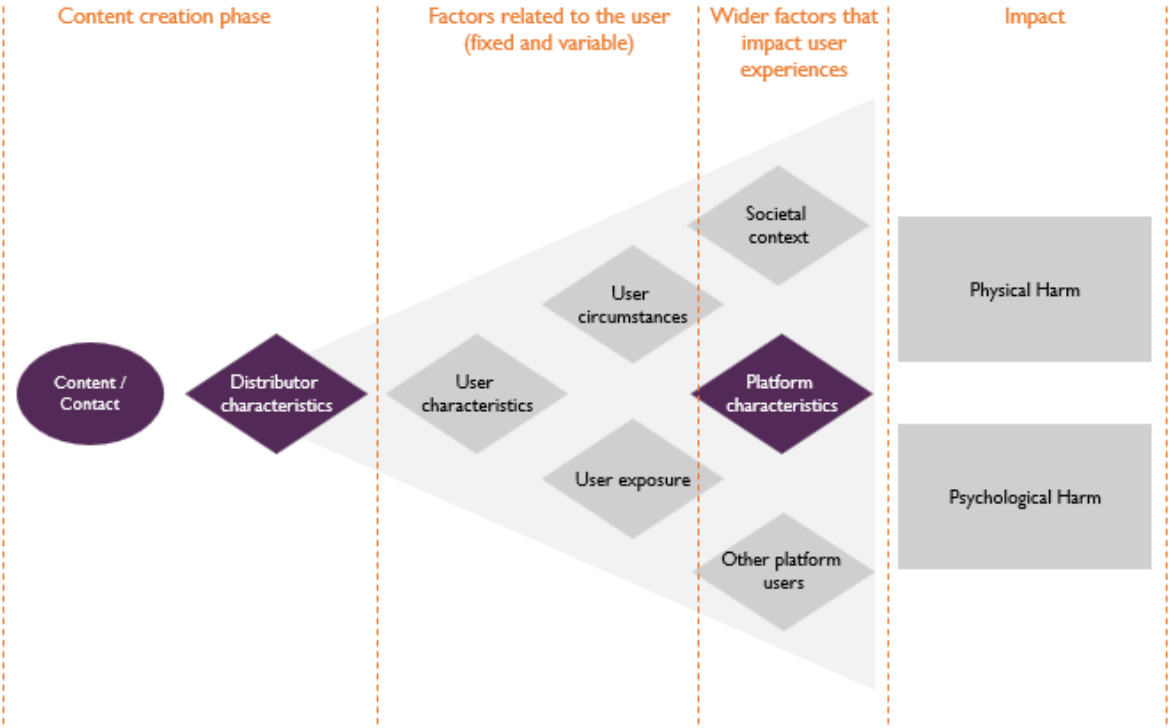
Case study 3: Maya (legal/possibly illegal hazards, unharmed)

Maya (18-20) was connected to several friends/acquaintances from her hometown on an image-sharing social media platform. She connected with these people via a function that enables users to add people they don't know from their local area.

Often, these contacts would share content showing people selling drugs in her hometown.

*Although she was not harmed by the content she saw, being able to add contacts in her geographic area increased the likelihood of being **exposed** to content relating to her local area. She could see how it might make her or others uncomfortable about going to certain places in her hometown.*

N.B. The function that enables users to add people they don't know was frequently used by children who had experienced harm in the Ofcom's forthcoming report 'Risk factors that may lead children to harm online'.



Risk factors

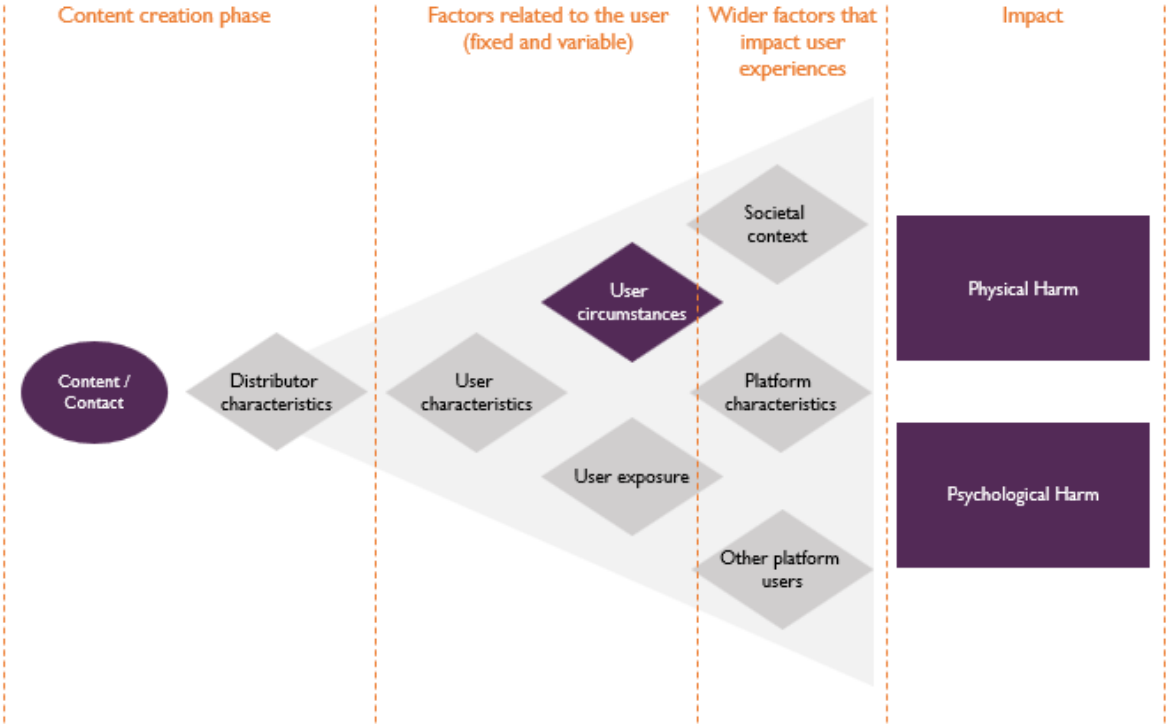
Content / Contact • Geographically local	**Distributor characteristics** • Known to the user • Geographically close	**User characteristics**	**User circumstances**
User exposure	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform	**Other platform users**

Case study 4: Simon (legal hazard, harmed)

Gore content

Simon (26-30) used to browse niche online discussion boards around Game of Thrones and certain anime. One day, he clicked on a piece of content about Game of Thrones, and it went to a page with gore content on it – a picture of dismembered bodies. Simon had no reason to believe that these bodies weren't real and found the experience extremely shocking.

Other factors exacerbated the harm Simon said he experienced. He spent a lot of time online because that's where he fulfilled most of his hobbies. He didn't have a large social circle and had been experiencing mental health issues for which he had been receiving counselling. Despite having felt better in recent months, seeing this content made his mental health decline again and he felt anxious, revolted and even guilty.



Risk factors

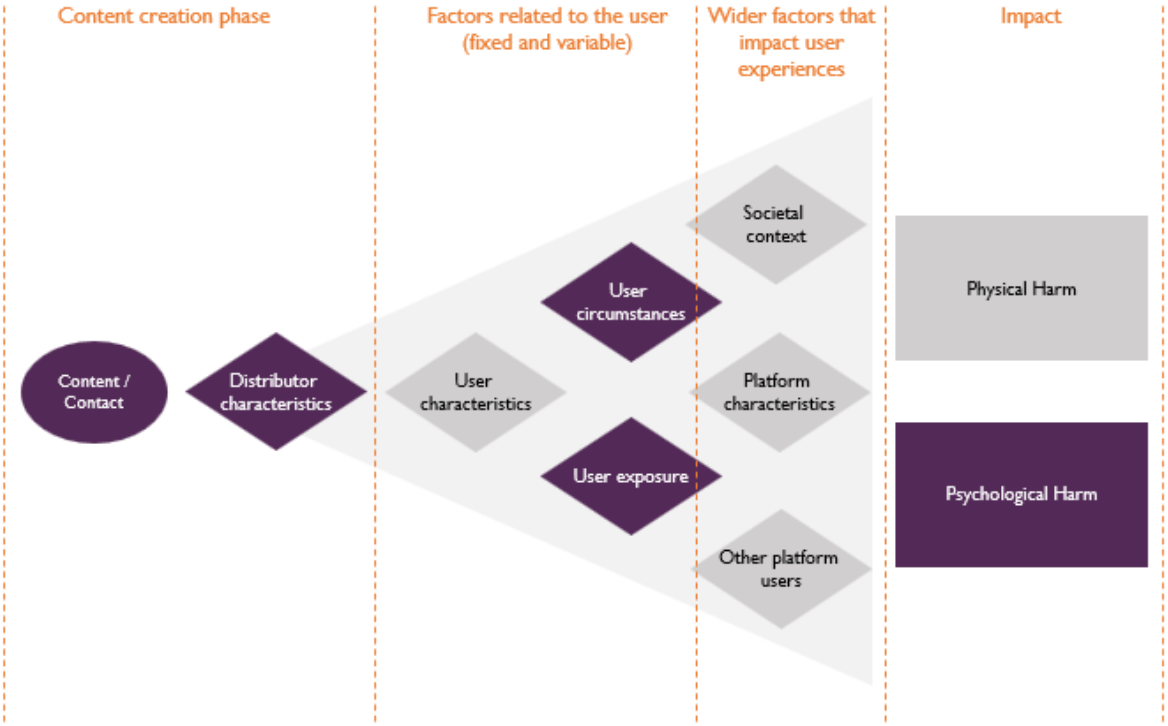
Content / Contact • Perceived as real / genuine • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space • Isolation or lack of social support • The mental health of the user • Past trauma or experiences of the user
User exposure	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 5: Tala (legal hazard, harmed)

Tala (21-25) sees a lot of glamourised lifestyle content shared by influencers who are similar in age to her and who she sees as having status. She engages with the content, liking and following people so that she sees a high volume of content like this in her feed.

She compares herself and her life to theirs, and finds it lowers her self-esteem and makes her question her own success.

Other factors exacerbated the harm Tala felt she experienced. She has recently moved back in with parents in a rural place away from her friends and feels quite isolated. She spends a lot of time online as a result of not having other activities to do or people to meet. Tala has also experienced periods of poor mental health in the past related to her body image.



Risk factors

Content / Contact • Perceived as aspirational or holds social status • Relatable or personal	**Distributor characteristics** • Holds status or a reputation	**User characteristics**	**User circumstances** • Isolation or lack of social support • The mental health of the user
User exposure User experiences: • Seeing lots of similar content User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users**

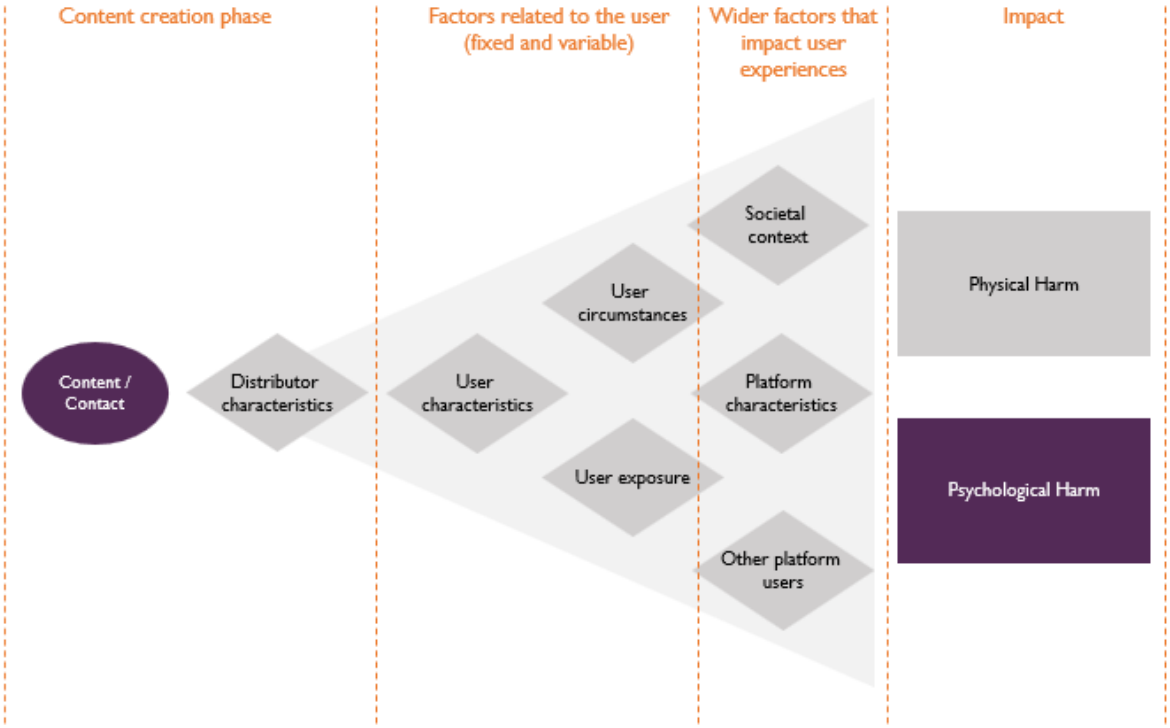
Case study 6: Luke (legal hazard, harmed)

Physical violence

Luke (41-45) saw a beheading video on a video-sharing platform. Usually, he watches videos of concerts from his favourite bands from the 90's and videos he finds funny. This video was different to the things he usually watches, but he clicked the content as he expected it to be a recommendation that was similar to content he enjoyed watching normally.

Luke was shocked, distressed, and felt physically sick after watching the clip. He has not used the video-sharing platform since he saw the video three months before the interview.

In this case, the unexpected factor within 'content' was a large reason why he believed he experienced harm. The harm was also exacerbated by the fact he believed the video was real.



Risk factors

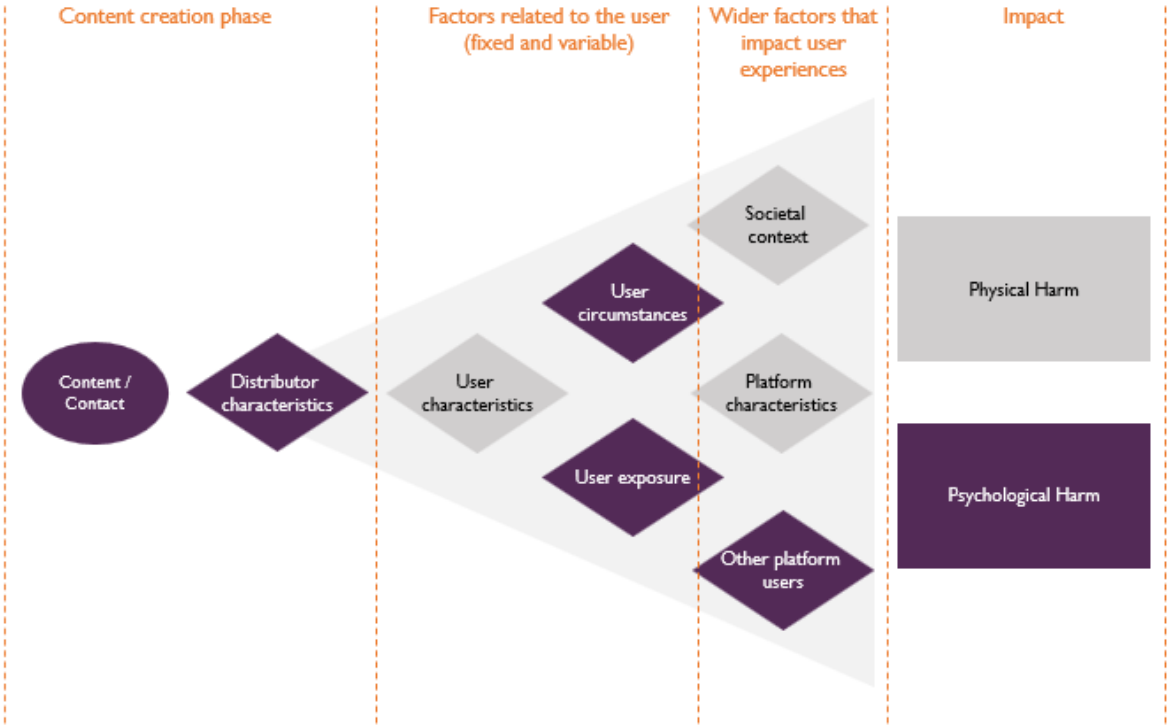
Content / Contact • Perceived as real / genuine • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances**
User exposure	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 7: Adil (illegal/likely to be illegal hazard, harmed)

Abusive and threatening comments and posts

Adil (26-30) received homophobic messages, threats, and abuse after coming out as bisexual on a social media platform. A lot of the comments were from people he knew through the South Asian society at university.

He felt ashamed and worried for his daughter, who was often featured in the abusive messages. He suffered from depression and anxiety following the incident and sought counselling.



Risk factors

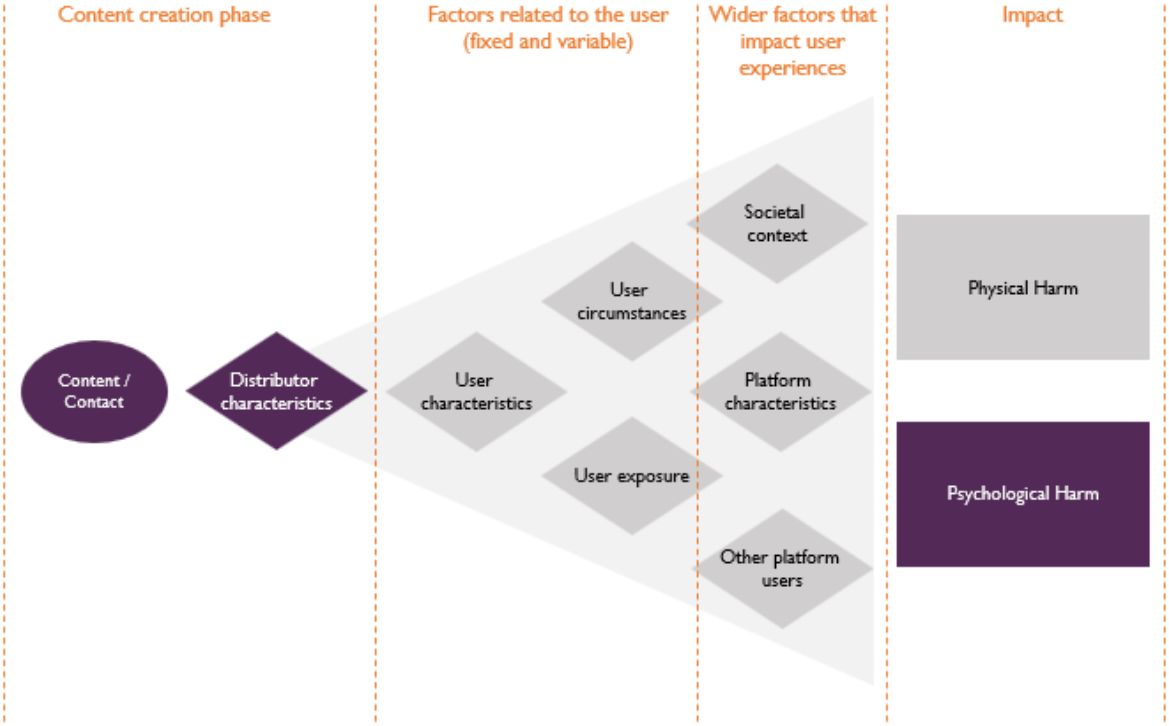
Content / Contact • Relatable or personal	**Distributor characteristics** • Known to the user • Geographically close	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space
User exposure User experiences: • Seeing lots of similar content	**Societal context**	**Platform characteristics**	**Other platform users** • Peer pressure or recommendations

Case study 8: Shruti (illegal/likely to be illegal hazard, harmed)

Abusive and threatening comments and posts

Shruti (26-30) was a victim of cyber-stalking and abuse from her ex-partner. He lived nearby and threatened to visit her house and reveal the details of their romance to her family, a strict conservative family who would not have approved of the relationship.

She felt terrified that he would reveal details of their relationship, and her anxiety affected her performance at work.



Risk factors

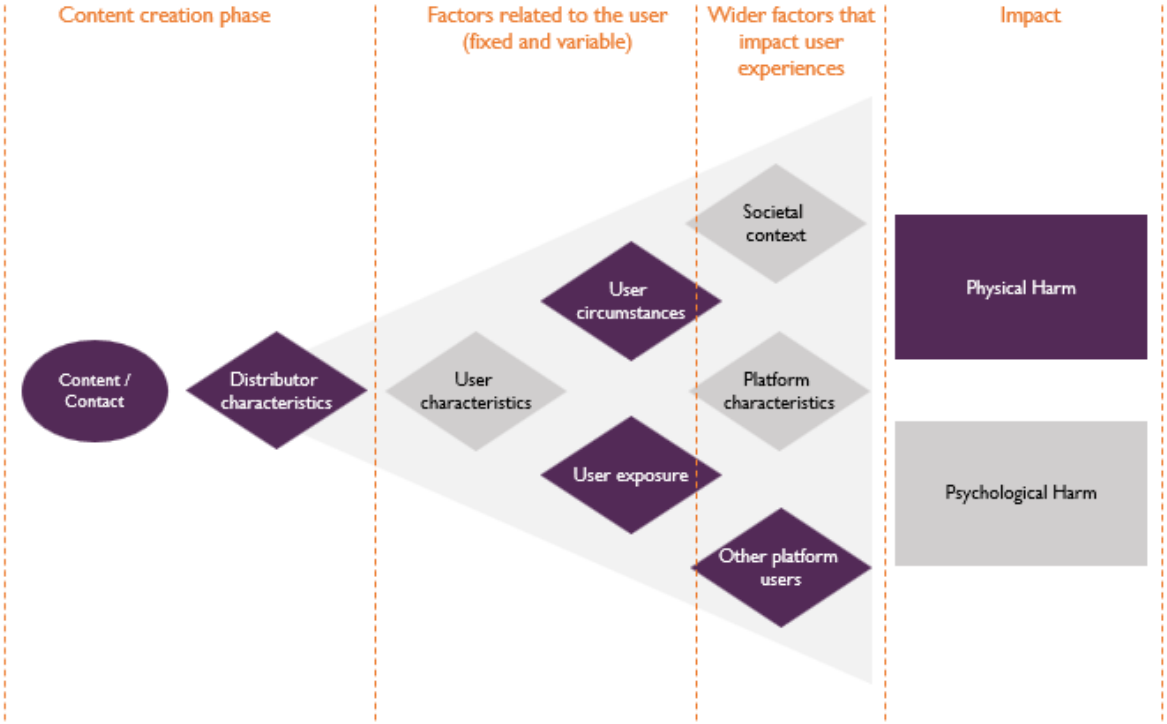
Content / Contact • Relatable or personal	**Distributor characteristics** • Known to the user • Geographically close	**User characteristics**	**User circumstances**
User exposure	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 9: Humza (legal hazard, harmed)

Humza (18-20) enjoys going to the gym as his main hobby. Outside of this, his main forms of socialisation happened online. Humza began to engage heavily with people sharing steroid-related content. On the forums, he described several authoritative figures who shared knowledge about how to use steroids to get the desired results. Often, these users were older and held status within the forums - Humza described these characters as acting as "self-elected moderators."

Humza's initial curiosity developed into a desire to take steroids. He eventually tried steroids himself, but his body reacted badly. He suffered from gynecomastia and depleted levels of testosterone, which he had to receive treatment for.

Other factors exacerbated the harm Humza felt he experienced. For example, his limited social circle meant he was not exposed to counter narratives and had few people to step in before he tried the steroids.



Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix' • Perceived as aspirational or holds social status	**Distributor characteristics** • Holds status or a reputation	**User characteristics**	**User circumstances** • Isolation or lack of social support
User exposure User experiences: • Not seeing counter narratives User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users** • Validation / approval of content

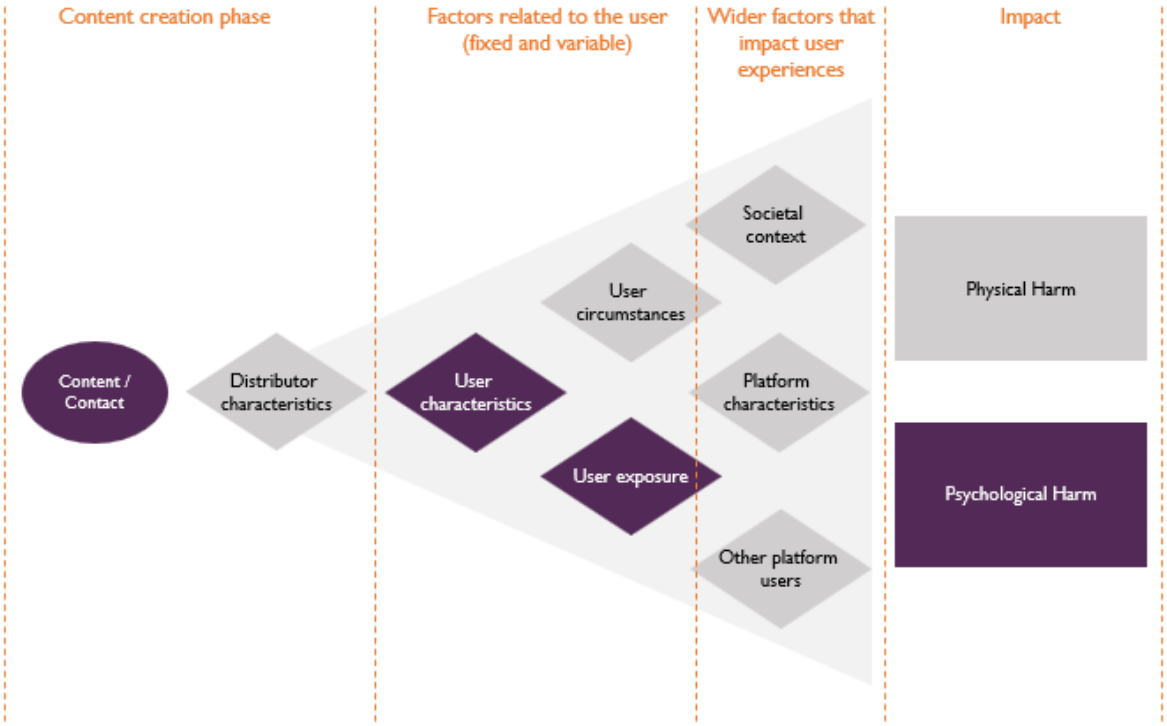
Case study 10: Ahmed (legal hazard, harmed)

Sexual violence

Ahmed (31-35) lives in London. He started watching porn when he was 12 years old. At the beginning, he mainly looked at images online. This progressed into watching what he described as "fairly normal porn." By the age of 17, Ahmed said he had grown bored of content like this and wanted to find things which were different. He started watching violent, hardcore porn, and continued to do so for ten years.

Ahmed reflected on the experience, stating that it was not a good idea. He believed that watching porn like this for such a long period of time had fundamentally shaped his view on relationships. He claimed he misunderstood the meaning of relationships, seeing them only as a pursuit of pleasure and sex, and used this to explain his inability to form a solid romantic relationship with anyone. He also said that his expectations of sex were warped to believe that everybody found pleasure in violence and pain during sex.

Although Ahmed continued engaging with increasingly violent pornographic content into adulthood, the gateway into hardcore porn arose in his youth and shaped the content he sought to consume in the following years.



Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix'	**Distributor characteristics**	**User characteristics** • The age of the user	**User circumstances**
User exposure User experiences: • Seeing lots of similar content User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users**

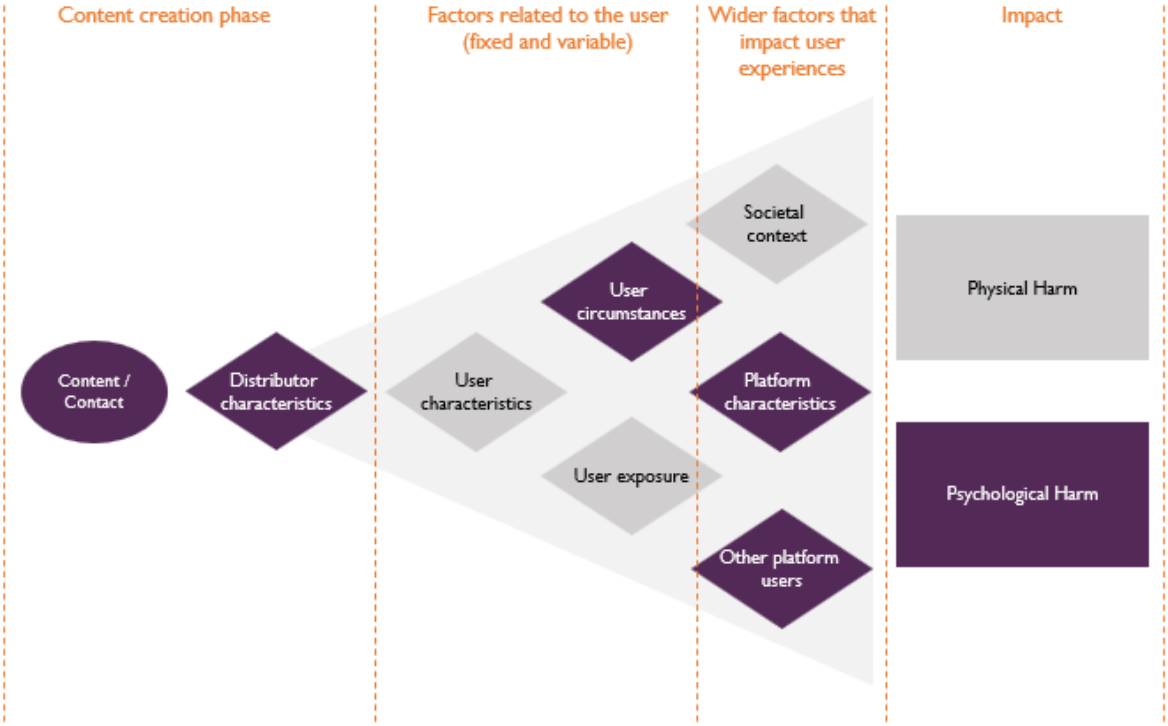
Case study 11: Lucy (legal hazard, harmed)

Abusive and threatening comments and posts

Lucy (26-30) relies on her income from livestreaming. She became a victim of an online scandal in the live streaming community and suffered several "Hate Raids" on her channel. Her fans also enacted retaliatory "Hate Raids" on other streamers' channels.

She said she suffered psychological, reputational, and economic harm. She wanted to leave the platform she used to stream and find another means of income but was reluctant as she was unsure how she would find an alternative income and has invested a lot in her current profession.

Other factors exacerbated the harm Lucy experienced. For example, she knew the people who caused the scandal. She also felt guilty about her fans retaliating against the content. She attributed their actions to the tribalism and status-politics of streaming culture, and felt let down by and isolated from the wider streaming community.



Risk factors

Content / Contact • Relatable or personal • Unexpected	**Distributor characteristics** • Known to the user • Holds status or a reputation	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space • Isolation or lack of social support
User exposure	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform • How the platform indicates popularity and status • Perceived permissibility of platform	**Other platform users** • Validation / approval of content

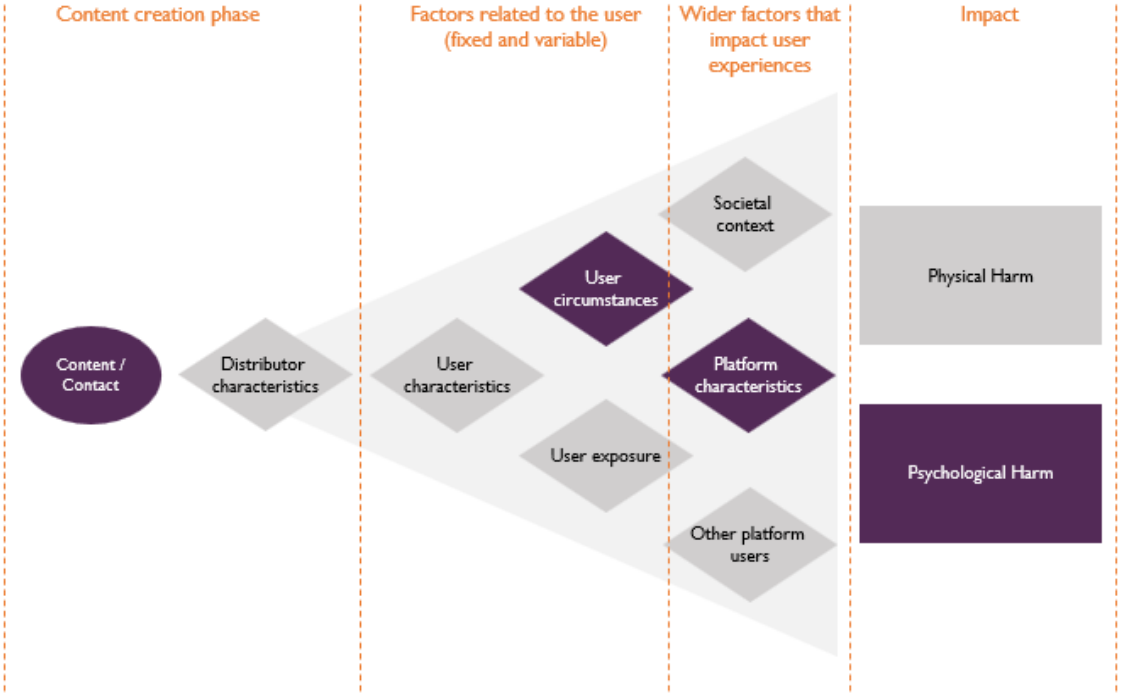
Case study 12: Michelle (legal hazard, harmed)

Sexual violence, misogynistic / sexist content

Michelle (21-25), lives with her husband and children, and has a history of sexual assault. She has a small social circle and gaming online is a big part of her social and family life: she and her family game together most evenings. They play an online multi-player game which also involves other players who are unknown to Michelle and her family, and who are often selected at random. All players can talk to each other in the chat function. She doesn't mind playing with strangers, although prefers it when she can play with users who she has met before online. The game has a voice-chat function designed to help players co-ordinate while playing.

Michelle reported receiving gendered abuse often over the voice-chat function from other players who are in the multi-player game on her team. She has frequently received derogatory comments about being a woman, as well as sexual threats and abuse. This usually escalated if her team was doing badly. With a history of sexual assault, Michelle finds this particularly harmful.

Michelle felt upset and disheartened by the abuse and worried about her children, especially her daughter, playing the games when sexist comments were frequently exchanged. For Michelle, however, the cost of withdrawing from gaming would be too great as she felt it was an activity the family did to spend quality time with one another. Her children now also love gaming, and she doesn't want to keep them out of that world.



Risk factors

Content / Contact • Relatable or personal • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space • Past trauma or experiences of the user
User exposure	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform • How easily accounts can be disposed of and remade on the platform	**Other platform users**

Case study 13: Jackie (legal hazard, harmed)

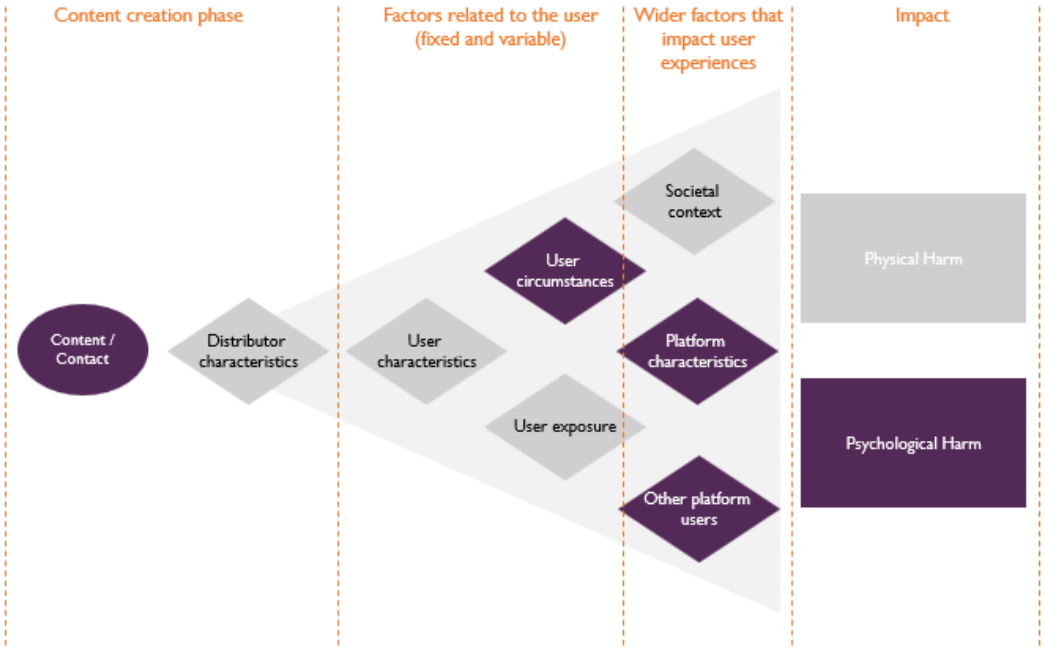
Physical violence

Jackie (51-55) is a single mum living with her son. She said she sometimes feels quite lonely and, at the time that harm occurred, was looking for ways to connect with people and spark social connections.

Her sister recommended she started using a platform for making connections. On the platform, users are encouraged to make connections with strangers by sending images and video clips which appear on the recipient's profile. Jackie said that some people use the website to make romantic and sexual connections, while others used it for socialising or meeting people from around the world.

Jackie didn't fully understand the nature of the platform when she started using it. During the interview, she reflected that the platform had a more sinister side to it. She felt it was dangerous to be able to connect with so many users, but at the time she joined the platform, she was eager to meet people and did not fully consider that there could be negative consequences to this.

*Jackie increased her likelihood of being **exposed** to potentially harmful content / contact by exchanging messages, images, and videos for several hours a day with people she did not know. Jackie began chatting to a user she did not know. The user then sent her a picture of animal abuse which opened on her home page. Because of the platform's design features, Jackie was not able to close or delete it. Jackie felt sick and shocked, and removed herself from the platform. During the interview, she said she still felt embarrassed that she had used the website, putting herself at risk of being contacted by strangers like this.*



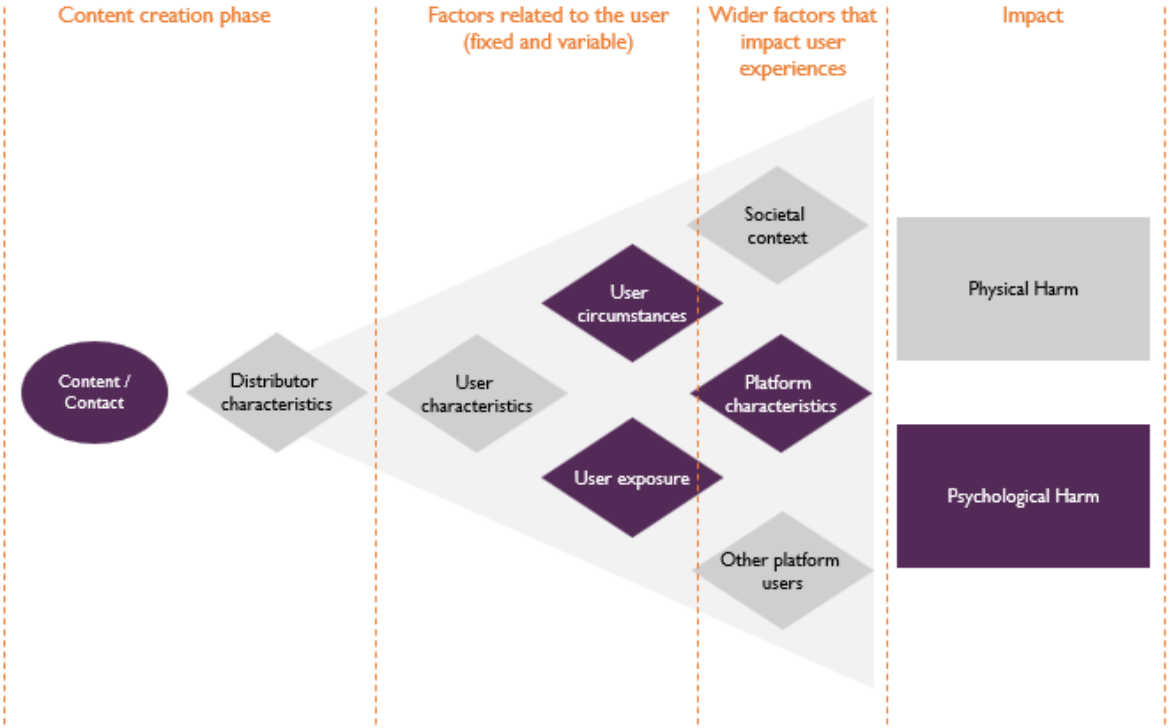
Risk factors

Content / Contact • Perceived as real / genuine • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • Isolation or lack of social support • The mental health of the user
User exposure	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform • Perceived permissibility of platform	**Other platform users** • Validation / approval of content

Case study 14: Neil (illegal/likely to be illegal hazard, harmed)

Financial scams

Neil (61-65) started a relationship with a French woman who approached him online, and later scammed him out of £3,000, making him feel hugely embarrassed and ashamed. Neil's isolation increased his likelihood of experiencing harm as he was more eager to build a connection and had few friends to challenge the relationship.



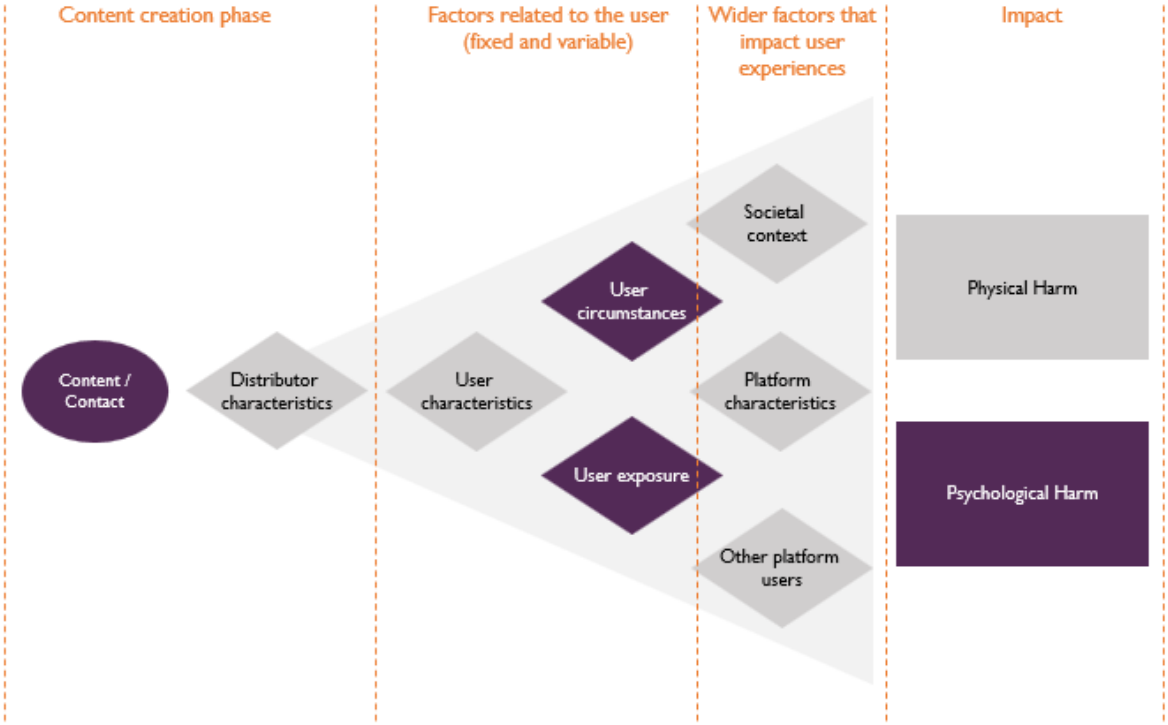
Risk factors

Content / Contact • Perceived as real / genuine • Relatable or personal	**Distributor characteristics**	**User characteristics**	**User circumstances** • Isolation or lack of social support
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform	**Other platform users**

Case study 15: Kevin (illegal/likely to be illegal hazard, harmed)

Suicide and self-harm content

Kevin (26-30) was struggling with his mental health earlier this year due to not being able to go outside during the Covid-19 lockdown. He was drawn to seek out content which gave instructions for suicide, including information encouraging suicidal actions. His mental state meant he sought the content out by directly searching for suicide methods using a search engine. Kevin by-passed the initial pages that showed support phonelines until he found forums that described suicide methods. His poor mental health increased his likelihood of ***exposure****, as it led him to seek out this content. It also meant he was more likely to experience* ***harm*** *from the content and the advice it gave.*



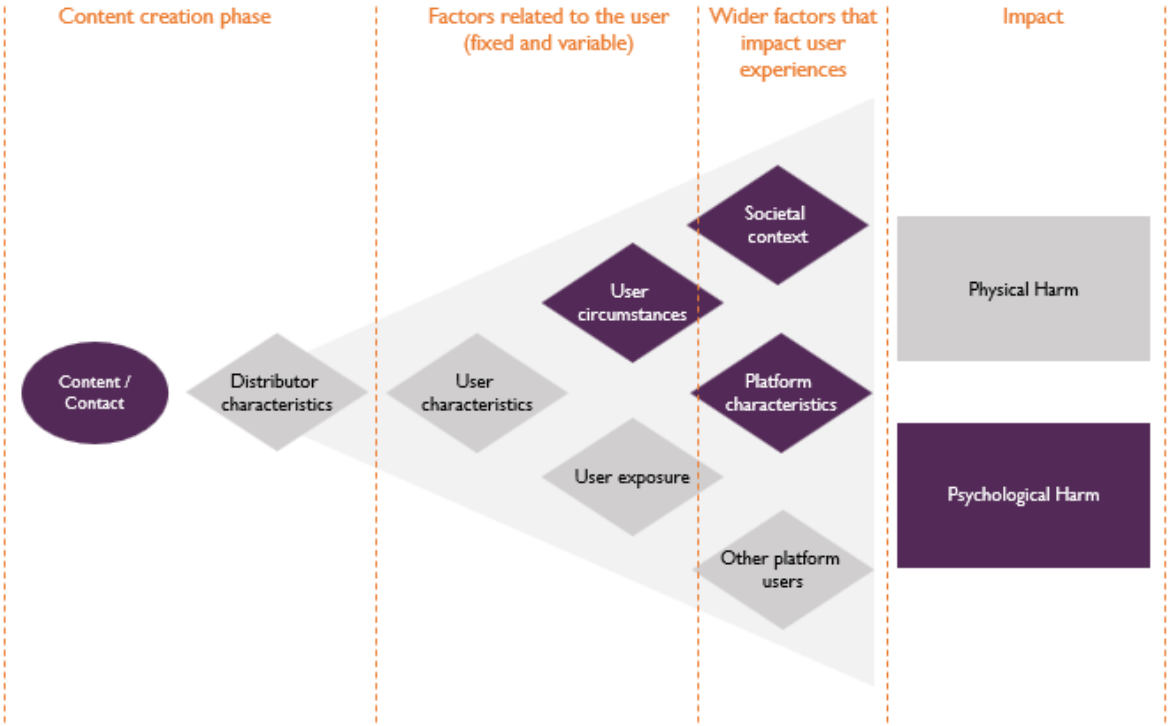
Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix'	**Distributor characteristics**	**User characteristics**	**User circumstances** • Isolation or lack of social support • The mental health of the user
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 16: Kylie (legal hazard, harmed)

⚠ Suicide and self-harm content, physical violence

Kylie (26-30) experienced emotional abuse when she was younger. At the age of 16, she was in an online relationship/friendship with someone she met online. They had an argument and this person made her believe they had taken their own life because of her. As a result, she is triggered by suicide content. There was a brief trend on a video-sharing platform where people were baiting users into watching a clip of someone shooting themselves in the head by showing it suddenly in a video of normal, mundane content. She saw this and it caused her to have a panic attack and feelings of anxiety.



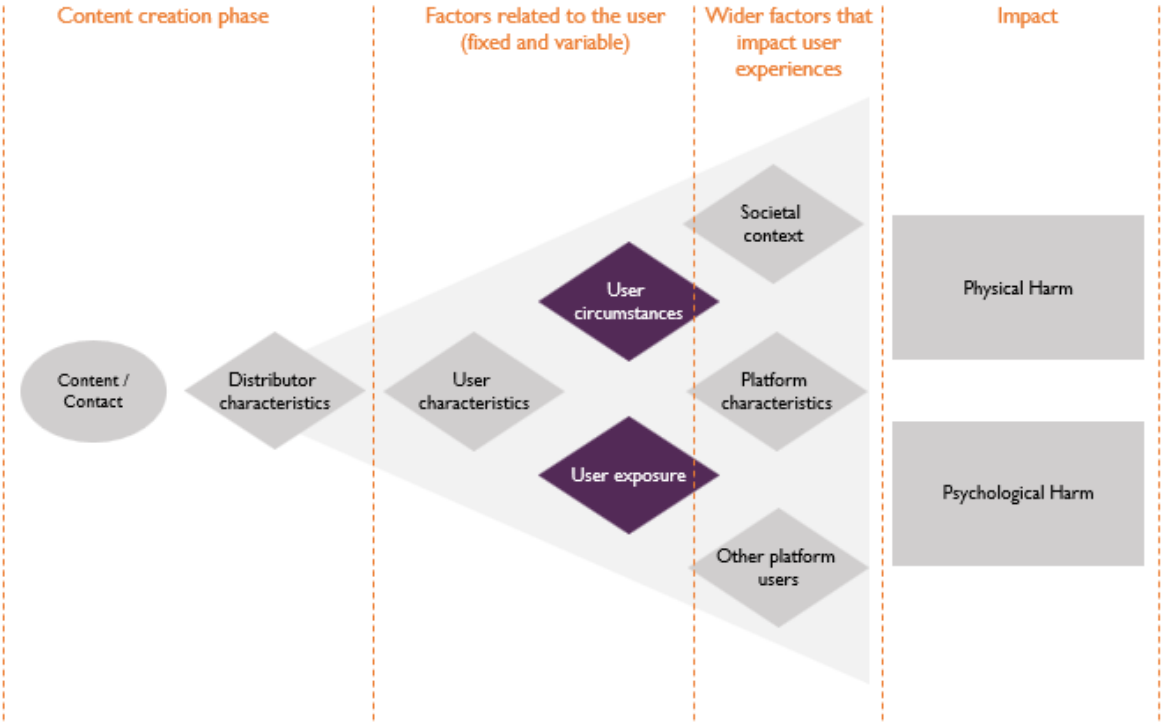
Risk factors

Content / Contact • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • Past trauma or experiences of the user
User exposure	**Societal context** • Trending topics	**Platform characteristics** • How the platform indicates popularity and status	**Other platform users**

Case study 17: Katya (legal hazard, unharmed)

Abusive and threatening comments and posts

Katya (36-40) was a former victim of domestic abuse via social media. As an activist and educator, she now comes across a large amount of content related to domestic violence, but this doesn't lead to her experiencing harm – she feels that as she is now advocating for domestic abuse survivors, her engagement with this content is in the context of positive action and being in control of the situation.



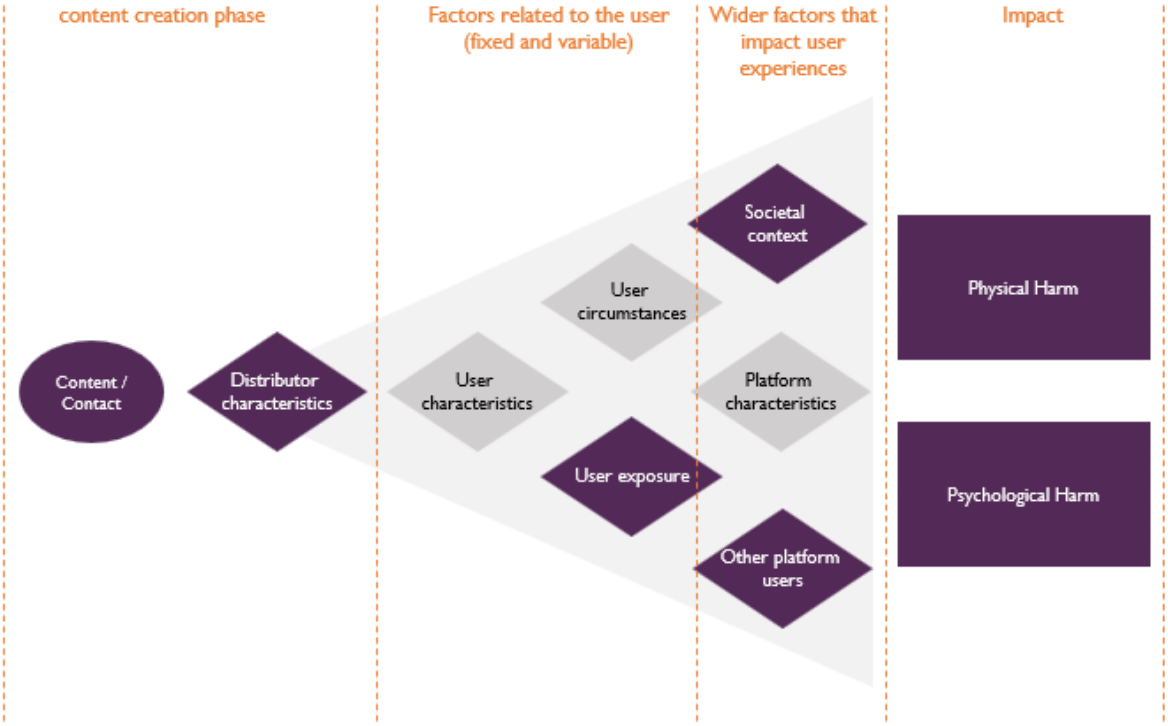
Risk factors

Content / Contact	**Distributor characteristics**	**User characteristics**	**User circumstances** • Past trauma or experiences of the user
User exposure User experiences: • Seeing lots of similar content User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 18: Josh (legal hazard, harmed)

Financial scams

Josh (26-30) started investing in cryptocurrency after being encouraged to invest by a former colleague, who had become a 'glamorous' lifestyle influencer by gaining their success from crypto currency and 'crypto-gurus' promoting investment. This friend encouraged Josh to invest in a new crypto currency. Josh did not seek or encounter counter-narratives dissuading or disproving the success of the currency. He invested in a new coin and experienced financial loss as a result, leading to depression and weight gain.



Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix' • Perceived as aspirational or holds social status	**Distributor characteristics** • Known to the user • Holds status or a reputation	**User characteristics**	**User circumstances**
User exposure User experiences: • Not seeing counter narratives User actions: • Active user engagement with the content/contact	**Societal context** • Acceptability of posting content • Trending topics	**Platform characteristics**	**Other platform users** • Validation / approval of content

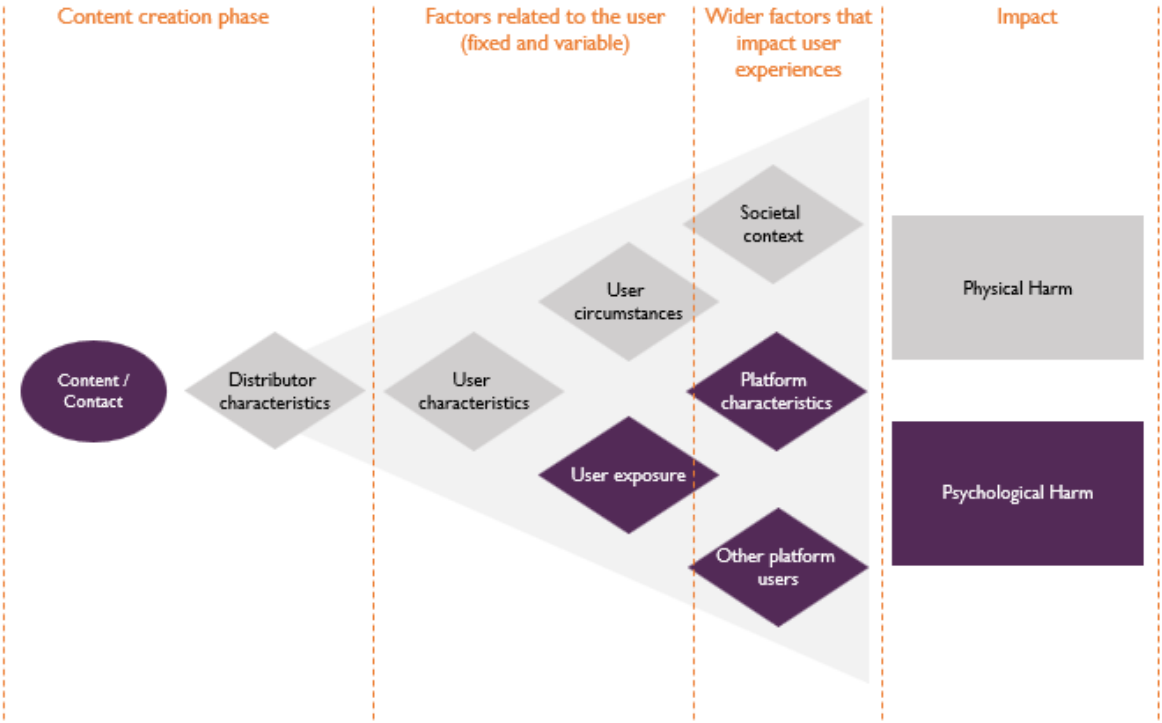
Case study 19: Kathleen (legal hazard, harmed)

Misogynistic / sexist content

Kathleen (21-25) receives derogatory comments about her gender on football-based discussion groups. This didn't bother her until one time when several other users also commented on her post with similar comments, giving the original perpetrator validation.

Once, she replied to a comment asking why the person had a certain view, which was followed by the same person messaging her directly the following day with more comments. After replying, "who are you?" to the messages, the messages were pushed into her mainstream "active conversations" inbox which meant that the person could then see when she was online. This made her feel vulnerable.

Overall, this has led to her self-excluding from these types of groups and feeling like she can't express herself the way she would like to.

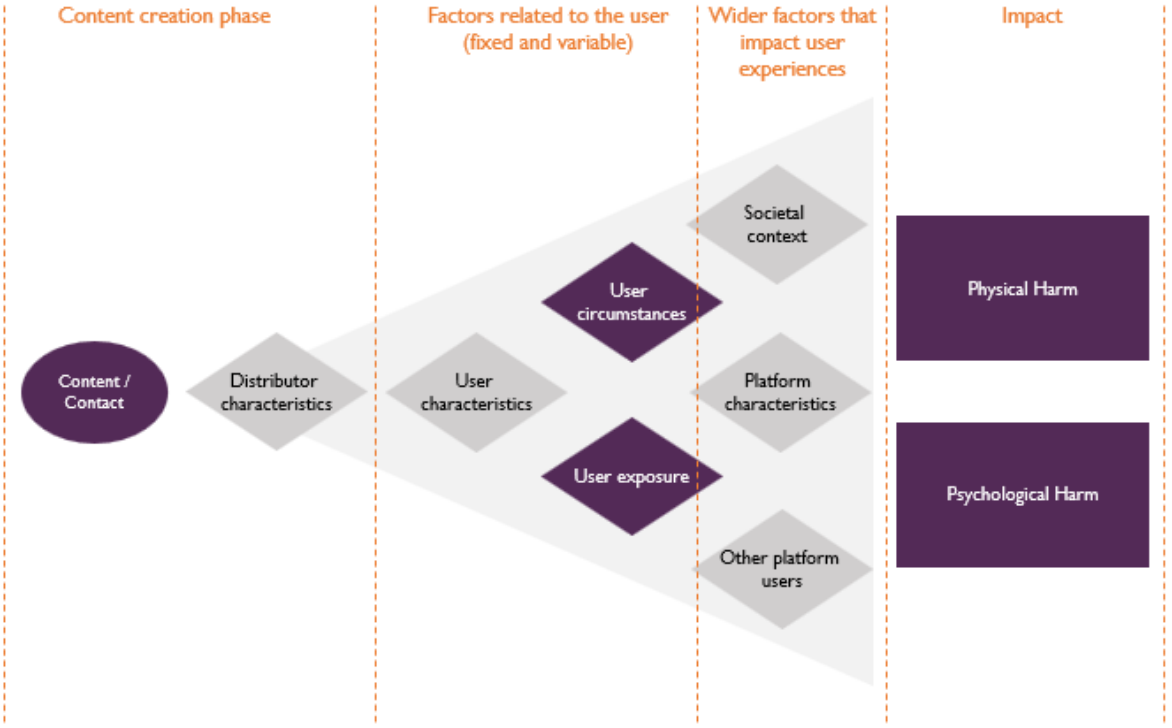


Risk factors

Content / Contact • Relatable or personal • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances**
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics** • Perceived permissibility of platform	**Other platform users** • Validation / approval of content

Case study 20: Lana (legal hazard, harmed)

Lana (18-20) has health anxiety. She engages with 'missed diagnosis' and 'sickness recovery' videos and searches out similar content. Her increased engagement with this content encourages the video-sharing platform's algorithms to feed her more similar videos, increasing her likelihood of ***exposure*** *to similar content. This exposure also leads to harm—she experiences panic-attacks, anxiety, and sleep deprivation.*



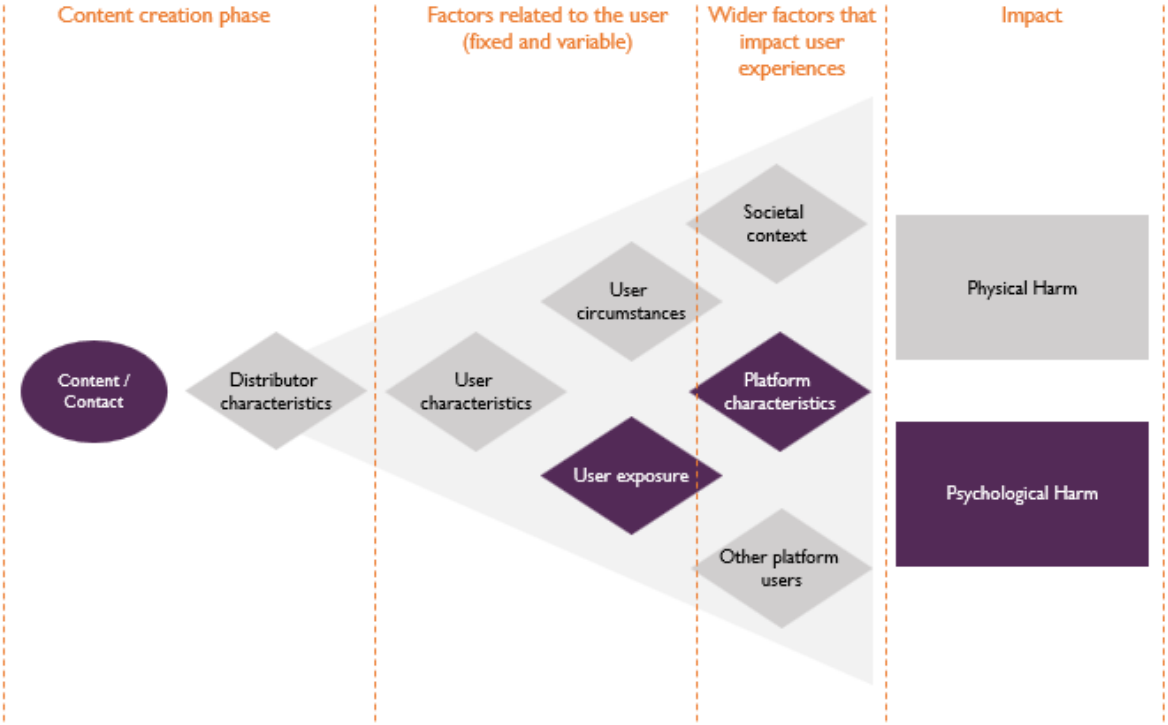
Risk factors

Content / Contact • Appeals to a desire or feels like a 'quick fix'	**Distributor characteristics**	**User characteristics**	**User circumstances** • The mental health of the user
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics**	**Other platform users**

Case study 21: Katie (legal hazard, harmed)

Cyber-flashing / unsolicited sharing of nudes

Katie (31-35) often receives unsolicited nudes from men via direct messages on an image-sharing platform. She has sometimes received multiple messages from the same person, but once she blocks them, she stops receiving messages from them. Despite this, she still receives new ones and questions if she is to blame for "encouraging" them.



Risk factors

Content / Contact • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances**
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics** • How easily users can be contacted on the platform • How easily accounts can be disposed of and remade on the platform	**Other platform users**

Case study 22: Jemma (legal and possibly illegal hazards, unharmed)

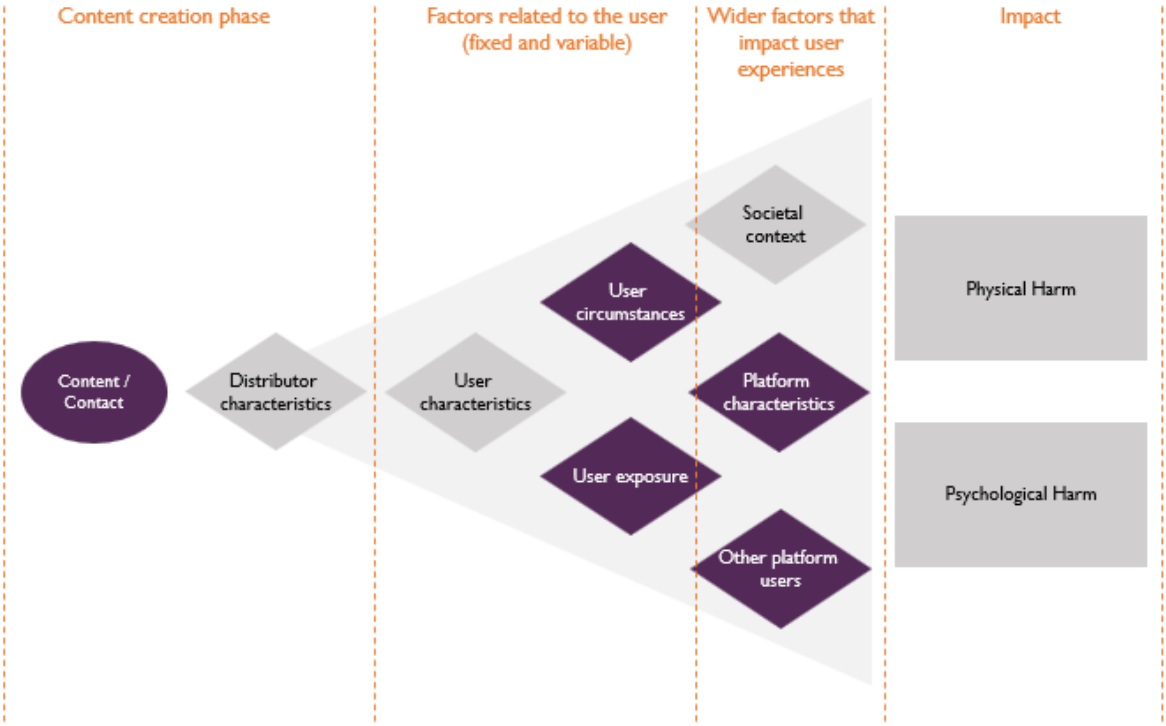
Abusive and threatening comments and posts, homophobic and transphobic content

Jemma (46-50) is a trans women who experiences harassment on a social media platform directed both at her and at the trans community in general, sometimes from accounts that are not 'live' for very long. Jemma feels like other users feel it is permissible to distribute content without repercussions on the platform.

A lot of her social life is online, so she doesn't want to stop using platforms, and she sometimes enjoys engaging in debates with other users.

Jemma sometimes reports derogatory comments when they are posted on other people's posts, but they are rarely taken down. She has taken to re-posting/ 'quoting' the posts with a note to say they should be blocked, so that multiple people report it. She thinks this makes the platform take more notice.

She says she isn't harmed by seeing this content – she likes seeking out debates, and if she really doesn't like something she reports the content and blocks the user.



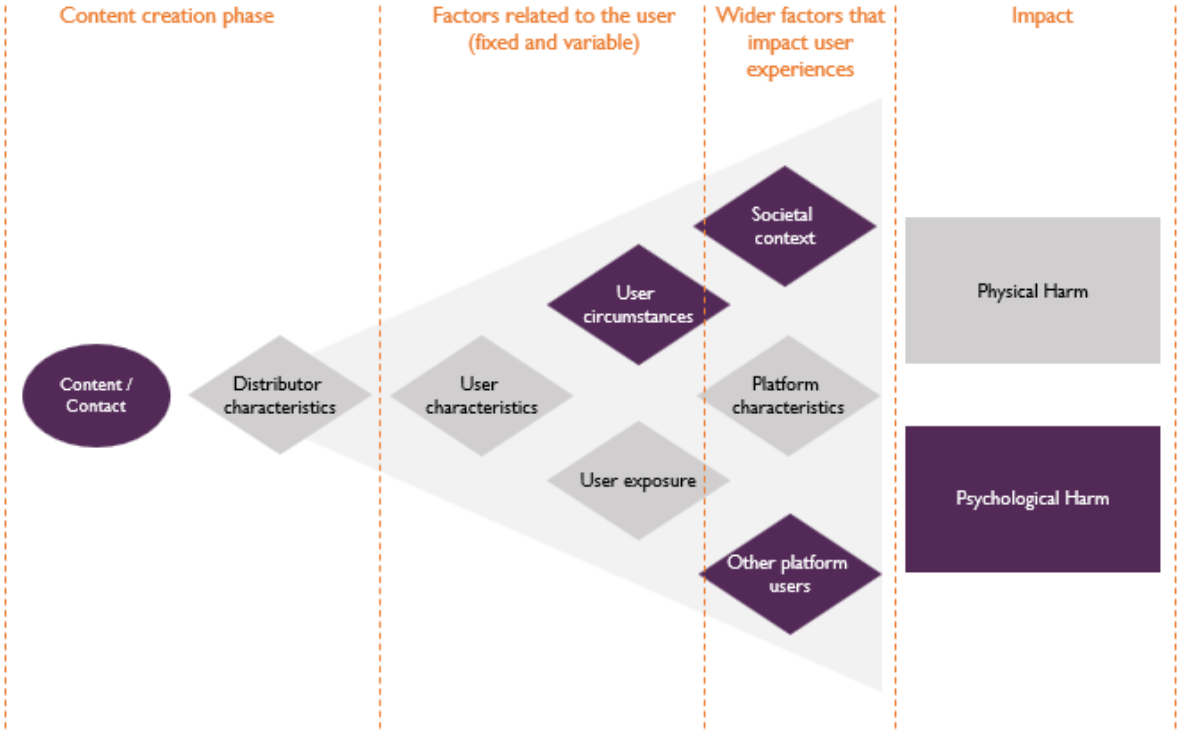
Risk factors

Content / Contact • Relatable or personal • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space
User exposure User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics** • How easily accounts can be disposed of and remade on the platform • Perceived permissibility of platform	**Other platform users** • Validation / approval of content

Case study 23: Adam (illegal hazard, harmed)

Abusive and threatening comments and posts

Adam (26-30) is a journalist who uses a social media platform to promote his writing. He experienced abuse and a pile-on online for posting an affirming comment on an innocuous, but potentially slightly right-leaning post which stirred controversy amongst BLM political activists. He was sent death threats via direct message.



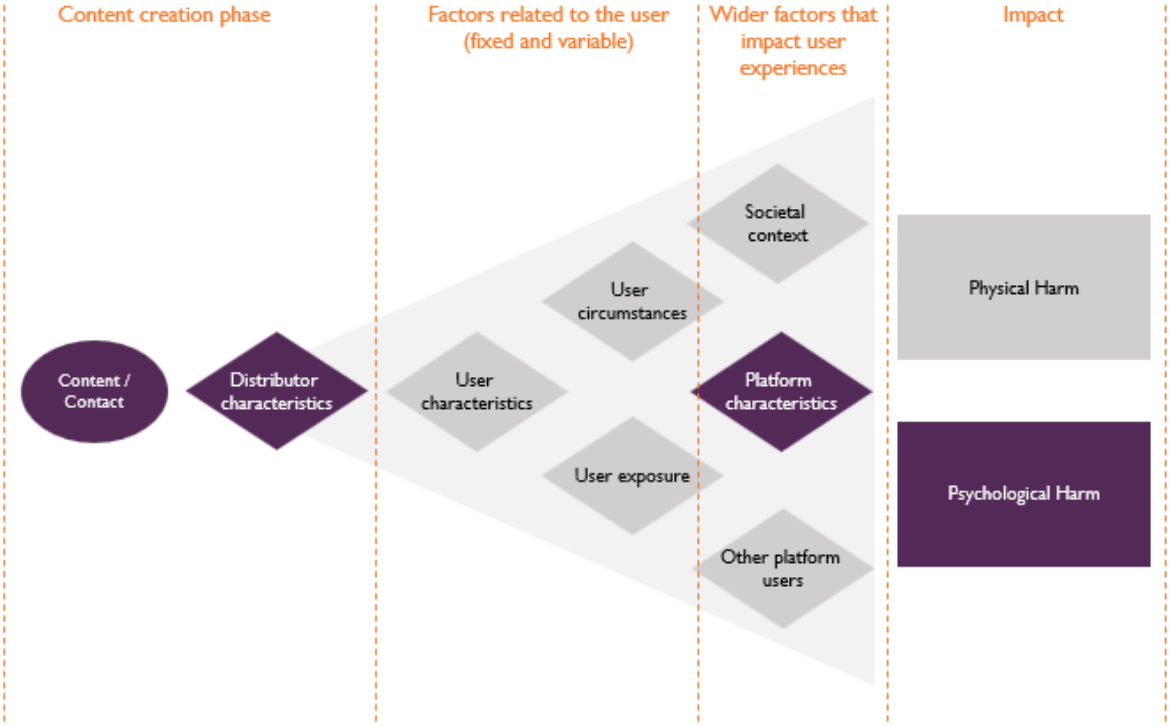
Risk factors

Content / Contact • Relatable or personal • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances** • The cost to the user of withdrawing from the online space
User exposure	**Societal context** • Acceptability of posting content	**Platform characteristics**	**Other platform users** • Validation / approval of content

Case study 24: Connie (legal hazard, harmed)

Abusive and threatening comments and posts

After an argument with Connie (26-30), Connie's friend created a false social media profile under the name 'Joe King'. The account had no profile picture or other personal information attached. The account was used by the friend to spread malicious and false rumours about Connie in the local area amongst friends and her places of work. Connie believes that the harassment would not have been carried out if her friend had not been able to make a fake account so quickly and easily, with no fear of consequences.



Risk factors

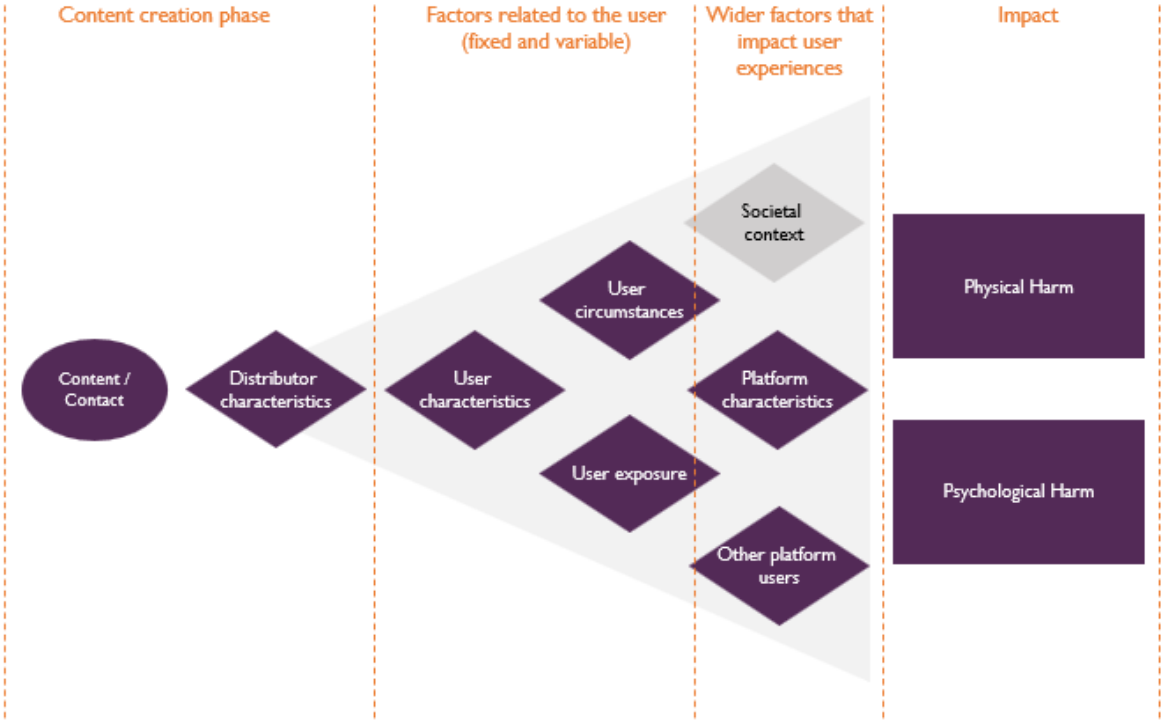
Content / Contact • Geographically local • Relatable or personal • Unexpected	**Distributor characteristics** • Known to the user • Geographically close	**User characteristics**	**User circumstances**
User exposure	**Societal context**	**Platform characteristics** • How easily accounts can be disposed of and remade on the platform	**Other platform users**

Case study 25: Aria (legal hazard, harmed)

Eating disorders, suicide and self-harm content

Aria (18-20) had previous experience of in-patient psychiatric care. Recently, she has noticed several video-sharing platform trends of other young people broadcasting their experiences of in-patient psychiatric care, which she finds triggering. She believes that other video-sharing platform users who don't have a history of in-patient care find the content intriguing and shocking, which increases engagement on the posts and content. This incentivises content creators to share similar content in the hope of also driving up their engagement and status on the platform, thus increasing the likelihood that Aria will be exposed to similar content.

At the age of 14, Aria was approached on a social media platform by an unknown user who encouraged her to join a group on a messaging service described as 'supporting girls with anorexia recovery.' The group encouraged the members to share their weights every week and self-harm if they had gained weight. Aria felt like there was a lot of pressure to conform. Aria was in the group for 1.5 years and became incredibly thin. She attempted to overdose and was admitted into full-time psychiatric care.



Risk factors

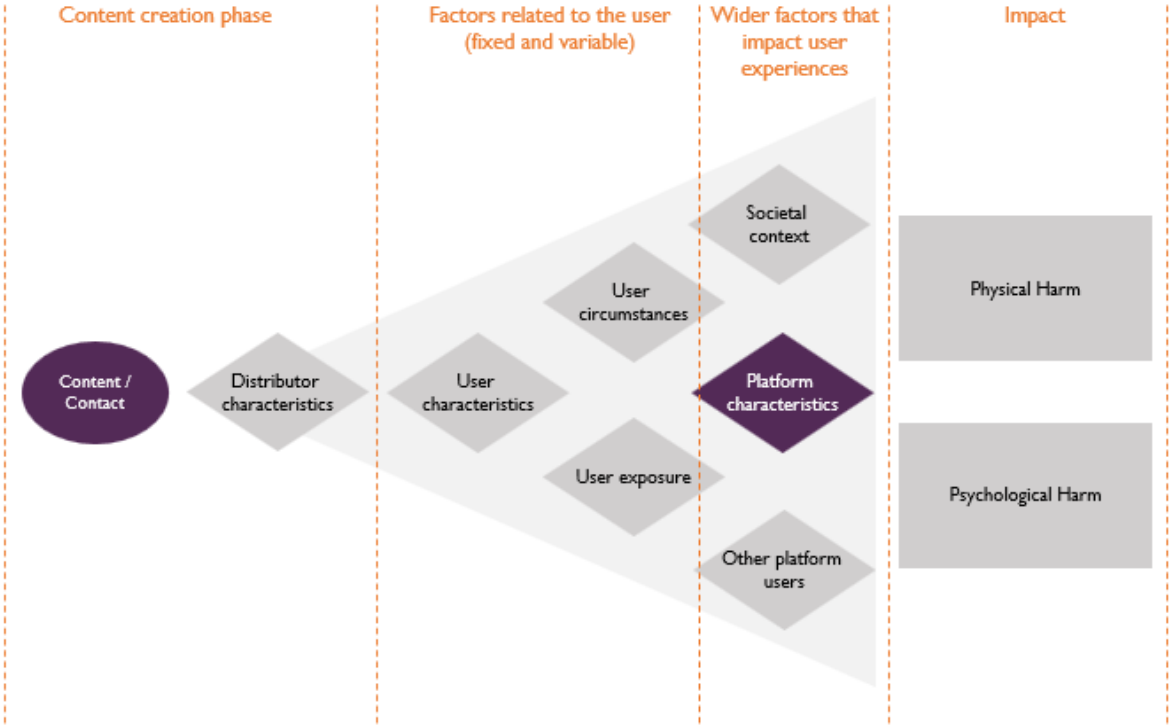
Content / Contact • Perceived as aspirational or holds social status • Relatable or personal	**Distributor characteristics** • Holds status or a reputation	**User characteristics** • The age of the user	**User circumstances** • The cost to the user of withdrawing from the online space • The mental health of the user • Past trauma or experiences of the user
User exposure User experiences: • Seeing lots of similar content User actions: • Active user engagement with the content/contact	**Societal context**	**Platform characteristics** • How the platform indicates popularity and status	**Other platform users** • Validation / approval of content • Peer pressure or recommendations

Case study 26: Hillary (legal hazard, unharmed)

Eating disorders

Hillary (31-35) works in HR and set out to research how to help a colleague struggling with an eating disorder. During her research, she clicked a link on an article which took her to a site encouraging sufferers to continue their disordered eating habits. The page had embedded links which took her to other sites, often with videos and diagrams relating to the encouragement of eating disorders.

She was shocked by the tone and genre of the content and concerned that she was able to access it easily, but was not harmed as she saw the content in a work context and does not suffer from an eating disorder herself.



Risk factors

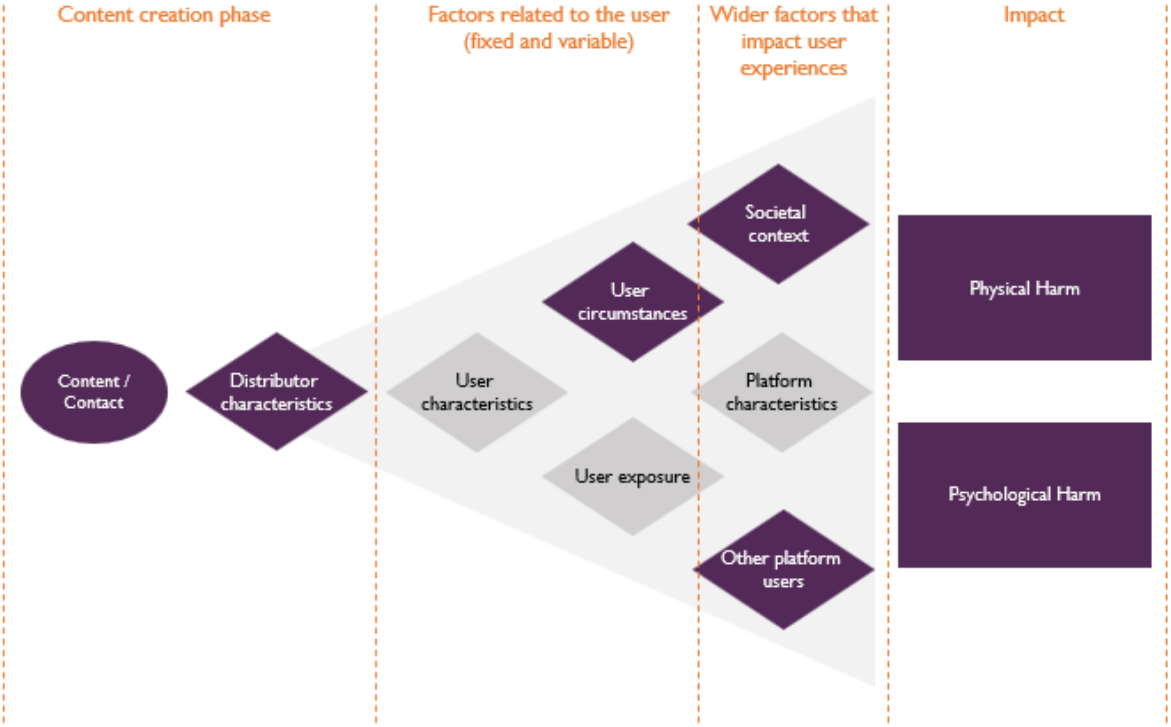
Content / Contact • Unexpected	**Distributor characteristics**	**User characteristics**	**User circumstances**
User exposure	**Societal context**	**Platform characteristics** • How easily users can be linked / signposted to other content or platforms	**Other platform users**

Case study 27: Damien (legal hazard, harmed)

Sexual violence, physical violence, gore content

Damien (51-55) was sexually and physically abused when he was a child. Instances of abuse bring up thoughts and emotions from his past. He has been exposed to ISIS beheadings and other gory human harm videos online from a link that a friend sent him.

A particular ISIS beheading has left a mark on him. He has had flashbacks of it and has instances when he has seen people who look similar in the street and had panic attacks.



Risk factors

Content / Contact • Perceived as real / genuine • Unexpected	**Distributor characteristics** • Known to the user	**User characteristics**	**User circumstances** • The mental health of the user • Past trauma or experiences of the user
User exposure	**Societal context** • Trending topics	**Platform characteristics**	**Other platform users** • Peer pressure or recommendations

Annex: Participant overviews

Audience 1 - Control Group			
1	Sasha	**26-30**	Sasha is a teaching assistant who received unsolicited nudes and scam messages on an image-sharing platform. She immediately blocks this kind of content and says it doesn't affect her.
2	Natalie	36-40	Natalie is a single mother who moved over from South Africa several years ago. She has modelling experience and frequently receives scam messages on an image-sharing platform asking her to model for different brands. She has fallen for an offline modelling scam in the past, so she is not harmed by these online interactions. She also frequently receives unsolicited explicit messages, which she is also unharmed by.
3	Jemma	46-50	Jemma is a trans woman who has experienced harassment via a social media platform. She transitioned a year ago and engages in (what she reports) are polite conversations with people on a social media platform who express anti-trans views to try and educate them. During one of these exchanges, she received abusive replies to her posts from people she did not know. She deleted her post and didn't engage in the conversation anymore. It made her feel a bit anxious and led to her not posting as frequently on a social media platform, but she didn't report that it had affected her greatly.
4	Jake	**26-30**	Jake enjoys gore and subscribes to several online discussion boards and image-sharing platform pages which pathologize injury and death. He also has paid-for subscriptions to more graphic content.
5	Maya	18-21	Maya sees people selling drugs on an image-sharing platform. The posts are often shared by people who she knows, or are friends of friends, from her hometown. She has never been tempted to buy from them and sees it as an inevitable part of being a young person. She is career-driven, works hard at school, has a supportive middle-class family, and says she would not want to jeopardise her prospects.
6	Hillary	**31-35**	Hillary works in HR and set out to research how to help a colleague struggling with an eating disorder. During her research, she accidentally came across an eating disorder forum which encouraged sufferers to continue their disordered eating habits. She was shocked by the tone and genre of the content but was not harmed as she does not suffer from an eating disorder herself.
7	Linda	46-50	Linda sees a lot of videos of bullying and knife/gun crime which are shared by community outreach workers to 'raise awareness'. She isn't personally affected by it, but it does take place in her local area, often in places she recognises around London. She also has 2 children, so she thinks about their proximity to the violence.
8	Katya	36-40	Katya was a former victim of domestic abuse via social media (intimate image abuse, coercive control via messages etc.). She is now an educator and raises awareness of digital domestic abuse like hers. For this reason, she engages with high volumes of potentially harmful content. She does not experience harm and instead says it makes her more confident about the way she is spending her time.
9	Jill	**22-25**	Jill got bullied a lot in school for her weight and appearance. Now Jill sees lots of bullying content online and it reminds her of what happened to her when she was younger. Whilst she recognises that this could be 'triggering' for her, she feels as if she has overcome it now and instead wants to spread messages of the importance of support.

Audience 2 – Illegal/likely to be illegal hazard[10], and harmed			
1	Brian	26-30	Brian received racially motivated abuse via direct and group messages from people he played Sunday league football against. He experienced anxiety and depression as well as receiving counselling for the event.
2	Dan	41-45	Dan was threatened online by an acquaintance from his local area, who sent messages to his friends and posted publicly on social media (which was later removed). Dan worried about his livelihood as a university lecturer and for the safety of his daughter, who lives with him. He was also physically assaulted as part of the incident by the perpetrator.
3	Adil	22-25	Adil received homophobic messages, threats, and abuse after coming out online as bisexual. A lot of the comments were from people he knew through the South Asian society at university. He felt ashamed and worried for his daughter, who was often featured in the abusive messages. He suffered from depression and anxiety following the incident and sought counselling.
4	Hayley	26-30	Hayley struggles financially and has been supporting herself and her partner on an unreliable income through the pandemic. She almost fell victim to an online scam impersonating HMRC, giving out her address, name, and date of birth. Her boyfriend stopped her before she gave across bank details. She finds the glamourous lifestyles presented online add to the anxiety and feelings of failure she has around her life, and fuel her need to find money-making schemes online.
5	Caleb	22-25	Caleb is a gay man who reports receiving kink death and rape threats over a dating app. He finds that the location setting of the dating app (which indicate how far you are from another person) make him feel particularly vulnerable. He also receives abuse over social media from people he doesn't know after he featured on a dating programme.
6	Shruti	26-30	Shruti is a victim of cyber-stalking form her ex-boyfriend. He sent aggressive messages as himself, and later used an anonymous alias online after she blocked him. It affected her work to the point she was recommended to seek professional support.
7	Neil	61-65	Neil fell for a romance scam online after being contacted by a French woman. They built up a relationship over the course of 2 months before she began asking him for money. He had spoken to her on video call and trusted her, so he continued sending her money when she requested it. He now believes it was all a scam, as her stories became more outlandish e.g. being arrested for drug trafficking and needing money for bail. He lost £3000 through the scam.
8	Reina	22-25	Reina was a victim of intimate image abuse. In the later stages of a breakup, Reina's ex would make multiple accounts and message Reina abuse and post abuse publicly. He then made an account pretending to be Reina and posted nude images of Reina after following her friends and family. This led to increased bouts of anxiety, sleep deprivation, Reina having to take time off work and negative perceptions of relationships and men.
9	Kevin	26-30	Kevin is autistic and found the second lockdown a particularly hard period in his life due to lack of space in his family home. During this time, Kevin was in a particularly dark place over a period of weeks. It resulted in him searching for

[10] Note that it was often not possible to gather sufficient information about a scenario to ascertain that the content was definitely illegal. Therefore, the research team have made an assumption, based on what the participants were able to tell us, about which content was likely to have been illegal.

			ways in which to kill himself. Kevin by-passed the initial pages that were support lines until he found forums that described methods to commit suicide and went on to consider the different methods. Amongst this advice there were comments encouraging suicide. In the moment, this content pushed Kevin toward following the advice, but in the longer term these pictures and methods also stayed in Kevin's mind and led to intrusive thoughts about suicide.
10	Adam	26-30	Adam is an aspiring journalist. He describes himself as right-leaning and has some socially conservative attitudes due to his faith. He spends a lot of time on a social media platform and often posts his journalism there. After commenting an innocuous comment supporting a post about the benefits of capitalism, he got targeted and piled upon by a mixture of anonymous and non-anonymous accounts. All these accounts were affiliated with the Black Lives Matter movement. Some of the abuse came through his direct messages as death threats, leaving him frightened, and reluctant to leave his house.

Audience 3 - Legal hazard and harmed

1	Laura	18-21	Lauren is a student and struggled with an eating disorder through lockdown, influenced by offline factors but fuelled by recovery image-sharing platform accounts and forums.
2	Aria	18-21	Aria was added to a messaging app grooming group chat for girls with eating disorders which kick-started her disordered eating and self-harm. Eventually she received in-patient psychiatric care. She now finds videos of self-harm and in-patient care triggering and makes her want to self-harm.
3	Tala	22-25	Tala sees a lot of glamourised lifestyle content which make her feel inadequate and depressed.
4	Connie	26-30	After an argument, Connie's friend impersonated her and sent messages to friends and acquaintances, disrupting Connie's livelihood, social life and long-term friendships.
5	Ahmed	31-35	Ahmed engaged with increasingly violent and hardcore porn which he believes has disrupted his relationship with women and sex. He also engaged with animal cruelty content which changed his worldview and made him turn vegetarian.
6	Scarlett	26-30	Scarlett has an eating disorder and finds social media prolongs her bad periods where she engages with eating disorder content which she finds using eating-disorder-related search terms.
7	Leyla	41-45	Leyla has been exposed to various pieces of harmful content online, including gore content and content relating to war and animal cruelty. She finds the content difficult to talk about and believes it has changed her worldview.
8	Eden	22-25	Eden is a non-binary streamer who experienced 'hate raids' on their birthday stream. Raiders filled their stream with distressing images, and redirected viewers to their messaging app channel, where more content was being shared. They were shocked and felt ashamed, targeted. Their moderators were also affected, having to clear the content on the messaging app server.
9	Lana	18-21	Lana has health anxiety and frequently sees health-related content on a video-sharing platform. She finds missed diagnoses and people's recovery journeys especially triggering. They give her panic-attack symptoms, trigger her anxiety, and cause sleep deprivation.
10	Fliss	18-21	Fliss receives unsolicited nudes via an image-sharing platform from people she has met on social media. She has been sexually assaulted several times throughout her life and finds that the unsolicited nature of some nudes are triggering, make her feel degraded and give her PTSD-like symptoms.
11	Kylie	26-30	Kylie has a history of being groomed online and emotional abuse. At the age of 16, she was in an online relationship/friendship with someone she met online. They had an argument and this person made her believe they committed suicide because of her. As a result, she is triggered by suicide content. There was a brief trend on a video-sharing platform where people where baiting users into watching a clip of someone shooting themselves in the head by showing it suddenly in a video of normal, mundane content. It caused her to have a panic attacks and feelings of anxiety, and found she struggled to function for a week afterwards.
12	Simon	26-30	Sam used to browse niche online discussion boards to explore his passions around Game of Thrones and certain anime shows. Three times within the past year he's come

			across extreme content disguised as seemingly safe posts. It's made him unfollow all online discussion boards and has ruined his enjoyment of those shows as he can no longer disassociate them from the extreme content that he saw.
13	Ashvi	18-21	Ashvi is a student. She is currently living at home, having recently returned from her placement year in France. She has found her time there hard and lonely, and often used a video-sharing platform as a crutch. She was spending up to 4 hours a day on the app. She engages with social activists and likes to stay informed via social media but is often confronted with content about violence towards women as a result. She finds this very triggering, reminding her of an assault she experienced at age 14. She feels scared, vulnerable and has a heightened sense of anxiety as a result. They also fuel night terrors and sleep deprivation, which she has experienced since the assault.
14	Damien	51-55	Damien was brought up in care and was sexually and physically abused. Instances of abuse bring up thoughts and emotions from his past. He has been exposed to ISIS beheadings and other gory human harm videos online from a link that a friend sent him. A particular ISIS beheading has left a mark on him. He has had flashbacks of it and has instances when he has seen people who look similar in the street and has had panic attacks. He has also been exposed to animal abuse.
15	Humza	18-21	Humza has been diagnosed with schizophrenia and is currently on antipsychotic drugs. Humza was going to the gym consistently for about a year and started researching methods to maximise results. This started off inconspicuously with diet and exercise but soon evolved into more extreme measures. Initially he was adamant he would never use steroids himself but was curious about the science behind it. Curiosity led to further research. He found video-sharing platform channels that discussed steroids and UK forums that were especially for discussion, information and advice about steroids. This eventually led to him ordering and using steroids, causing depleted testosterone levels, erectile dysfunction and gynecomastia. He is currently in the process of receiving testosterone replacement therapy with the NHS. Despite this, he still engages with steroid/performance-enhancing drugs content.
16	Becky	22-25	Over lockdown, Becky has been quite isolated and put on weight. She met a guy on a dating app in late 2020. They were chatting on an image-sharing platform, a messaging app and a dating app. On an image-sharing platform, she saw that he was liking other girl's photos who were all glamourous and skinny. She got into a spiral of fasting to try and lose weight, taking photos that made her look skinny and posting them for his approval. She obsessively checked if he had liked her photos, even when she was at work. She also had her profile picture and name stolen by someone and associated with a porn account. She found this funny, and it didn't really affect her – she just told all her friends to ignore it.
17	Yash	26-30	Yash saw an advert offering a free bet last year, shortly after moving to the UK for the first time. He became addicted to gambling and at one point had accounts on 20 casino websites. He has since 'self-excluded' meaning he can no longer make accounts with casinos but struggles with his addiction to in-person gambling.
18	Ellie	26-30	Ellie is an illustrator who uses an image-sharing platform to share her art. She receives racial and sexist abuse over her platform via direct messaging and live-streaming. More widely, she also has a complex relationship with an image-sharing platform, feeling like she has an addiction. The number of likes her art receives has a direct impact on her mood for the day. She is involved in several activist circles on the site and feels like the sharing of content and education about race is an obligation, and sometimes a large burden which has led her to burn out in the past.
19	Lucy	26-30	Lucy is a professional streamer who focuses on IRL (in real life) content with occasional gaming streams. She has been streaming for approximately 5 years and made it her full-time occupation two and a half years ago. She regularly has 200-300 people watching her content, many of whom she recognises in her chat. She had a wide social network through the streaming community but was involved in a scandal involving two other streamers which became popular gossip in the streaming community. She suffered 'hate raids', which she found targeting, making her feel vulnerable, isolated, and confused about her friendships. It has made her question her future in streaming and she has withdrawn from her streaming social network.

20	Mahalah	26-30	Mahalah was born in Israel and is Jewish. The content she saw was antisemitic content and individuals, and companies taking anti-Israel stances on social media. Mahalah also talked about news platforms presenting news in this manner. This content resulted in Mahalah feeling that her heritage made her part of a hated community. This isolated her and has led to her withholding where she is from and her religion.
21	Kelly	51-55	Kelly's daughter has a history with anorexia that led to her spending time in the hospital. Kelly was repeatedly recommended a documentary on her social media platform feed for two weeks. The documentary contained themes of anorexia and would bring up thoughts and emotions related to her daughter's past. Kelly also talked about negative news articles, specifically about death, that frequently feature on her feed from news platforms that she does not follow. Such news being intermixed with content from her friends and family has an added impact on Kelly. She talked about this content leading to a negative world view and excessive worry/fear about her own death.
22	Marta	31-35	Marta met a man online who she dated for a year and a half. He later turned out to be married. She received abuse online and in-person from her boyfriend's wife and friends.
23	Jackie	51-55	Jackie admits she is quite lonely. She started using a social media platform, which her sister recommended to her. The platform is designed to connect people who don't know each other for romantic or platonic relationships. She used it every evening for a period of 6 months, chatting with people. The platform encourages users to exchange messages, emoticons, and images sent directly to people which pop up without notification for the recipient, in a friendly way. She was sent a photo of a beheaded cat which she found horrific. She felt physically sick and was upset. She no longer uses the site and feels embarrassed and ashamed that she used a site which made her so vulnerable to risk.
24	Michelle	22-25	Michelle is a female gamer who lives with her husband and two children. Gaming is a major part of the family's life - Michelle said that gaming is their equivalent to watching TV. When gaming online Michelle receives gendered abuse ranging from 'get back into the kitchen' to sexual threats. This most commonly occurs on a player versus player game. Michelle says that this abuse often comes from players on her own team when things are not going well. This abuse will be via voice chat which often means it is hard to document and subsequently report. She is part of a female gaming community that supports and sometimes mass reports abusers after such instances. This abuse has led to Michelle gaming less although it is one of her passions and fears for her daughter if she games when she is older. Michelle also implied that she was a victim of sexual abuse growing up and that comments can bring up old emotions related to this.
25	Josh	26-30	Josh invested in a new cryptocurrency after engaging with the content of a 'crypto currency expert'. This content was a mixture of glamorous 'influencer' lifestyle pictures and pictures discussing and recommending crypto. Josh engaged with this content for three months before deciding to invest in a new launching coin. This led to an extreme monetary loss, depression, and weight gain.
26	Nassir	26-30	Nassir has received sustained online abuse from someone that he went to secondary school with. This abuse ranges from direct messages across multiple platforms to the abuser posting in local community social media platform pages. The nature of this abuse/posting wrongly accused Nassir of being adulterous, a domestic abuser and a rapist. In one instance the abuser messaged someone from Nassir's work on a social media platform. This led to a meeting and questioning at work. The police have been involved in the situation and the abuser was sectioned for a period. The result of this abuse was a breakdown in the relationship between Nassir and his partner, his work life being compromised, isolation and depression.
27	Jayesh	26-30	Jayesh is a British Muslim. His passion is to travel and keep a video blog about it on his video-sharing platform channel. On one video he posted of himself in Iraq, someone left an Islamophobic comment that left Jayesh feeling distressed and angry. He had received light criticism and people questioning him before, but this was very out of the blue and left him angry, disheartened and confused for days.
28	Vera	26-30	Vera uses her social media to share images of her work, which is important for her networking. She frequently receives misogynistic abuse in response to the images. She also receives abuse when commenting or replying to news stories. She has recently been a victim of sexual harassment, which alongside online content, has led her to feeling distrustful of men, and vulnerable when out.

29	Lisa	41-45	Lisa has a history of sexual assault and a learning disability, for which she receives 24-hour carer support. She was contacted by an anonymous account who sent her physical threats and abuse about her appearance and learning difficulties. She contacted the police with the help of her carers, but was so scared to leave the house, she began going out with a knife and reducing her time outdoors.
30	Nora	22-25	Nora has been diagnosed with anxiety, depression, borderline personality disorder and avoidant personality disorder. Nora regularly reads true crime content on an online discussion board, sometimes for 2-3 hours a day, and usually late into the evening. Recently she opened a link on a comment in an online discussion board thread and came across a gore website. The main photo was one of dead bodies, and Nora was disgusted. She felt physically sick. She was already "hyper-vigilant", but this incident has exacerbated her perceived need to be careful wherever she goes as she's aware there are people out there who would happily post and engage with this kind of content, and people who may do her harm.
31	Lydia	66-70	Lydia lives alone. Lydia suffers from MS and spends the majority of her time at home on her own. Her screen time is now mostly made of up of a social media platform and a news website. Previously, Lydia used another social media platform, but during the fallout of the Brexit vote and Covid Lydia stopped using this platform as elderly people were frequently targeted and blamed for the political climate and labelled 'senile', 'bedwetters' and 'bedblockers'. Lydia now uses a different social media platform, and frequently sees ageist news content and comments on this platform. She gave the example of the recent news story about the pensioner who had to ride the bus all day to stay warm. In the comments beneath this story, elderly people were derided and labelled as 'burdens' and 'senile'. Similarly, on a recent news article, comments asked 'who is this menopausal woman'. Lydia believes this content is a contributing factor in her depression, leads to isolation and undermines her self-worth and her perception of what people think of her.
32	Hattie	18-21	Hattie sees homophobic and transphobic comments on videos. She also occasionally gets served videos on a video-sharing platform of people perpetuating negative stereotypes of the LGBT community because she engages with LGBT content on the platform. These play on her mind and make her feels concerned that if you engage in any way, you'll be targeted yourself.
33	Iboh	31-35	Iboh has seen 'jungle justice' videos on a social media page from her home country of Nigeria. One she recalls well featured a man who stole from a shop being burned alive by local people. She was and continues to be distressed by the video. Videos like this change her perception of her home country, making her unwilling to return and distancing her from her family there.
34	Luke	41-45	Luke saw an execution video on video-sharing platform a few months ago after it appeared on his home page as a suggested video. He was shocked, upset, and has not used this video-sharing platform since. His perception of the platform has changed as he thought it would be regulated to cut out content like that. There were some online 'goods' from the incident, including informing him of how vulnerable he is online and motivating him to make his accounts and information more secure.
35	Katie	31-35	Katie has started receiving videos of men masturbating via direct message on an image-sharing platform. They seem to be from random strangers – never the same person twice. Overall, she uses logical thinking and humour to try not to let it affect her, such as just thinking "it's a sad pervy person sitting by themselves in a room". But she now keeps wondering why it's happening, thinking she is to blame, and looking back through her posts to see if there is anything that might encourage it. She has changed her behaviour and now won't post any photos of herself on the beach or anything that would show any skin – just in case. She gets a bit of a dread feeling whenever she opens her image-sharing platform inbox.
36	Kathleen	22-25	Kathleen would join common interest social media groups around things like the football World Cup. Sometimes when she would post things, she would get derogatory comments such as 'women don't know anything about football', 'get back in the kitchen' etc. One time someone also then messaged her directly on a social media platform with similar comments and tried to get her to engage. This all means that she no longer feels comfortable posting on these types of groups - she would mainly 'like' posts instead of

			writing things as then she is more anonymous. Kathleen is annoyed that she feels she can't do what she wants because these people will harass her.

Participant from Ofcom's forthcoming report on *Risk factors that may lead children to harm online*			
1	Noah	12-14	Noah lives with his mum and two older brothers. Recently, Noah reported experiencing bullying by some other boys at school whom he had previously been friends with.
2	Lucy	15-17	Lucy lives with her family. When she was 13, she received a few unsolicited nude pictures from someone she knew from school. She believes that the experience has negatively shaped her perceptions of men, and has changed her behaviour on social media in that it is shaping what she posts.

Annex: Methodology

Data collection

Data for the research was collected through in-depth interviews with 55 participants. Due to Covid-19 restrictions at the start of the research process, all interviews were conducted remotely by video-call to protect participants and researchers from any risk of infection.

Each interview lasted between 1-2.5 hours. A semi-structured approach was followed, with the researcher using a topic guide as a prompt throughout the call. Key themes in the topic guide included:

- Participant background
- Online habits and device use
- Overview of the hazard/s experienced
- Mapping the experience through each component of the model

Ofcom's research on children's online harms

On behalf of Ofcom, Revealing Reality are simultaneously carrying out research into the online harms that children experience. This research involves interviews with 40+ children aged 8-17 and explores similar topics. The model was designed to reflect how harm happens to adults. However, the research with children enabled us to check whether the same routes are true for younger people. Revealing Reality found that the model was equally applicable to the experiences of children and therefore, to make findings more robust, the data set from the project with children has also been drawn on during analysis and included in findings where relevant.

Recruitment

A sample of 55 participants was selected for the project to stress test Ofcom's model of how online harms manifest.

The majority were recruited because they believed they had recently experienced severe/significant harm as a result of online content / contact within the last six months. The sample was therefore skewed towards people who had had bad experiences online, and not reflective of the general population. The requirement for the exposure to be recent was included to ensure that participants could still recall the experience in detail—especially as, to minimise the risk of further harm, participants were not asked to show us the content / contact during the interview so they had to rely on recall.

Of the 55 individuals:

- 36 people had experienced significant harm from legal online hazards
- 10 people had experienced significant harm from online hazards likely to be illegal[11]
- 9 people had experienced exposure to similar hazards but were unharmed – this group was recruited as a control group

Recruitment focused on ensuring that participants across the sample had experienced a broad range of hazards, including:

[11] During the research we found that concluding whether or not a particular experience involved legal or illegal content was difficult. Few of the participants had approached the police, and none of the cases had led to a prosecution. Therefore, participants were allocated to the 'likely to be illegal' category based on whether it is likely, given what we had been told, that the content was illegal. This was more straightforward in cases of selling illegal goods, but less for harassment, bullying, hate and abuse. The task was made harder as we intentionally did not ask participants to seek out the original content to minimise the risk of algorithms mistaking this for an interest in being served more similar content.

- Content / contact attacking protected characteristics
- Content / contact of a sexual nature
- Content / contact selling or promoting illegal behaviour/objects e.g. weapons, drugs
- Eating disorder content
- Glamourised lifestyle content
- Gore and violent content
- Misinformation
- Online abuse and harassment
- Online fraud and scams
- Suicide content[12]

Similarly, recruitment sought to include a range of harms experienced as a result of different hazards across the sample. This included both physical and psychological harms which were categorized as significant by participants.

Other factors considered during recruitment to ensure a varied sample included people's online activity (e.g. types of online activity, devices used, time spent online) and demographic criteria (e.g. household income/social grade, age, gender, ethnicity, accessibility needs, sexuality, religion, and region within the UK). These factors were of secondary importance to ensuring a range of hazards and harms were accounted for in the sample.

Note that because the recruitment criteria required people to be aware they had been harmed (except in the control group) we did not include people who might have been subject to mis/disinformation in the sample. A different study design will be required to understand how harms manifest for this audience, given they may not self-identify as being harmed.

The research team used two recruitment strategies. Firstly, working with a free-find recruitment partner, who used their connections and database to recruit 20 participants. Secondly, the research team themselves recruited 35 participants. The team tested several routes to find potential participants and iterated their approach as they went. Routes were strategically followed to ensure the study included a sample of people who had experienced a range of harms online, often tailoring the recruitment approach and materials to reach people who had experienced one specific hazard.

The team used social media support groups, community groups, grassroots organisations, and paid advertising. In total, the team contacted 114 organisations. Paid advertising on social media platforms and online noticeboards were the most successful recruitment routes, followed by student and community groups on social media. Despite the number of grassroots organisations and support groups for victims of online harms that were contacted, the research team only received responses from four organisations and no participants were recruited this way

Individuals who were interested in taking part in the research filled in a form expressing interest and with some basic details. The research team then screened these people through a 10-minute phone call, asking them more about their experiences and characteristics. Interviews were scheduled with those who were suitable. Care was taken to follow ethical and data protection procedures as agreed with Ofcom and as outlined in the project Data Protection Impact Assessment.

Findings from recruitment

Although not all potential participants were taken through for the research, the recruitment process revealed some interesting findings as well.

[12] Note: experiences of terror-related content, and child sexual abuse material (CSAM) was not represented in the sample. People in the sample who experienced exposure to illegal content / contact included death threats, content assisting suicide, illegal harassment, racist abuse, homophobic abuse, online scams, cyberstalking and intimate image abuse.

- The number of individuals coming forward who had experienced harm from engaging with eating disorder content / contact was disproportionately high. Not all of these people were interviewed to ensure a good mix in the final sample in order to fully test the model.
- There was a high volume of people reporting short-term and low-impact harm based on having seen content relating to news and current affairs. These people were not included as most of this was not user-generated content / contact and they hadn't experienced significant harm.
- Individuals were most frequently reporting that they had experienced harm via what the research team later identified as route 1 (isolated), compared to the other routes. The research team suspects that the other routes may have been harder for individuals to identify, or the full effects of an experience may not yet have been realised by the individual. Alternatively, the other routes may simply be less common.

Analysis

Researchers recorded interviews and took detailed fieldnotes. Data was transferred into an analysis grid. The data from each participant was mapped against the components of the model. Risk factors and other data that didn't fit into the pre-defined components of the model were highlighted and discussed by research team members. Patterns in risk factors were identified. Interviews were also discussed in depth between the research team members, with bigger themes being drawn out, and then sense-checked against the data.

Safeguarding and ethics

The nature of this project required researchers to discuss sensitive issues with participants. Ethics and safeguarding were of paramount importance. The research team went through a rigorous process of discussing and planning around ethical issues before any recruitment or fieldwork took place. This included producing a comprehensive DPIA, safeguarding protocol and protocol around disclosure of illegal activity. Some key components of the approach were:

- Ensuring participants were fully informed about what the research would involve and felt comfortable discussing their experience
- Frequently checking in with participants about how they were feeling and giving them opportunities to take breaks or withdraw
- Ensuring that participants understood the team's obligations around safeguarding and that confidentiality may be overridden in cases where there is risk of harm
- Avoiding the sharing of any illegal content (e.g. photos) from the participant to the researcher

The research team were also fully briefed on what the research would involve, and a policy was in place in the event of researchers feeling distressed when conducting the research.

No ethical issues were recorded during the project.

Limitations of the methodology

This was a large sample for a qualitative research project. However, there were some limitations of this methodology and conclusions that can be drawn from the data collected:

- The majority of the sample were recruited because they believed they had recently experienced severe/significant harm as a result of online content / contact within the last six months. The sample was therefore skewed towards people who had had bad experiences online, and not reflective of the general population.
- The data does not include examples of all types of hazards and harm, so there are likely to be additional factors or findings that may emerge from a larger/different data set
- The researchers were relying on self-reporting from individuals, and their ability to remember and analyse what they had engaged with. To mitigate this, we interviewed people who had had their experience in the past six months, but there may still be inaccuracies in participants' testimony

- Whilst the qualitative methodology demonstrates that certain things happen/exist, it cannot lead to an estimate of the number of cases, or the probability/likelihood of harm occurring.

Thank you