

<b>Title:</b> The Online Safety Bill <b>RPC Reference No:</b> RPC-DCMS-4347(2) <b>Lead department or agency:</b> Department for Digital, Culture, Media and Sport <b>Other departments or agencies:</b> Home Office	<b>Impact Assessment (IA)</b>
	<b>Date:</b> 26/04/2021
	<b>Stage:</b> Consultation
	<b>Source of intervention:</b> Domestic
	<b>Type of measure:</b> Primary Legislation
	<b>Contact for enquiries:</b> soh-analysis-team@dcms.gov.uk
<b>Summary: Intervention and Options</b>	<b>RPC Opinion: Informal review<sup>1</sup></b>

Cost of Preferred (or more likely) Option (in 2019 prices)			
Total Net Present Social Value	Business Net Present Value	Net cost to business per year	Business Impact Target Status Qualifying provision
N/Q	N/Q	N/Q	
<b>What is the problem under consideration? Why is government action or intervention necessary?</b> Internet use is growing and evidence indicates that Covid-19 has increased this even further. While the internet is a powerful force for good, illegal and harmful content and activity is widespread online. On the whole, there is a lack of transparency on the potential for experiencing harm online which makes it more difficult for consumers to make an informed choice. In addition, an inconsistent approach towards fighting harms has limited the effectiveness of voluntary efforts to disrupt criminals from using platforms. Therefore, without government intervention, limited progress will be made at reducing online harms.			

<b>What are the policy objectives of the action or intervention and the intended effects?</b> The policy aims to make online platforms a safer place for all. The policy objectives are as follows: <ul style="list-style-type: none"> <li>- <b>User safety:</b> improved safety of users online, through reduced risk and incidence of specific online harms, especially with respect to vulnerable groups.</li> <li>- <b>Preserving freedom of speech:</b> ensuring sufficient safeguards for freedom of expression.</li> <li>- <b>Law enforcement:</b> improving the efficacy of law enforcement and crime prevention with respect to illegal content and behaviour online.</li> <li>- <b>Efficiency:</b> increased coherence and clarity of government activity to tackle online harms and build the capability of users to stay safe online.</li> <li>- <b>Evidence:</b> a culture of transparency enhancing the amount and quality of information in relation to online harms that is available to government, industry, civil society, and wider society.</li> </ul>
--

<b>What policy options have been considered, including any alternatives to regulation? Please justify preferred option (further details in Evidence Base)</b> This impact assessment (IA) considers three policy options (in addition to a <i>do nothing</i> option); however, a range of options have been considered as part of the policy development process. <ul style="list-style-type: none"> <li>● <b>Option 0:</b> The option to “do nothing” is also considered within this IA, this would entail a continuation of platforms being liable for illegal content that they “host” only. So far, this approach has not provided sufficient incentive for platforms to reduce online harms.</li> <li>● <b>Option 1:</b> A duty of care for user generated content and activity addressing illegal harms and safeguarding children from both illegal and harmful content activity. Duties are set out in primary legislation and guidelines or codes of practice.</li> <li>● <b>Option 2 (preferred):</b> A duty of care for user generated content and activity addressing both illegal and legal but harmful content, and safeguarding children from illegal and legal but harmful content. Duties are set out in primary legislation (and subsequent secondary) and guidelines or codes of practice.</li> </ul>
--

<sup>1</sup> [The RPC - how we work with departments](#)

- **Option 3:** Detailed safety duties setting out organisations' responsibilities in addressing illegal harms and legal but harmful content, and the safeguarding of children from both illegal and legal but harmful content. These safety duties are detailed in primary legislation (and subsequent secondary) and are uniformly applied across all harms and organisations in scope.

**Option 2** is the preferred option as it is likely to achieve reductions in online harms while maintaining a proportionate and risk-based approach.

Does implementation go beyond minimum EU requirements?		N/A		
Is this measure likely to impact on international trade and investment?		Yes (minimal)		
Are any of these organisations in scope?	<b>Micro</b> Yes	<b>Small</b> Yes	<b>Medium</b> Yes	<b>Large</b> Yes
What is the CO <sub>2</sub> equivalent change in greenhouse gas emissions? (Million tonnes CO <sub>2</sub> equivalent)		<b>Traded:</b> N/Q		<b>Non-traded:</b> N/Q
<b>Will the policy be reviewed?</b> It will be reviewed. <b>If applicable, set review date:</b> within 5 years				

*I have read the Impact Assessment and I am satisfied that, given the available evidence, it represents a reasonable view of the likely costs, benefits and impact of the leading options.*

Signed by the responsible : Catherine Colebrook Date: 26/04/2021

## Summary: Analysis & Evidence

## Policy Option 1

**Description:** A duty of care for user generated content and activity addressing illegal harms and safeguarding children from both illegal and harmful content activity.

### FULL ECONOMIC ASSESSMENT

Price Base Year 2019	PV Base Year 2020	Time Period Years 10	Net Benefit (Present Value (PV)) (-£1,689m)		
			Low: -	High: -	Best Estimate: -£1,689m (illustrative only)

COSTS (£m)	Total Transition (Constant Price) Years		Average Annual (excl. Transition) (Constant Price)	Total Cost (Present Value)
Low	Optional		Optional	Optional
High	Optional		Optional	Optional
Best Estimate	£40.5m		£197.0m	£1,689.4m

#### Description and scale of key monetised costs by 'main affected groups'

Businesses are expected to incur the following transition costs (all in 10 year PV): reading and understanding the regulations (£9.2 million), ensuring they have a user reporting mechanism in place (£12.4 million), and updating terms of service (£14.7m)

Business are expected to incur the following ongoing compliance costs (all in 10 year PV): producing risk assessments (£31.0 million), additional content moderation (£1,271.5 million), transparency reporting (£3.6 million), industry fee (£346.7 million)

Government is expected to incur the following costs (all in 10 year PV): justice impacts (£0.4 million)

#### Other key non-monetised costs by 'main affected groups'

The following costs to business have not been monetised: fines for non-compliance, cost to internet service providers (ISPs) and payment service providers (PSPs) of business disruption measures, potential requirement for some businesses to adopt age verification systems, cost to industry and government stemming from the requirement to report online CSEA. Where possible, this IA provides an indication of the likely scale of these impacts.

There are a number of indirect costs and wider impacts on society which have not been monetised, these include innovation impacts, competition impacts, freedom of expression implications, privacy implications, and trade impacts - these have all been assessed qualitatively.

BENEFITS (£m)	Total Transition (Constant Price) Years		Average Annual (excl. Transition) (Constant Price)	Total Benefit (Present Value)
Low	Optional		Optional	Optional
High	Optional		Optional	Optional
Best Estimate	-		-	-

#### Description and scale of key monetised benefits by 'main affected groups'

Based on a subset of quantified online harms<sup>2</sup>, this IA estimates that this option would need to reduce online harms by 3.1% (average, annual) in order to break even, this equates to around £201 million average annual benefit over the appraisal period. Given the difficulties in monetising the impact of online harms, this represents a very conservative approach to benefit estimation and the break even point is likely much lower. These potential benefits are included only for the break-even analysis and have not been included in the illustrative Net Present Social Value.

<sup>2</sup> Cyberbullying, cyberstalking, intimidation of public figures, child sexual abuse and exploitation, modern slavery, hate crime, and the sale of illegal drugs online.

<b>Other key non-monetised benefits by 'main affected groups'</b>		
<b>Key assumptions/sensitivities/risks</b>	<b>Discount rate(%)</b>	3.5%
The key assumptions for this option are: the number of <b>businesses</b> in scope of the regulations and the incremental cost (in terms of percentage of turnover) of complying with the requirements - all key assumptions are tested in the risks and sensitivity section.		

**BUSINESS ASSESSMENT (Option 1)**

<b>Direct impact on business (Equivalent Annual) £m:</b> illustrative only at this stage			<b>Score for Business Impact Target (qualifying provisions only) £m:</b> to be scored at secondary
<b>Costs: 156.0</b>	<b>Benefits: -</b>	<b>Net: 156.0</b>	

## Summary: Analysis & Evidence

## Policy Option 2

**Description:** A duty of care for user generated content and activity addressing both illegal and legal but harmful content, and safeguarding children from illegal and legal but harmful content.

### FULL ECONOMIC ASSESSMENT

Price Base Year 2019	PV Base Year 2020	Time Period Years 10	Net Benefit (Present Value (PV)) (-£2,103m)		
			Low: -	High: -	Best Estimate:-£2,118m (illustrative only)

COSTS (£m)	Total Transition (Constant Price) Years	Average Annual (excl. Transition) (Constant Price)	Total Cost (Present Value)
Low	Optional	Optional	Optional
High	Optional	Optional	Optional
Best Estimate	£40.5m	£248.1m	£2,118.1m

#### Description and scale of key monetised costs by 'main affected groups'

Businesses are expected to incur the following transition costs (all in 10 year PV): reading and understanding the regulations (£9.2 million), ensuring they have a user reporting mechanism in place (£12.4 million), updating terms of service (£14.7 million).

Businesses are expected to incur the following ongoing compliance costs: producing risk assessments (£31.0 million), additional content moderation (£1,700.2 million), transparency reporting (£3.6 million), industry fee (£346.7 million)

Government is expected to incur the following costs (all in 10 year PV): justice impacts (£0.4 million)

#### Other key non-monetised costs by 'main affected groups'

The following costs to business have not been monetised: fines for non-compliance, cost to ISPs and PSPs of business disruption measures, potential requirement for some businesses to adopt age verification systems, cost to industry and government stemming from the requirement to report online CSEA. Where possible, this IA provides an indication of the likely scale of these impacts.

There are a number of indirect costs and wider impacts on society which have not been monetised, these include innovation impacts, competition impacts, freedom of expression implications, privacy implications, and trade impacts - these have all been assessed qualitatively.

BENEFITS (£m)	Total Transition (Constant Price) Years	Average Annual (excl. Transition) (Constant Price)	Total Benefit (Present Value)
Low	Optional	Optional	Optional
High	Optional	Optional	Optional
Best Estimate	-	-	-

#### Description and scale of key monetised benefits by 'main affected groups'

Based on a subset of quantified online harms<sup>3</sup>, this IA estimates that this option would need to reduce online harms by 3.9% (average, annual) in order to break even, this equates to around £253 million average annual benefit over the appraisal period. Given the difficulties in monetising the impact of online harms, this represents a very conservative approach to benefit estimation and the break even point is likely much lower. These potential benefits are included only for the break-even analysis and have not been included in the illustrative Net Present Social Value.

<sup>3</sup> Cyberbullying, cyberstalking, intimidation of public figures, child sexual abuse and exploitation, modern slavery, hate crime, and the sale of illegal drugs online.

<b>Other key non-monetised benefits by 'main affected groups'</b>		
<b>Key assumptions/sensitivities/risks</b>	<b>Discount rate(%)</b>	3.5%
The key assumptions for this option are: the number of businesses within scope of the regulations, and the incremental cost (in terms of percentage of turnover) of complying with the requirements - all key assumptions are tested in the risks and sensitivity section.		

**BUSINESS ASSESSMENT (Option 2)**

<b>Direct impact on business (Equivalent Annual) £m:</b> illustrative only at this stage			<b>Score for Business Impact Target (qualifying provisions only) £m:</b> to be scored at secondary
<b>Costs: 205.8</b>	<b>Benefits: -</b>	<b>Net: 205.8</b>	

## Summary: Analysis & Evidence

## Policy Option 3

**Description:** Detailed safety duties setting out organisations responsibilities in addressing illegal harms and legal but harmful content, and the safeguarding of children from both illegal and legal but harmful content.

### FULL ECONOMIC ASSESSMENT

Price Base Year 2019	PV Base Year 2020	Time Period Years 10	Net Benefit (Present Value (PV)) (-£7,355m)		
			Low: -	High: -	Best Estimate: -£7,355m (illustrative only)

COSTS (£m)	Total Transition (Constant Price) Years	Average Annual (excl. Transition) (Constant Price)	Total Cost (Present Value)
Low	Optional	Optional	Optional
High	Optional	Optional	Optional
Best Estimate	£64.6m	£867.9m	£7,355.5m

#### Description and scale of key monetised costs by 'main affected groups'

Businesses are expected to incur the following transition costs (all in 10 year PV): reading and understanding the regulations (£9.2 million), ensuring they have a user reporting mechanism in place (£33.2 million), updating terms of service (£14.7 million).

Businesses are expected to incur the following ongoing compliance costs: producing risk assessments (£31.0 million), additional content moderation (£5,999.3 million), transparency reporting (£921.0 million), industry fee (£346.7 million)

Government is expected to incur the following costs (all in 10 year PV): justice impacts (£0.4 million)

#### Other key non-monetised costs by 'main affected groups'

The following costs to business have not been monetised: fines for non-compliance, cost to ISPs and PSPs of business disruption measures, potential requirement for some businesses to adopt age verification systems, cost to industry and government stemming from the requirement to report online CSEA. Where possible, this IA provides an indication of the likely scale of these impacts.

There are a number of indirect costs and wider impacts which have not been monetised, these include innovation impacts, competition impacts, freedom of expression implications, privacy implications, and trade impacts - these have all been assessed qualitatively.

BENEFITS (£m)	Total Transition (Constant Price) Years	Average Annual (excl. Transition) (Constant Price)	Total Benefit (Present Value)
Low	Optional	Optional	Optional
High	Optional	Optional	Optional
Best Estimate	-	-	-

#### Description and scale of key monetised benefits by 'main affected groups'

The benefits presented in this IA are purely illustrative and have not been included in the main metrics. Based on a subset of quantified online harms<sup>4</sup>, this IA estimates that this option would need to reduce online harms by 13.5% (average, annual) in order to break even, this equates to around £878 million average annual benefit over the appraisal period. Given the difficulties in monetising the impact of online harms, this represents a very conservative approach to benefit estimation and the break even point is likely much lower.

#### Other key non-monetised benefits by 'main affected groups'

<sup>4</sup> Cyberbullying, cyberstalking, intimidation of public figures, child sexual abuse and exploitation, modern slavery, hate crime, and the sale of illegal drugs online.

<b>Key assumptions/sensitivities/risks</b>	<b>Discount rate(%)</b>	3.5%
The key assumptions for this option are: the number of businesses in scope of the regulations and the incremental cost (in terms of percentage of turnover) of complying with the requirements - all key assumptions are tested in the risks and sensitivity section.		

**BUSINESS ASSESSMENT (Option 3)**

<b>Direct impact on business (Equivalent Annual) £m:</b> illustrative only at this stage			<b>Score for Business Impact Target (qualifying provisions only) £m:</b> to be scored at secondary
<b>Costs: 814.2</b>	<b>Benefits: -</b>	<b>Net: 814.2</b>	



## Table of contents

Policy Rationale .....	11
Background.....	11
Wider international and regulatory context .....	12
Regulatory context.....	12
Domestic context .....	13
International context.....	13
Rationale for intervention .....	15
Context.....	15
Negative Externalities .....	18
Information Asymmetry .....	18
Government Intervention .....	19
Harmful content.....	19
Online fraud .....	20
Policy objectives .....	21
Options considered.....	21
Option 0 – do nothing (baseline).....	22
Option 1 – limited risk-based scope.....	24
Option 2 - Full risk-based scope .....	25
Option 3 - uniformly applied safety duties .....	26
Preferred option and implementation plan.....	26
Costs and Benefits.....	27
Main sources of evidence .....	27
Approach.....	28
Summary of options.....	29
Option 0 – <i>do nothing</i> (baseline).....	29
Costs to business.....	30
Number of businesses in scope.....	30
Risk categorisation of organisations in scope .....	33
Duties and requirements on businesses.....	34
Transition costs.....	36
Compliance costs.....	43
Industry fees .....	60
Regulator enforcement powers.....	61
Costs to individuals .....	64
Costs to government.....	65
Justice impacts .....	65
Requirement to report online CSEA .....	66
Benefits .....	68
Treatment of benefits .....	68

Methodology .....	68
Quantifying online harms in the baseline.....	70
Intimidation of public figures .....	78
Break-even analysis.....	80
Indirect Costs and Benefits .....	84
Freedom of expression .....	84
Privacy impacts.....	85
Summary of impacts .....	85
Business Impact Target Calculations .....	86
Calculations .....	87
Risks and Assumptions .....	88
Policy Risks.....	88
Analytical risks and assumptions.....	90
Sensitivity Analysis .....	94
Small and Micro Business Assessment.....	98
Justification for non-exemption of SMBs under the preferred option .....	98
Impacts on SMBs.....	99
Mitigations for SMBs.....	100
Wider impacts .....	104
Trade impacts .....	104
WTO notification .....	106
Innovation Test .....	106
Equalities Impact Assessment.....	108
Competition.....	110
Devolution Test.....	111
Monitoring and Evaluation .....	112
Evaluation plans.....	112
Review clause.....	112
Who will conduct the review? .....	112
What will the review consider? .....	112
Annex A: Changes to the policy since the OHWP .....	114
Annex B: Business and User Support Measures .....	117
Annex C: A detailed overview of the chosen policy position .....	119
Annex D: Business stakeholder survey questions.....	138
Annex E: Rapid evidence assessment of NetzDG methodology .....	143
Annex F: Consultation questions .....	145

# Policy Rationale

## Background

1. As the use of the internet has changed, there has been a recognition that the piecemeal and inconsistent regulatory oversight of providers of services online is no longer sufficient. The online safety framework aims to increase the safety of users online, primarily by protecting them from harms that would be perpetrated against them by other users, and doing so by ensuring platforms have the right systems and processes in place to protect their users.
2. In October 2017, DCMS published the Internet Safety Strategy green paper. The strategy considered the responsibilities of organisations to their users, the use of technical solutions to prevent online harms and the government's role in supporting users.
3. The Online Harms White Paper (OHWP) was published in April 2019 and set out the government's ambition to make the UK the safest place in the world to go online, and the best place to grow and start a digital business.<sup>5</sup> It described a new regulatory framework establishing a duty of care on businesses to improve the safety of their users online, overseen and enforced by an independent regulator. The OHWP proposed that regulation should be focussed on platforms that allow users to share or discover user-generated content or interact with each other online. Focusing on the services provided by companies, rather than their business model or sector, limits the risk that online harms simply move and proliferate outside of the ambit of the new regulatory framework. Further policy work has refined this definition and the new regulatory framework will now apply to platforms that:
  - Host user generated content which can be accessed by users in the UK; and/or
  - Facilitate private or public interaction between service users, one or more of whom is in the UK; and
  - Provide search engines

### User generated content and user interaction

#### User Generated Content

- Digital content (including text, images and audio) produced, promoted, generated or shared by users in an online service.
- Content may be paid-for, or free, time-limited or permanent. It must have the potential to be accessed, viewed, consumed or shared by people other than the original producer, promoter, generator or creator.

#### User Interaction

- Any public or private online interaction between service users with the potential to create and promote user generated content.
- Interaction may be one-to-one or one-to-many and may involve means other than text, images and audio.

In both cases 'user' refers to any individual, business or organisation (private or public) that produces, promotes, generates, shares or accesses content on a service. Users may be members, subscribers or visitors to the service and may generate content or interact directly through an intermediary, such as an automated tool or bot.

---

<sup>5</sup> Throughout this IA the terms 'business', 'platform' and 'organisation' are used interchangeably to refer to all in-scope businesses.

4. The regulation will be structured to be proportionate and risk-based, ensuring organisations have appropriate systems and processes in place to tackle harmful content and activity. The OHWP also made clear that the framework will protect users' rights, including freedom of expression online.
5. The government subsequently undertook formal consultation on the policy and gave an indication of its direction of travel in a number of key areas in the OHWP - Initial Government Response<sup>6</sup>, published in February 2020. Here, the government reconfirmed its commitment to the duty of care approach set out in the OHWP and announced a number of further measures to guarantee proportionality and protect freedom of expression. It also indicated that the government was minded to appoint OFCOM as the regulator.
6. Following this, further work was undertaken to develop and refine policy with a number of important changes made to the policy. The full intended policy position was set out in the full government response published on 15 December 2020 along with confirmation that OFCOM would be named as the regulator.
7. Since the OHWP was published in April 2019 there have been numerous changes to the policy position in response to stakeholder concerns. Key changes included:
  - The assurance of robust protections for journalistic content
  - Specific exemptions for low-risk services including reviews and comments on directly published content
  - Certain categories of harmful content (for example, advertising) to be excluded from regulatory scope in order to prevent regulatory duplication<sup>7</sup>
  - A refined definition of duty of care covering harms to individuals but not to society more broadly to provide more clarity for businesses
  - The introduction of specific provisions targeted at building understanding and driving action to tackle disinformation and misinformation
  - Further detail and clarity on what enforcement powers will look like
  - Further developing the differentiated approach to tackling harms; only the highest risk and highest reach organisations providing Category 1<sup>8</sup> services will have to take action in respect of adult users accessing legal but harmful content on their services.
8. See **Annex A** for further detail on changes made to the policy in response to stakeholder feedback.

## Wider international and regulatory context

### Regulatory context

9. Online harms is part of the government's wider strategic approach to regulating digital technology. It will help improve user safety online, build public trust in digital services and support innovation to drive digital growth. Action is being undertaken in a range of different areas - including data, cybersecurity, competition and protecting quality journalistic content - to improve online safety and security, support fair and efficient digital markets and protect our democratic values online. The forthcoming DCMS digital strategy will set-out how the government is bringing these strands together.

---

<sup>6</sup> [Online Harms White Paper - Initial Response \(December 2020\)](#)

<sup>7</sup> In 2019, the Secretary of State for Digital, Culture, Media and Sport announced a review of the way that the online advertising market is regulated in the UK, which is being considered through the Online Advertising Programme. As part of the Online Advertising Programme, the Department for Digital, Culture, Media and Sport will launch a public consultation on measures to enhance how online advertising is regulated in the UK in the first half of 2021.

<sup>8</sup> Category 1: High risk, high reach businesses

10. Tackling online harm cannot be done alone, it requires close cooperation with our international partners across the world to coordinate resources and expertise. The UK, with its strengths in digital innovation, highly respected legal system, business friendly environment and world class regulators has an opportunity to shape the agenda and act as a global leader in this space.

## Domestic context

12. The domestic regulatory environment currently stems from EU law and directives which are set-out below. The EU has recently undertaken significant work across a range of digital regulatory issues. Although we are no longer bound by EU law, these directives form the basis of the current regulatory landscape.
13. **e-Commerce Directive:** The 2000 e-Commerce Directive (Directive 2000/31/EC) (eCD) applies to information society services, which covers the vast majority of online service providers and includes provisions that protect platforms from liability for illegal content they host, provided they remove or disable access to illegal material 'expeditiously' once they have 'actual knowledge' of it.
14. **Audio Visual Media Services Directive (AVMSD or Directive):** The Directive is designed to provide minimum standards and market access for cross border broadcasters throughout the European Economic Area. In 2010, AVMSD expanded in scope to include Video on Demand services (such as Netflix etc). AVMSD 2020 (Directive (EU) 2018/1808) introduced rules for video sharing platforms for the first time. Last year, the government announced OFCOM as the national regulator for UK-established video sharing platforms. The UK transposed the revised Directive through the Audiovisual Media Services Regulations 2020 which came into force on the 1st of November 2020. The revised Audiovisual Media Services Regulations 2020 place requirements on UK-established video sharing platforms to protect all users from illegal content through taking appropriate measures. UK-established video sharing platforms are also required to take measures to protect minors from harmful content which may impair their physical, mental or moral development. As the regulations share broadly similar objectives to the online safety regime, the government's preference is for the requirements on UK-established video sharing platforms to transition to, and be superseded by, the online safety regulatory framework, once the latter comes into force.
15. **e-Privacy Directive:** The ePrivacy Directive (Directive 2002/58/EC) was agreed at EU level in 2002, and transposed in the UK as the Privacy and Electronic Communications (EC Directive) Regulations 2003 (SI 2003/2426) (PECR). The Directive, which has been amended several times since, aims to protect the privacy of electronic communications, reduce the incidence of nuisance calls, and restrict website and app developers' use of 'cookies' to track user activity.

## International context

16. Many countries are considering how to make the internet safer for users and some governments are taking action by introducing legislative measures to tackle harmful online content. Internet safety is also being discussed in a range of multilateral and multi-stakeholder fora. Recent initiatives include the Christchurch Call to Action, the work of the Global Internet Forum to Counter Terrorism, the WeProtect Global Alliance and the Technology Coalition Fighting Child Sexual Abuse, and the upcoming OECD Recommendation on the Protection of Children in the Digital Environment.
17. The government is working closely with many international partners to address this shared challenge in order to build consensus around shared approaches to internet safety and to learn from others nations' experiences of tackling online harms. Through the UK's presidency of the G7, the government is bringing countries together to promote proportionate and risk-based solutions to address harmful online activity that uphold our shared values. Some examples of international legislative approaches to tackling harmful online content are set out below.

18. **Ireland:** The Irish Online Safety and Media Regulation Bill (General Scheme published December 2020) is designed to implement both the AVMSD and new online safety provisions. Ireland intends to create a Media Commission which will take on both the new online safety responsibilities and the functions of the existing Broadcasting Authority of Ireland, and proposes to create a new Online Safety Commissioner. The Online Safety Commissioner would have the power to designate as in scope any online service or categories of online services that allow users to share, spread or access content that other users have made available. The Irish Bill includes provisions empowering the proposed Commissioner to draft online safety codes; assess the compliance of online services with those safety codes; direct online services to make changes to their systems, processes and policies and design, and seek to apply financial sanctions to services who fail to comply.
19. **Germany:** The German Act to Improve Enforcement of the Law in Social Networks (NetzDG), which came into full force in January 2018, requires social media platforms with more than 2 million registered users in Germany to remove 'manifestly unlawful' content within 24 hours of receiving a notification or complaint, and remove all other 'unlawful' content within seven days of notification or risk receiving a fine of up to 50 million euros.
20. **Australia:** The Australian Online Safety Bill, introduced to Parliament in February 2021, aims to promote the online safety of Australians, and grants enhanced powers to the eSafety Commissioner (Australia's online content regulator) to administer complaints related to cyber bullying of children, serious online abuse of adults, and to order the take down of harmful online content. The Bill contains a set of core online safety expectations for social media services, relevant electronic services and designated internet services, clearly stating community expectations, with mandatory reporting requirements. It also includes new abhorrent violent material blocking arrangements that allow the eSafety Commissioner to respond rapidly to an online crisis event such as the Christchurch terrorist attacks, by requesting internet service providers block access to sites hosting seriously harmful content.
21. **France:** France's law against hate content online (also known as the Avia Law), which aimed to push online platforms to remove hateful content effectively, is being reconsidered following a Constitutional Council ruling in June 2020 which removed a number of its key aspects because of concerns around freedom of expression online. The French Government has subsequently put forward a new legislative approach (presented to Parliament January 2021), based on the EU's Digital Services Act. The new law will require online platforms that sort, reference or share third party content to be clear about how they tackle illegal content and be accountable to the French Communications Regulator or face fines. Under the proposed law, companies would be required to have clear terms and conditions including moderation and user redress processes, companies would be responsible for conserving content for law enforcement and for assessing risks around both tackling illegal content and breaching freedom of speech.
22. **European Union:** The European Commission in December 2020 published the Digital Services Act, which, once adopted, will be directly applicable across the EU and will update liability and safety rules for digital platforms. The Act proposes new rules to increase the responsibilities of online intermediary services and reinforce oversight over platforms content policies. It has three key objectives; to protect consumers and their fundamental rights online, establish a transparency and accountability framework for online platforms, and foster innovation, growth and competitiveness within the single market. These rules will apply to intermediary services provided to recipients of the service that have their place of establishment or residence in the European Union, irrespective of the place of establishment of the providers of those services.

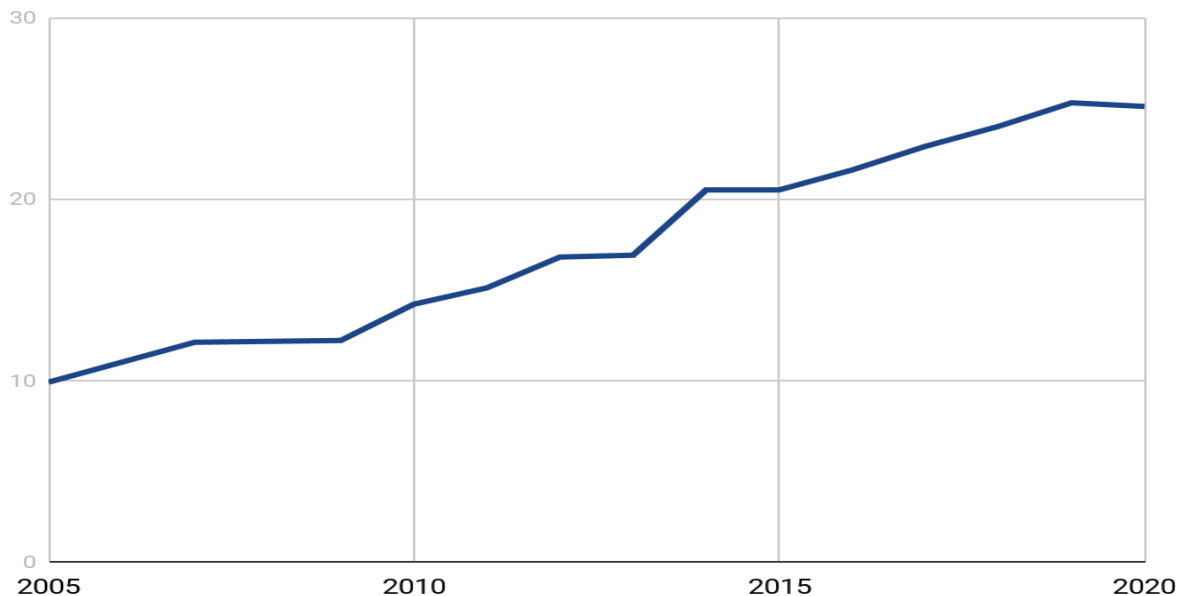
# Rationale for intervention

## Context

23. **Internet use is growing and evidence indicates that Covid-19 has increased this even further.** During April 2020 internet users in the UK spent an average of 4 hours and 2 minutes online each day, this is up from 3 hours 29 minutes in September 2019.<sup>9</sup> This increase in internet use during the pandemic is part of a wider trend. Adult internet use in the UK, among those who had been online in the last three months, increased from 80.9% in 2012 to 90.8% in 2019.<sup>10</sup> The internet is an integral part of everyday life for most people in the UK. In 2020, adult internet users estimated they spent an average of 25.1 hours online per week.<sup>11</sup>

**Figure 1: Weekly time spent online by UK adults (hours)<sup>12</sup>.**

This line graph illustrates the increasing amount of time spent online each week by UK adults, from 9.9 hours in 2005 to 25.1 hours in 2020.



24. **Parents find it increasingly difficult to ensure their children stay safe online.** Only 57% of parents of 12-15 year olds think that their child has a good balance between screen time and other things.<sup>13</sup> OFCOM's research reveals that there has been a steady decline over the past few years in the proportion of parents of online 5-15 year olds who agree that 'the benefits of the internet for my child outweigh any risks'; just over half (55%) agree with this in 2019, compared to two-thirds in 2015<sup>14</sup>. Just under half of parents of online 3-4 year olds agree (43%). Since 2017, there is less awareness of content filters provided by ISPs among parents of online 3-4 year olds and 5-15 year olds (from 66% to 56% for parents of 3-4s; and from 62% to 60% among parents of 5-15s).<sup>15</sup>

25. **Covid-19 is very likely to have driven these numbers up.** Three-quarters of British parents have reported that their child's screen time averaged nine hours per day at the height of the first

<sup>9</sup> [Online Nation 2020 - summary report](#) - OFCOM

<sup>10</sup> [Internet Users \(May 2019\)](#) - ONS. Internet use here refers to respondents who have used the internet in the last three months.

<sup>11</sup> [Adults' Media Use and Attitudes Report 2020](#) - OFCOM

<sup>12</sup> [Adults' Media use and Attitudes report' \(2015-2020\)](#) - OFCOM

<sup>13</sup> [Children and parents: media use and attitudes report 2019](#) (OFCOM)

<sup>14</sup> [Children and parents: media use and attitudes report 2019](#) (OFCOM)

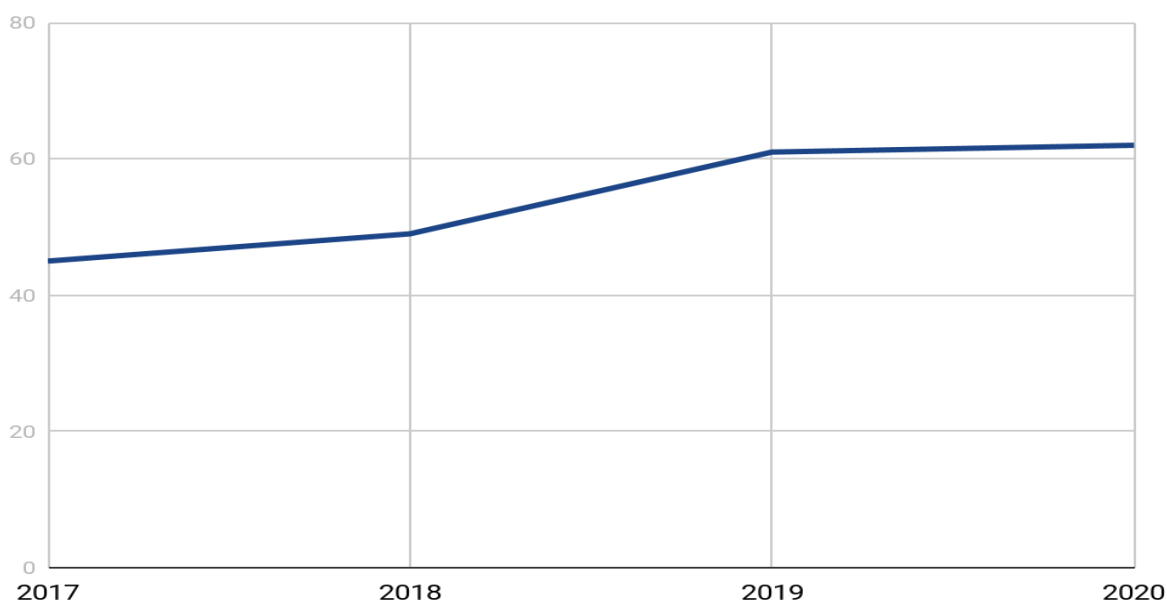
<sup>15</sup> [Children and parents: media use and attitudes report 2019](#) (OFCOM)

lockdown – nearly double the screen time average prior to the outbreak.<sup>16</sup> Furthermore, the Internet Watch Foundation has had a record number of public reports of suspected child sexual exploitation and abuse (CSEA). In September 2020 analysts processed 15,258 reports from members of the public - 45% more than the previous year.<sup>17</sup>

26. **While the internet is a powerful force for good, illegal and harmful content and activity is widespread online.** UK users are concerned about the content they interact with and their experiences on the internet. 62% of adult internet users have had at least one potentially harmful online experience in the last 12 months - worryingly this figure increases to over 80% for 12-15 year olds.<sup>18</sup> There has also been a 27% increase of reported online abuse during COVID-19.<sup>19</sup>
27. **The ease with which online platforms can be used allows for criminals to act undetected.** Online platforms are consequently used as a tool for abuse, acting as a medium for the sale of illegal drugs and promotion of extremist content. 3% of UK adults and 5% of children aged 12-15 have encountered material online promoting terrorism/radicalisation.<sup>20</sup>

**Figure 2: Adult internet users that have had at least one potentially harmful experience online in the past 12 months (per cent)<sup>21</sup>.**

This line graph illustrates the increasing percentage of adult internet users that have had at least one potentially harmful experience online in the past 12 months, from 45% of adults in 2017 to 65% of adults in 2020.



28. **The scale of CSEA content online is alarming.** There were more than 69 million images and videos related to child sexual exploitation and abuse referred by US technology companies to the National Center for Missing and Exploited Children in 2019,<sup>22</sup> an increase of more than 50% on the previous year.<sup>23</sup> During the Covid-19 pandemic, increases have been seen in online harms, for example in online activity relating to CSEA material.<sup>24</sup> Europol have also reported a surge in

<sup>16</sup> [Study suggests lockdown could have permanently altered families' tech habits](#) (October 2020)

<sup>17</sup> [IWF has record month as public reports of child sexual abuse surge](#) (October 2020)

<sup>18</sup> Internet Users' Experience of Harm Online, 2020, OFCOM and ICO

<sup>19</sup> Glitch and EVAW 'The Ripple Effect, COVID-19 and the Ripple Effect of Online Abuse' (2020)

<sup>20</sup> [Internet users' concerns about and experience of potential online harms](#) - OFCOM (June 2020)

<sup>21</sup> [Internet users' concerns about and experience of potential online harms](#) - OFCOM (2017-2020)

<sup>22</sup> [CyberTipline](#)' NCMEC (2019)

<sup>23</sup> [Tech Companies Detect a Surge in Online Videos of Child Sexual Abuse](#) - The New York Times (February 2020)

<sup>24</sup> [Covid-19 Child Sexual Abuse and Exploitation threats and trends](#) - Interpol (2020)

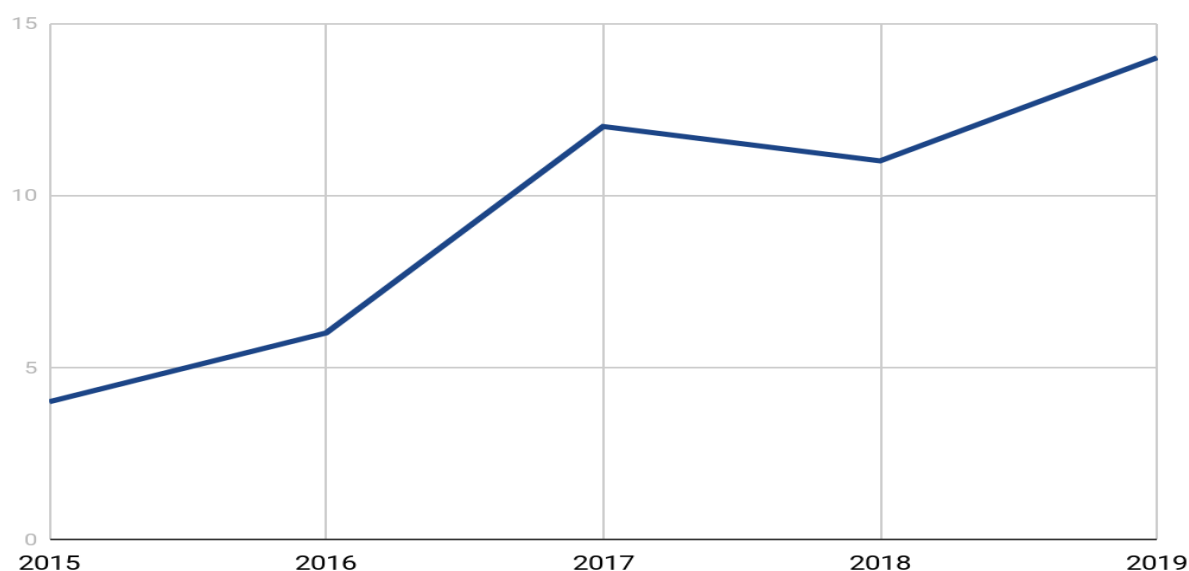


CSEA material online during the pandemic.<sup>25</sup>

29. **Legal but harmful content is also widespread.** Online advocacy of self-harm poses a clear threat to people's wellbeing. In 2015, a study of 4,000 young adults (aged 21) found that self-harm/suicide related internet use<sup>26</sup> (coming across, searching for or discussing self-harm/suicide) was reported by 22.5% of participants, of those, 8.2% and 7.5% had actively searched for information about self-harm and suicide respectively.<sup>27</sup> The prevalence of using the internet to view related content has been found to be higher in children than adults. One study of those presenting to hospital following self-harm found that 26% of children had viewed self-harm and suicide content, compared to 8.4% of adults.<sup>28</sup>
30. **In recent years there has been a rise in abuse, harassment and intimidation directed towards public figures.** An international survey of female journalists found 64% had experienced online abuse – death or rape threats, sexist comments, cyberstalking, account impersonation, and obscene messages.<sup>29</sup> Almost half (47%) did not report the abuse they had received, and two fifths (38%) said they had self-censored in the face of this abuse.
31. **Cyberbullying is also a cause for concern.** In 2017, one in five children surveyed aged 11-19 reported having experienced cyberbullying in the past year.<sup>30</sup> The prevalence of cyberbullying is higher for some groups, such as women, religious minorities, LGBT+, BAME and people with disabilities.<sup>31</sup> 1 in 4 young people now have anticipatory anxiety about being abused online, this highlights how detrimental this issue is to the wellbeing of children.<sup>32</sup>

**Figure 3: Children 12-15 that experienced cyberbullying (per cent)<sup>33</sup>.**

This line graph illustrates the increasing percentage of children aged 12-15 that have experienced cyberbullying, from 4% of children in 2015 to 14% of children in 2019.



<sup>25</sup> [Europol Covid Report](#) - Europol (2020)

<sup>26</sup> Encouraging or assisting suicide is illegal under Section 2 of the Suicide Act 1961; however, content that relates to suicide but is not an offence under the Act could be considered legal but harmful.

<sup>27</sup> [Exposure to, and searching for, information about suicide and self-harm on the Internet: Prevalence and predictors in a population based cohort of young adults](#) - Mars et al (2015)

<sup>28</sup> [Suicide and Self-Harm Related internet Use](#) -Padmanathan et al. (2018).

<sup>29</sup> [IFJ global survey shows massive impact of online abuse on women journalists](#) - IFJ (November 2018)

<sup>30</sup> [Mental Health of Children and Young People in England, 2017](#) - NHS Digital (2018)

<sup>31</sup> [The Annual Bullying Survey 2017](#) - Ditch the Label (2017).

<sup>32</sup> Ibid.

<sup>33</sup> [Adults' Media use and Attitudes report' \(2015-2019\)](#) - OFCOM

## Negative Externalities

32. **Online harms encompass a number of negative externalities, both illegal and legal but harmful content has consequences beyond that of the direct impact upon the victim.** For example, online platforms are used by criminals to sell illegal goods and services and the sale of weapons online is a contributing factor to violence, such as knife crime.<sup>34</sup> The online sale of dangerous and unregulated opioids, such as fentanyl, is of particular concern given the wide ranging impacts on the health service and wider society - there were over 160 deaths with illicit fentanyl appearing on the toxicology report between 2016 and 2019 in the UK<sup>35</sup>.
33. **Harms such as these disproportionately affect young and vulnerable users of online platforms.** Vulnerable people, including those with chronic pain and those suffering depression, are targeted and put at risk of accidental overdose.<sup>36</sup> There is also a risk that fentanyl may be used to assist suicide.<sup>37</sup> Whilst these products are readily available online, the potential for harm will remain.
34. **CSEA has damaging impacts not only on the victim's welfare but to the wider society.** Research has found that being a victim of CSEA is associated with an increased risk of adverse outcomes in all areas of a victim's life. These adverse outcomes include physical injuries, problems related to childbirth, trauma, anxiety, depression, substance misuse, offending, lower educational attainment, higher unemployment, and homelessness - the majority of which affect not only the victim but wider society as well. In addition, long-term longitudinal research indicates that in many cases these outcomes can endure over a victim's lifetime.<sup>38</sup>
35. **Legal but harmful activity, such as cyberbullying, can also lead to impacts beyond the direct effect on the victim's welfare.** Secondary effects of cyberbullying include depression, self-harm and life-long impacts for the victims. An estimated 37% of victims go on to suffer depression as a result and 41% develop social anxiety.<sup>39</sup> In some cases these harms deter people from using online platforms, 26% of people deleted their social media profile after experiencing cyberbullying.
36. **The risk of abuse and cyberbullying also impacts how an individual uses online platforms.** Half of girls aware of sexist abuse on social media say this has restricted what they do or aspire to in some way.<sup>40</sup> The House of Commons Petitions Committee has highlighted the extreme abuse experienced online by disabled people, which has forced some of them to leave social media.<sup>41</sup> Due to the large user bases of online platforms, and increasing dependence on technology, these harms affect a considerable proportion of the population.

## Information Asymmetry

37. **In addition to negative externalities, there is an information asymmetry between users of online platforms and the platforms themselves.** While some businesses do claim that they are transparent as possible, on the whole, there is a lack of transparency between the businesses and consumers of online platforms about the potential harms. The extent of harmful content on online platforms remains unclear for many individuals, especially parents of children who regularly use the internet.
38. **OFCOM and the Information Commissioner's Office (ICO) reported 62% of adults and 81% of 12-15 year old internet users have had at least one potentially harmful experience online**

---

<sup>34</sup> [Police announce new partnership to help combat knife crime](#) - NPCC (February 2020)

<sup>35</sup> [The use of the clearnet and social media for the advertising and sale of fentanyl analogues](#) - NCA (July 2019)

<sup>36</sup> [Online Harms White Paper](#), NCA analysis (2019)

<sup>37</sup> [Online Harms White Paper](#), NCA analysis (2019)

<sup>38</sup> [The impacts of child sexual abuse: A rapid evidence assessment](#) - IICSA Research Team (2017)

<sup>39</sup> [The Annual Bullying Survey 2017](#) - Ditch the Label (2017).

<sup>40</sup> [Girls Attitudes Survey 2016](#) - Girl Guiding

<sup>41</sup> [Online abuse and the experience of disabled people](#). The Petitions Committee, 2019.

**in the past 12 months.**<sup>42</sup> Children's attitudes towards the privacy of their online profiles highlights their lack of understanding of potential exposure to online harms. Around one-third of 12-15 year olds know how to change settings on their social media profile so fewer people can view it or know how to block junk email or spam with these actions actually being done by only approximately 15%.<sup>43</sup> It is therefore difficult for users to make an informed decision as to how they use online platforms and what content they access. There is also insufficient transparency about the level of investment and the effectiveness of different interventions carried out by online platforms in addressing these harms.

## Government Intervention

39. **Online platforms have failed to effectively address online harms.** In the absence of regulations, harmful online content is addressed on a voluntary basis, predominantly through business terms of use. While there have been examples of platforms putting in place safety mechanisms, such as more robust content moderation, these moves have been inconsistently applied across industry and are often rolled out due to regulatory or government pressures. The limitations of the current model are acknowledged by the industry itself: nearly half of tech industry workers (45%) believe that the industry is currently under-regulated. Only 2% see voluntary commitment as the most effective way of mitigating potential harms.<sup>44</sup>
40. **There is potentially a trade-off between encouraging traffic to a site and ensuring the safety of all users.** For example, there is a potential economic incentive for platforms not to address content such as fake news. Research suggests that false news is 70% more likely to be retweeted than real news, generating a higher profit for platforms.<sup>45</sup> Removing this content could therefore result in a short-term loss of profit and reduced user engagement. However, there is also a question of sustainability. It is in the businesses' long-term interest to prevent the platforms from an overcrowding of fake and unreliable news, diminishing the quality of the service provided. This creates a trade-off between short-term profit and long-term sustainability of the quality of content. There is a lack of incentive for businesses to introduce new systems to keep their platforms safe.<sup>46</sup> Therefore, without government intervention limited progress will be made at reducing online harms.
41. **A clear, proportionate and predictable regulatory framework is a good thing for businesses looking at where to start up, grow and invest.** Many other countries are also planning to introduce online regulation. By acting first we will be able to set the benchmark & reduce uncertainty. We have the opportunity to set global standards, unlock investment and influence the global approach.
42. **This proposal will rebuild public confidence and set clear expectations of businesses, allowing UK users to enjoy more safely the benefits that online services offer.** Given the prevalence of illegal and harmful content online, and the level of public concern about online harms, the digital economy urgently needs a new regulatory framework to improve UK users' safety online.

## Harmful content

43. The Online Safety Bill (OS Bill) seeks to address the following broad categories of harmful online content:

---

<sup>42</sup> [Internet users' experience of potential online harms: summary of survey research](#) - OFCOM and ICO (2020)

<sup>43</sup> [Report on Internet Safety Measures](#) - OFCOM (2015)

<sup>44</sup> [People, Power and Technology: The Tech Workers' View](#) - DotEveryone (May 2019)

<sup>45</sup> [The spread of true and false news online](#) - Vosoughi et al. (March 2018)

<sup>46</sup> [Hate crime: abuse, hate and extremism online](#) (2017)

- **illegal user generated content and activity:** user generated content and activity which is an offence under UK law - such as child sexual exploitation and abuse, terrorism, hate crime and sale of illegal drugs and weapons
  - **legal but harmful user generated content and activity:** user generated content and activity which may not be illegal under all circumstances, but which gives rise to a foreseeable risk of psychological and physical harm to adults - such as abuse or eating disorder content.
  - **underage exposure to user generated content and activity which gives rise to a foreseeable risk of psychological and physical harm to children** - such as pornography, violent content.
44. The OHWP set out that this Bill would not seek to address user generated content which gives rise to a foreseeable risk of harm to corporations and organisations and their interests (e.g. copyright offences, competition law).
45. In addition, and in line with the position set out in the OHWP, a number of categories of user generated content and activity will be specifically excluded from the scope of the Bill because there are existing legislative, regulatory and other governmental initiatives in place - for example breaches of data protection legislation, breaches of consumer protection law, and cyber security breaches or hacking.
46. The OHWP provided an initial more specific list of 'harms' (specific offences and other sub categories of user generated content and activity) that would be in scope such as harassment and cyber-bullying. However, it stated that this list was, by design, neither exhaustive nor fixed. A static list could prevent swift regulatory action to address new types of harms.

## Online fraud

47. In its full response to the OHWP consultation, the government reiterated its deep concern about the growth, impact and scale of online fraud and recognised the considerable harm these types of fraud can cause. There were 3.7 million instances of fraud in England and Wales in the year ending March 2020<sup>47</sup> and over half of these had some online element<sup>48,49</sup>. It is clear that action must be taken across government to tackle the UK's most common crime type.
48. However, the government determined at that time that the fraud threat would be most effectively tackled by other mechanisms and, as such, clarified that the legislation would not require companies to tackle online fraud.
49. Having engaged extensively with a broad range of stakeholders, including the financial industry, consumer groups and representatives, law enforcement, and other public bodies, the Secretary of State has listened to stakeholder views and confirmed their intention to bring fraud into scope of the Online Safety Bill.
50. As the new regulatory framework will be limited to tackling harm facilitated through user-generated content, the government expects the regulatory framework to have a particular impact on some types of fraud, for example, romance scams. In its 2020 strategic assessment of serious and organised crime<sup>50</sup>, the NCA noted that romance fraud continues to have a high financial and emotional impact on victims, with reported victim losses of over £60 million in the year ending February 2020. However, the inclusion of user-generated fraud in scope of the Online Safety Bill

---

<sup>47</sup> [Crime Survey for England and Wales](#) - year ending March 2020

<sup>48</sup> [Nature of crime: fraud and computer misuse](#) - year ending March 2020

<sup>49</sup> Cyber fraud represents cases where the internet or any type of online activity was related to any aspect of the offence.

<sup>50</sup> [National Strategic Assessment of Serious and Organised Crime](#) - 2020

alone will not solve the problem of online fraud. The government will continue its work exploring other legislative and non-legislative options to tackle fraud coherently and holistically.

51. Estimates presented in this impact assessment do not yet reflect the inclusion of user-generated fraud. The government is continuing to assess the likely costs and benefits, including wider impacts, potential justice impacts, and exploring the extent to which user-generated content is used as a vector for fraud.
52. Including fraud in scope is likely to increase estimated costs to businesses presented within this IA through, for example, potential additional content moderation and user reporting costs. However, given the high cost of online fraud on individuals and society, even a relatively modest reduction resulting from its inclusion in the Bill would represent significant cost savings against a do nothing baseline. The government will work with stakeholders before introduction of the Bill to improve the evidence base and will seek to reflect the inclusion of fraud within the final stage impact assessment.

## Policy objectives

53. The policy objectives are:

- **User safety:** improved safety of users online, particularly through reduced risk and incidence of specific online harms, especially with respect to vulnerable groups. The regulatory response will rebuild public confidence when using online platforms allowing citizens to enjoy more safely the benefits that online services offer.
- **Preserving freedom of speech:** ensuring sufficient safeguards for freedom of expression will be evidenced through the collection and reporting of transparency data and user satisfaction.
- **Law enforcement:** improving the efficacy of law enforcement and crime prevention with respect to illegal content and behaviour online. This can be measured using crime data and will also rely on better understanding of the drivers of crime including the specific role of activities in scope in facilitating those crimes.
- **Efficiency:** increased coherence and clarity of government activity to tackle online harms and build the capability of users to stay safe online.
- **Evidence:** a culture of transparency enhancing the amount and quality of information in relation to online harms that is available to government, industry, civil society and wider society.

54. **DCMS has focused on commissioning research in order to support the evidence base for the Online Safety Bill.** DCMS is funding a two stage project investigating the feasibility of research to assess the drivers and impact of online harms and then leading to specific research to assess child online safety and online abuse (including anonymous abuse) in more detail. The initial findings from this research will help to provide a more robust evidence baseline of online harms to support OFCOM in its implementation of the regime. In addition, the work DCMS is undertaking on online harms data through HMT's Shared Outcome Fund will develop universal taxonomies for online harms data which should improve the evidence base in the future.

## Options considered

55. **This consultation stage IA considers three distinct policy options; however, a range of options were considered as part of the earlier policy development process.**

56. The short-list of options which this IA appraises are:

- **Option 0 - do nothing (baseline)**
- **Option 1 - limited risk-based scope:** a duty of care for user generated content and activity. The duty of care sets out organisations' responsibilities in addressing illegal harms and the safeguarding of children from both illegal and legal but harmful content and activity. Unlike **Option 2**, this option does not address legal but harmful content and activity accessed by adults. Under **Option 1**, duties are set out in primary legislation (and subsequent secondary) and guidelines or codes of practice. The regime is overseen, monitored and enforced by an independent regulator, which applies a risk-based approach to its core activities.
- **Option 2 - full risk-based scope:** a duty of care for user generated content and activity. The duty of care sets out organisations' responsibilities in addressing illegal harms and legal but harmful content, and the safeguarding of children from both illegal and legal but harmful content and activity. In addition to the scope of **Option 1**, the preferred option addresses legal but harmful content and activity accessed by adults. Under **Option 2**, duties are set out in primary legislation and codes of practice. The regime is overseen, monitored and enforced by an independent regulator, which applies a risk-based approach to its core activities.
- **Option 3 - uniformly applied safety duties:** Detailed safety duties setting out organisations' responsibilities in addressing illegal harms and legal but harmful content, and the safeguarding of children from both illegal and legal content. These safety duties are detailed in primary legislation and are uniformly applied across all harms and organisations in scope.

57. **Option 2 is the preferred option.** All options are assessed against a *do nothing* counterfactual.

58. This IA does not specifically consider a non-regulatory option as an alternative, although this was considered in the long list appraisal. Self-regulation and voluntary approaches to tackle harms were considered but given the wide-ranging and significant societal impacts of online harms, inconsistent current voluntary actions, and competing market incentives (as evidenced in the rationale for intervention), the government does not consider non-regulatory approaches on their own to be appropriate. However, regulation will only be one part of the solution. Alongside online safety legislation, the government is pushing forward with a number of non-regulatory business and user support measures. These non-regulatory measures will both support legislative implementation and growth as well as innovation across the UK's burgeoning safety-tech sector, creating the right conditions for UK safety tech businesses to deliver cutting edge technologies to safeguard users and prevent harm. A full description of the proposed business and user support measures can be found in **Annex B**.

## Option 0 – do nothing (baseline)

59. **The do nothing option is not able to deal with the current policy problem.** Where legal frameworks do exist around illegal content online (such as the intermediary liability provisions under the eCommerce Directive, or existing criminal law for specific harms), a significant increase in resources for reporting and law enforcement would be needed to tackle the problem sufficiently. There is no existing legal framework to tackle the policy problem of harm being caused to children or adults through content and activity which is harmful but not illegal.

60. **Alongside reporting to the internet platform, there are a number of other routes for individuals to report content that they believe to be illegal online.** For example, the UK based Internet Watch Foundation provides a mechanism for individuals to anonymously and confidentially report online child sexual abuse content. True Vision provides an online mechanism for the reporting of hate crimes and incidents online. There are also government website tools for the reporting of online material promoting terrorism or extremism.

61. **The current systems, especially relating to legal but harmful content and activity rely on voluntary action by social media companies.** Under existing regulations, there is very little a



user can do in terms of seeking redress and there is no regulatory oversight of a platform's enforcement of their own terms of service.

62. **In principle, an individual can bring a claim for breach of contract (either in the local small claims court or, for larger, more complex cases, the High Court) if they consider that an internet platform has breached any of the terms of service.** Broadly, the individual would need to demonstrate that: (i) a contract exists between the individual and the internet platform, (ii) the contract was breached as the platform failed to fulfil its obligations satisfactorily, (iii) directly as a result of the breach, the individual suffered a loss and, (iv) should be compensated.
63. **An individual - who need not be a user in a contractual relationship with a platform - could also bring an action in negligence if they can demonstrate: (i) the internet platform owed them a duty of care, (ii) which it breached, (iii) which caused the individual to suffer loss or harm, and (iv) which was reasonably foreseeable.**
64. **In the event of a contractual breach, an individual can seek to recover damages for consequential loss, including personal injury.** Damages for non-monetary loss which don't amount to personal injury (e.g. mental distress or loss of amenity) are awarded only in exceptional cases. Awards of damages for non-monetary loss are more common in negligence claims. Pain, suffering and loss of amenity, and mental distress, are recognised as separate heads on which to bring a claim for non-monetary losses in tort.
65. **Although an individual could bring a claim against an internet platform to seek redress, the government is not aware of any cases having been brought on contractual or negligence grounds (whether successful or otherwise).** This likely reflects the practical and evidential challenges of bringing such claims, the difficulty in showing loss of a sort for which damages can be claimed, and the inevitable costs involved in legal action.
66. **Alternatively individuals have the opportunity to report harmful content to, and raise complaints and concerns about harmful online activity with, the internet platform.** But it is entirely up to the internet platform as to how it will respond, and how effective that will be, as a means of redress.
67. **The legal incentive for firms (through potential legal liability) to address both illegal and legal but harmful harms is insufficient.** There are multiple barriers to consumers seeking redress, resulting in limited legal action taken against platforms that may have been in breach of contract when failing to address harmful content. On top of this, the existing legal framework for online harms solely addresses illegal harms and not those that are legal but harmful. It is consequently up to the individual platforms to voluntarily address legal but harmful content.
68. **Under the *do nothing* option, platforms face perverse and competing incentives in relation to content moderation.** Harms such as dis- and misinformation have wide-ranging negative impacts on society; however, such content also generates a significant amount of user engagement on social media platforms. Given that false news was found to be 70% more likely to be retweeted than the truth<sup>51</sup>, there is the potential for perverse incentives to delay removal of harmful content.
69. **In contrast, some platforms will face incentives to address harmful content in order to maintain advertising revenue; however, demand for advertising spaces on the main social media platforms is relatively inelastic.** In 2019 it was projected that by 2020 UK advertisers would be spending almost two-thirds of their budget online<sup>52</sup> and it is unlikely that this would be significantly affected by a platform's moderation activities. Further to this, it is difficult for advertisers to move away from popular platforms, smaller platforms cannot offer advertisers such a large and engaged user base. This is also the case for video-sharing platforms where

---

<sup>51</sup> [The spread of true and false news online](#) - Vosoughi et al. (March 2018)

<sup>52</sup> [Almost two-thirds of UK ad spend to be online by 2020](#) - Hammett (2019)

consumers are known to use a limited number of platforms, and historically only a small number of platforms have ever achieved scale.<sup>53</sup> The main social media platforms also provide a unique method of marketing, namely UGC. UGC has been shown to have a significantly stronger impact than marketing generated content (MGC) on consumer behaviour<sup>54</sup> and there are a limited number of platforms through which this form of marketing can take place. Therefore, given the limited options available, advertisers are unlikely to migrate away from platforms should they not address harmful content.

70. **Public pressure can act as a driver of content moderation processes but this could ultimately lead to delayed and reactive approach to addressing harms.** A study of video-sharing platforms highlighted that public pressure (as it relates to brand integrity) is a driver of investment in user safety measures.<sup>55</sup> While it is right for platforms to react to user sentiment, this leaves open the possibility that approaches are delayed and only reactive to harms which attract media attention. Public pressure and a desire to maintain brand integrity is insufficient in ensuring a transparent and proactive approach to addressing harms.

## Option 1 – limited risk-based scope

71. **Option 1** would see the introduction of a risk-based regulatory regime. Under this option, organisations would have a duty of care to protect users from illegal harms and safeguard children where the service is likely to be accessed by children. Platforms would adhere to codes of practice, enforced by an independent regulator which applies a risk-based approach to its core activities. The number of businesses in scope is unchanged across each policy option and is based on research conducted for DCMS by Revealing Reality (the methodology is explained in later sections). While the number of businesses in scope does not change across the options, the harms and duties on businesses do. The table below outlines how the scope, duties and implementation of **Option 1** will work:

Table 1: **Option 1** - Scope, duties and implementation

Option 1		
Scope	Duties	Implementation
<p><b>Businesses:</b> around 24,000 businesses in scope of <b>Option 1</b></p> <p><b>Harms:</b> All harms outlined in the OHWP are addressed under <b>Option 1</b> but it does not include duties on businesses to address legal but harmful content accessed by adults.</p>	<p><b>All in-scope businesses</b> have duties to address illegal harms</p> <p><b>All in-scope businesses</b> have duties to safeguard children (if likely to be accessed by them)</p>	<p>Primary legislation with codes of practice. All codes of practice will be subject to an IA which will specifically consider the impacts on small and micro businesses.</p>

72. **For illegal offences organisations will have a duty to put in place systems and processes to identify, minimise and remove the presence of illegal content.** Businesses must also consider whether they are likely to be accessed by children. If they are, they have a duty to put in place systems and processes to protect children from harmful content.

<sup>53</sup> Understanding how platforms with video sharing capabilities protect users from harmful content online - Ernst & Young (not yet published)

<sup>54</sup> [Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content](#) - Goh et al. (2013)

<sup>55</sup> Understanding how platforms with video sharing capabilities protect users from harmful content online - Ernst & Young (not yet published)



## Option 2 - Full risk-based scope

73. **The preferred option would also see the introduction of a risk-based regulatory regime.** Under this option, in-scope businesses' duty of care would be three-fold: there would be duties on organisations to undertake risk assessments and to protect users from illegal harms; duties on the highest risk platforms to address legal but harmful content accessed by adults; and duties on all businesses to safeguard children (if the platform is likely to be accessed by children).
74. **As in Option 1, under the preferred option, platforms would adhere to codes of practice, enforced by an independent regulator which applies a risk-based approach to its core activities.** The key difference between the preferred option and **Option 1** is that under the preferred option, the highest risk organisations would have an additional duty to address legal but harmful content accessed by adults. Legal but harmful content is wide ranging and results in significant impacts. Only requiring the highest risk platforms to operate under this duty is the right balance between safety (the majority of harm occurs on high risk platforms) and proportionality (minimising the costs to smaller and lower risk businesses).
75. **Importantly, Option 2 would be implemented through primary legislation with codes of practice set out by the independent regulator.** This would ensure that the measures that service providers are required to implement would be proportional to the cost and impact of the remedial action, as well as the impact and severity of harm the measures are seeking to remedy.
76. **Option 2 is likely to achieve reductions in online harms while maintaining a proportionate and risk-based approach.** It would not place undue burdens on businesses where there are low, or no, risk of harms.

Table 2: **Option 2** - Scope, duties and implementation

Option 2 (preferred option)		
Scope	Duty of care	Implementation
<p><b>Businesses:</b> around 24,000 businesses in scope of <b>Option 2</b></p> <p><b>Harms:</b> All harms but with differentiated duties for illegal content and activity, safeguarding children and legal but harmful content and activity</p>	<p><b>All in-scope businesses</b> have duties to address illegal harms</p> <p><b>All in-scope businesses</b> have duties to safeguard children (if likely to be accessed)</p> <p><b>Additional duty on the highest risk platforms</b> (Category 1 in the legislation) to address legal but harmful content and activity accessed by adults.</p>	<p>Primary legislation with codes of practice. All codes of practice will be subject to an IA which will specifically consider the impacts on small and micro businesses.</p>

77. **This Bill will set out general definitions of illegal and harmful content and activity.** In addition, a limited and non-exhaustive list of priority categories of harmful content, posing the greatest risk to users, will be set out in secondary legislation. This will provide legal certainty for organisations and users. However businesses will still have a duty to address harmful content and activity which falls outside of these priority harms, where there is a risk of it manifesting via their service. An exhaustive list could prevent swift action to address new or uncommon types of harmful user generated content and activity.

78. **For illegal offences organisations will have a duty to put in place systems and processes to identify, minimise and remove the presence of illegal content.** Businesses must also consider whether they are likely to be accessed by children. If they are, they have a duty to put in place systems and processes to protect children from legal harms. The additional duty on Category 1 (high risk and high reach) platforms means they are likely to be required to assess risks on their platforms and update and enforce their terms of service. There is also a requirement on Category 1 platforms to publish transparency reports.

79. A detailed overview of the current policy position can be found in [Annex C](#).

## Option 3 - uniformly applied safety duties

80. **Option 3 would see the introduction of a direct and uniform regulatory regime.** Under this option, businesses would have detailed safety duties, to protect users from both illegal and legal but harmful harms, including the safeguarding of children. As this approach would be uniform, rather than risk-based, the requirements would be applied consistently across all harms and organisations in scope. The actions organisations could take to comply would be set out in primary and secondary legislation and would be enforced by an independent regulator.

81. **We estimate that the effect of this approach is that the overall reduction in online harms would be only marginally greater (the majority of harms occur on high risk businesses) but many low risk businesses would incur costs from actions they would not be expected to take under the other two options.**

Table 3: **Option 3** - Scope, duties and implementation

Option 3		
Scope	Safety duties	Implementation
<p><b>Businesses:</b> around 24,000</p> <p>Harms: All harms with uniform responsibility</p>	<p><b>All in-scope businesses</b> have duties to address illegal harms</p> <p><b>All in-scope businesses</b> have duties to safeguard children (if likely to be accessed)</p> <p><b>All in-scope businesses</b> have duties to address legal but harmful content and activity.</p>	<p>Detailed safety duties set out in primary and secondary legislation.</p>

82. **For illegal offences platforms will have a duty to put in place systems and processes to identify, minimise and remove the presence of illegal content.** Businesses must consider whether they are likely to be accessed by children. If they are, they have a duty to put in place systems and processes to protect children from harms. All in-scope businesses also have a duty to put in place systems and processes to address legal but harmful content.

## Preferred option and implementation plan

83. **Option 2 is the preferred approach as it balances the need for action to reduce the harms experienced online with a need to maintain a proportionate and risk-based approach.** Importantly, this does not place undue burdens on businesses where there is a low, or no, risk of harm.

84. **The government is working towards the legislation being introduced in Autumn 2021 and expects passage to take 10-12 months, which means Royal Assent is expected in 2022.** These timelines are dependent on parliamentary scheduling and capacity. The government estimates that it will take 18 - 24 months following Royal Assent for the full regime to enter into force. This estimate takes into account the time needed to set the regulator up and to pass the necessary secondary legislation, including codes of practice produced by the regulator.
85. **The appraisal period in this IA runs for 10 years from 2023.** The first year of the appraisal period is a transition year - the point at which businesses are expected to familiarise themselves with the regulations. While in reality familiarisation and transition costs may be incurred earlier, for appraisal purposes this IA assumes these are incurred in the first year of the appraisal period. In this IA business are expected to ensure compliance from 2024 onwards and therefore, costs related to compliance are estimated to be incurred from the second year of the appraisal period onwards.
86. **This programme is co-sponsored by DCMS and the Home Office: DCMS will be responsible for the delivery of the programme, alongside the designated regulator candidate OFCOM.** The Home Office will play a role in the governance and assurance of the programme but will not be directly involved in delivery.

## Costs and Benefits

### Main sources of evidence

87. This consultation stage IA draws on a number of evidence sources to provide an indication of the likely scale of the impact. The government expects to improve these estimates through further engagement with businesses, both as part of this IA and through the pre-legislative scrutiny process.
88. **Revealing Reality (RR) research:** In 2020, DCMS commissioned consultancy firm Revealing Reality (RR) to estimate the number of organisations in scope of the online safety framework and to determine the likely incremental costs of compliance. To determine the number of organisations in scope, RR took a stratified sample of the Inter-Departmental Business Register and manually assessed the features on each platform, the results were extrapolated up to the UK economy using BEIS' Business Population Estimates. For compliance costs, RR interviewed a range of businesses ranging from small low risk platforms to large higher risk platforms. Estimates provided by businesses are then applied to the estimates for in-scope organisations to calculate an incremental cost of compliance. More details on the methodology are included later in this IA, and the results form the basis of our current estimates.
89. **Business stakeholder survey:** In January 2021 following the publication of the full government response, DCMS sent cost surveys to a sample of 36 business stakeholders to understand in greater detail how organisations and platforms are preparing for regulation and any costs associated with the preparations. The sample consisted of 10 of the 16 largest social media platforms in the UK, review platforms, games organisations, retail sites, dating sites, and forums. DCMS received 32 responses, 15 were usable (i.e. answered at least one question), 7 included information on actions taken to prepare for regulation, and 3 provided cost information. While on the whole businesses were hesitant (or unable<sup>56</sup>) to provide cost information at this stage, the

---

<sup>56</sup> It was noted by one respondent that the full government response does not provide enough detail to estimate expected costs. This is somewhat expected given that requirements will be set through future codes of practice and subsequent secondary legislation.

survey does provide some insight into whether the businesses surveyed expect to take actions and incur costs to ensure compliance with the future regime. The transcript of questions can be found in [Annex D](#) and the results are discussed throughout to supplement current estimates.

90. **Rapid evidence assessment (REA) of NetzDG**: Despite numerous countries considering how to make the internet safer for users (see ‘international context’ section above), international policies addressing this issue are either planned and not yet implemented or have not been fully assessed. As such, comparisons between the OSB and similar international policies have been limited to NetzDG (2018). The Network Enforcement Act (NetzDG) is a German law aimed at combating hate speech online which came into effect on 1 January 2018. NetzDG has been in force for a reasonable amount of time and while there are significant differences between NetzDG and the Online Safety Bill, both address (or aim to address) online harms to some extent, and it is a useful proxy. DCMS conducted a REA with the aim of providing an overview of the impact of NetzDG in Germany specifically in relation to compliance costs faced by businesses, the impact of the law upon market innovation and whether it has reduced online harms. The main objective of this REA was to enhance the knowledge base in regards to existing online regulation, to provide an informed assessment of the impact the Online Safety Bill will have on businesses once it comes into force. An overview of the methodology can be found in [Annex E](#).
91. **Audiovisual Media Services Directive (AVMSD) research**: DCMS commissioned EY consultants to research the implementation of the AVMSD for video-sharing platforms. The Directive sets requirements on video-sharing platforms (e.g. YouTube) to protect users from harm and therefore, actions taken and costs incurred by in-scope businesses represent another reasonable proxy for the costs of the Online Safety Bill. The research looked at both UK and non-UK based platforms and included a current state assessment, an assessment of measures taken by businesses to ensure compliance (and their associated costs), and consumer research on the user experience.

## Approach

92. This section sets out the approach to estimating the costs to industry (including civil society), government and individuals of the proposed regulation. At this primary stage, it is not possible to predict with certainty the actions of OFCOM or the steps businesses may take to ensure they are compliant with the regulation. While this Bill sets out a duty of care for businesses, the specific requirements on businesses and the actions they can take to comply will be set out in codes of practice laid by OFCOM and where necessary secondary legislation. All future codes of practice will be subject to an IA, including an assessment of the impacts on small and micro businesses and innovation (this goes beyond current regulator requirements under the Small Business, Enterprise and Employment Act 2015).
93. Given that specific business requirements are unknown at this stage, impact estimates included here are largely illustrative and aim to indicate the potential scale or nature of impacts of the whole policy (scenario 2 in the RPC’s primary legislation guidance<sup>57</sup>). To do this without knowing the details of future codes of practice, this IA develops (and estimates the costs of) a plausible set of actions businesses may take, based on the policy intention, the size of the business and the risk of online harms on its services.
94. In line with requirements for impact assessments, all per business and total costs presented below are in 2019 prices and 2020 present value base year.

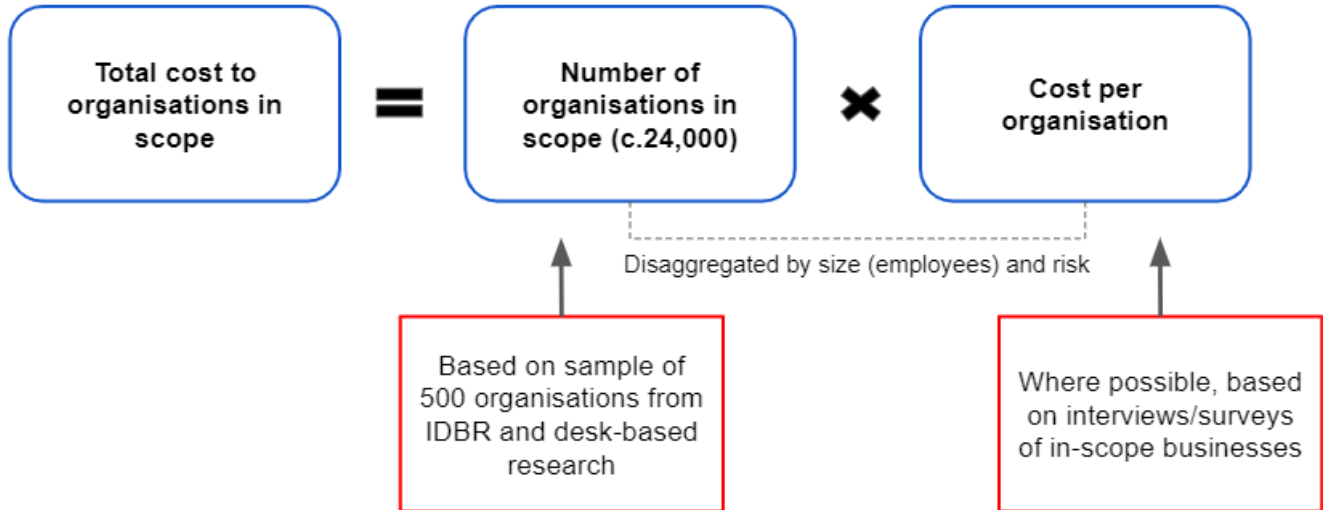
---

<sup>57</sup> [RPC case histories: assessment and scoring of primary legislation measures](#) - RPC (August 2019)

95. The approach to business cost estimation is outlined in the figure below

**Figure 4: Approach to business cost estimation.**

This flow chart demonstrates the approach taken to estimate costs to industry. The total cost to organisations in scope is equal to the number of organisations in scope (approximately 24,000) multiplied by the cost per organisation. The number of organisations in scope is based on a sample of 500 organisations from the IDBR and desk-based research. The cost per organisation where possible is based on interviews and surveys of in-scope businesses. Both of these variables have been disaggregated by size (employees) and risk.



## Summary of options

- **Option 0 - do nothing (baseline)**
- **Option 1 - limited risk-based scope:** a duty of care for user generated content and activity. The duty of care sets out organisations responsibilities in addressing illegal harms and the safeguarding of children from both illegal and legal but harmful content and activity. Under **Option 1**, duties are set out in primary legislation and codes of practice. The regime is overseen, monitored and enforced by an independent regulator, which applies a risk-based approach to its core activities.
- **Option 2 - full risk-based scope:** a duty of care for user generated content and activity. The duty of care sets out organisations responsibilities in addressing illegal harms and the safeguarding of children from both illegal and legal but harmful content and activity. In comparison to **Option 1**, this option includes an additional requirement for high risk organisations to address illegal but harmful content and activity accessed by adults. Under **Option 2**, duties are set out in primary legislation and codes of practice. The regime is overseen, monitored and enforced by an independent regulator, which applies a risk-based approach to its core activities.
- **Option 3 - uniformly applied safety duties:** Detailed safety duties setting out organisations responsibilities in addressing illegal harms and legal but harmful content, and the safeguarding of children from both illegal and legal. These safety duties are detailed in primary legislation and are uniformly applied across all harms and organisations in scope. The regime is overseen, monitored and enforced by an independent regulator.

## Option 0 – do nothing (baseline)

96. Under the *do nothing* option, the level of protection that platforms have in place to keep users safe is highly variable. **RR research** found that, in general, the mitigations an organisation had in place were proportionate to the organisation's risk of potential online harm, i.e. higher risk platforms had many more protections in place than low risk platforms. Human and automated moderation was present across all risk categories of platforms, whereas processes such as reporting functions, paying for access to databases, such as Photo DNA, and publishing transparency reports, were only present in higher risk businesses.

97. Different types of mitigation were implemented to varying degrees. For instance, while automated moderation was used throughout, the complexity and tailoring of this to the specific platform varied. For example, a low risk organisation was using 'off the shelf' automated moderation to detect spam, whereas a high risk organisation had developed their own bespoke automated software tailored to detect specific harms present on their site.

98. Most organisations were already investing in protecting their users in the absence of regulation and expected this investment would continue to increase over time. The reasons for investing in protecting users from online harm included:

- Creating a positive environment for users, to retain existing users and attract new ones
- To meet the requirements of advertisers and third-party suppliers, such as payment providers, who do not want to be associated with harmful platforms
- To remain competitive in the industry and keep up with their competitors

99. For the few who were not investing significant amounts in mitigation, some appeared to be ideologically opposed to the idea of collecting and moderating the content of their users.

100. Findings from the **RR research** on high variability in current mitigations was corroborated by EY's **AVMSD research** of video-sharing platforms. They found that the measures employed by each platform depended on the nature of the risks, the level of resources of the platform, the type of content on the platform, the impact on the platform's brand, and competitive considerations.

101. While the government does not have accurate data on the current UK-wide level of harm mitigation under the *do nothing* option, this IA - where at all possible - attempts to incorporate this in the costing of alternative options, i.e. only the incremental costs of regulation have been included.

102. As outlined in the above rationale for intervention, many online harms have been increasing in prevalence and increased screen time resulting from Covid-19 has likely exacerbated this. This IA estimates that under the *status quo* online harms result in a societal cost of at least £54 billion (PV) across the appraisal period (the calculations underpinning this estimate can be found in the benefits section below) - this is based only on a small subset of harms that this IA was able to quantify, so the actual cost is likely to be much larger .

## Costs to business

### Number of businesses in scope

103. **Under all three options, the number of affected businesses (and CSOs) within scope of the regulations is estimated to be around 24,000<sup>58</sup>.** This assessment assumes that the number of in-scope businesses grows in line with the average rate of annual growth in the business population (3% between 2000-2020)<sup>59</sup>.

---

<sup>58</sup> The exact estimate is 24,311

<sup>59</sup> [Business Population Estimates for the UK and the regions 2020](#) - BEIS



104. In order to estimate the number of organisations in scope, RR extracted a stratified sample of 500 organisations from the IDBR. The sample consisted of 100 randomly selected organisations in each of the following size categories (sole traders, micro (not including sole traders), small, medium and large<sup>60</sup>). A key advantage to this approach was to guarantee a minimum number of organisations to be tested within each size group. If a purely random sample had been taken, it would contain only a few large organisations—not nearly enough for any reasonable analysis.

105. A sample of 500 is considered to be large enough to provide robust estimates as it ensured a relatively small margin of error at the 95% confidence level (between  $\pm 2.6$  to 4.4 percentage points). Additionally, every organisation within the sample had to be manually reviewed and categorised to determine what in-scope features and mitigation practices they had in place on their website or app. This IA presents comprehensive sensitivity analysis on these estimates in later sections.

106. For each organisation in the sample, RR manually reviewed whether it offers activities within the scope of the regulation as outlined in the OHWP. The findings from the sample (e.g. percentage of in-scope businesses in each size category) were then extrapolated using BEIS' BPE<sup>61</sup> to estimate the total number of in-scope businesses in the UK. This is in line with RPC guidance on defining a business by taking a 'GDP approach', i.e. the assessment of impacts on business are in terms of the location of the economic activity being in the UK. This provided an initial estimate of approximately 18,300 businesses.

107. The OS Bill will affect CSOs as well as businesses and BEIS' BPE excludes voluntary organisations. To address this, the findings from the sample were further extrapolated using data on CSOs in the UK Civil Society Almanac.<sup>62</sup> Given that businesses and CSOs are treated the same under the Business Impact Target, this IA adds the in-scope CSOs to the number of affected businesses (approximately 550 CSOs were added). Throughout the IA, references to in-scope businesses include CSOs. This is a reasonably reliable methodology for determining the number of CSOs in scope; however, it does have limitations:

- As mentioned, the same methodology used for all businesses is applied to CSOs. This is therefore an approximation as the actual size and risk-level of CSOs will be slightly different to that of all platforms in scope.
- For all organisations in scope, 'size' was quantified in terms of number of employees (as in the SBEE Act). This is not possible for CSOs, largely because a large amount of the workforce are volunteers. Instead, and in line with standard appraisal practice in this area, CSOs are ranked by annual revenue<sup>63</sup>, meaning there will be some variation within organisation size.
- Specific actions resulting in transition costs and compliance costs are assumed to be the same for businesses and CSOs (differing only on the basis of organisation size and the risk of harms occurring on the platform). This is a reasonable approach at this stage given that requirements apply equally to CSOs and businesses. However, the government will seek further evidence through pre-legislative scrutiny and engagement with stakeholders and if evidence becomes available that CSOs will incur different costs than businesses, estimates will be revised.

108. Acknowledging the potential for gaps in the random sample (for example the lack of in-scope small businesses), additional types of organisations were identified and included in the estimates. For example, crowdfunding or fundraising sites, dating sites and forums were added to the sample on the assumption that all (or at least most) of these would fall within scope. Approximately 3,000 small businesses were added to the estimates for a total of around 21,600. It is important to note that these additions do not represent an exhaustive list of all types of

---

<sup>60</sup> The definition is in line with SBEE Act.

<sup>61</sup> [Business Population Estimates](#) - BEIS

<sup>62</sup> [UK Civil Society Almanac 2020](#) - NCVO

<sup>63</sup> CSOs are matched to the business size categories based on average revenue by business size as presented in BEIS' BPE.

organisation that could be in scope, but are an attempt to deal with some of the larger groups to provide a more realistic estimate. Estimates for the number of in-scope businesses provided by the **RR research** were based on 2019 data from the IDBR. These were uplifted by the average annual growth in the business population (3% between 2000-2020<sup>64</sup>) to account for an implementation date of 2023. For modelling purposes, this growth rate continues throughout the appraisal period.

**109. Following the above steps, the final estimate for the number of in-scope businesses is c24,000 organisations.**

Table 4: Steps to attain an estimate for the number of in-scope businesses

	Micro	Small	Medium	Large	Running total
<b>Percentage in-scope within sample</b>	0.3 % <sup>65</sup>	0 %	2 %	4 %	-
<b>Number of in-scope businesses within UK economy (nearest hundred)</b>	17,100	0	800	400	18,300
<b>Number of in-scope CSOs within UK economy<sup>66</sup> (nearest hundred)</b>	400	0	100	<100	18,900
<b>Accounting for gaps in sample with known types of businesses</b>	-	~1,000	~2,000	-	21,600
<b>Number of in-scope organisations uplifted to 2023<sup>67</sup> (nearest hundred)</b>					<b>24,300</b>

110. Given the difficulty in determining the exact number of businesses in scope, this IA conducts sensitivity analysis (the results of this can be found in the sensitivity section). The vast majority of UK organisations (around 99%) fall out of scope because:

- They have no online presence; or
- Their only internet presence is a website without the functionalities described above; or
- Where they do offer the functionalities described above, it is via a third party provider<sup>68</sup>, for example Etsy, Trustpilot, Feefo, etc; or
- They are out of scope due to one of the list of exemptions below.

111. Following the OHWP and in response to stakeholder consultation, the government incorporated a number of exemptions for specific types of services:

- **‘Low risk functionality’ exemption:** The Online Safety Bill will exempt user comments on digital content provided that they are in relation to content directly published by a platform/service. This will include reviews and comments on products and services directly delivered by a business, as well as ‘below the line comments’ on articles and blogs.
- **Services used internally by businesses:** This is defined as a service (or distinct part of a service), managed by an organisation, whose primary purpose is to host members' user-generated content and enable interactions between members within that organisation. This encompasses online services which are used internally by

<sup>64</sup> [Business Population Estimates for the UK and the regions 2020](#) - BEIS

<sup>65</sup> Weighted data combining 0 employee and 1-9 employee stratas

<sup>66</sup> Note the size of CSOs is determined by annual revenue in line with appraisal practice in this area.

<sup>67</sup> Start of the appraisal period and expected date of implementation

<sup>68</sup> In this case, the third-party provider would be in scope rather than the platform using their services.



organisations such as intranets, customer relationship management systems, enterprise cloud storage, productivity tools and enterprise conferencing software.

- **Network infrastructure:** Any service which doesn't have direct control over the User Generated Content on their platform. In practice, this takes out network infrastructure such as Internet Service Providers, Virtual Private Networks and content delivery services as they don't have any control over an individual piece of content. This also rules out business to business services e.g. white label or software as a service (SaaS) services offered to businesses where again the business to business business doesn't actually have control over specific pieces of content or activity.
- **Educational platforms:** Online services managed by educational institutions, including early years, schools, and further and higher education providers. This includes platforms used by teachers, students, parents and alumni to communicate and collaborate. This includes platforms like intranets and cloud storage systems, but also "edtech" platforms.
- **Email and telephony:** Email communication, voice-only calls and SMS/MMS remain outside the scope of legislation.

112. Furthermore, business-to-customer interactions are not considered user generated content and will also be out of scope (for example video and email interactions between a user and a business). An example of this would be a complaints box where users can interact with a business as well as patient-doctor virtual services where users can have a virtual appointment with a physician.

113. The steps described above to calculate the number of in-scope businesses was initially conducted prior to the announcement of these exemptions. This original research found that under a broad interpretation of the OHWP, around 3% of all UK businesses could be considered in scope, equating to approximately 180,000 businesses. DCMS commissioned RR to reassess the initial sample of in-scope businesses, incorporating the new exemptions. **The new analysis found that the exemptions removed approximately 160,000 businesses from the scope of the regulation resulting in a final estimate as described above of approximately 24,000 organisations.**

## Risk categorisation of organisations in scope

114. The regulatory regime under both **Option 1** and the preferred option will be risk-based and proportionate. Rather than a one-size-fits-all approach, service providers will take a variety of different actions depending on the characteristics of their services and the risk of online harms on their platforms. Organisations which offer services with the lowest risk of online harms will face the lowest regulatory burdens and businesses offering high-risk services will be required to take the most action. To reflect this in the analysis of businesses' impacts, businesses from the sample were split into three risk tiers (low, mid and high) which determined the type of likely actions they would take in complying with the regulatory framework.

115. In addition to businesses taking different actions to ensure compliance under the general proportionality principle, a small number of the largest and highest risk businesses will have additional duties, namely to take action with regard to legal but harmful content and activity accessed by adults. This very small number of the highest-risk in-scope services will be designated as Category 1 services. The designation process for Category 1 services consists of three steps.

- First, high-level criteria are set out in this Bill specifying factors which lead to significant risk of harm (namely the size of a service's audience and the presence of functionalities on a service that are likely to give rise to harms).
- Second, the Secretary of State will determine these thresholds in secondary legislation, with non-binding advice provided by OFCOM.
- Third, OFCOM will assess in-scope services against the thresholds and publish a list of Category 1 services.

116. By definition all Category 1 services will fall into the high risk tier, but not all services in the high risk tier will be Category 1 services. The thresholds are expected to be set at such a point that only the largest services will be designated as Category 1 - the current estimate based on the policy intention is that only up to 20 of the largest and highest risk services will meet the Category 1 thresholds, likely to be large social media platforms and potentially some gaming platforms and online adult services. While these businesses are likely to be large multinationals, this IA - in line with better regulation guidance - includes all impacts on businesses in terms of the location of the economic activity being in the UK, i.e. a 'GDP-approach'.

117. The categorisation of in-scope businesses in the analysis was done through a 'scoring' system where in-scope features add to the service's risk score as does an organisation's reach - this approach is in line with how the legislation's thresholds will work in practice. In addition, services targeted at or used primarily by children are assigned a higher score (this reflects the additional requirements on services 'likely to be accessed by children').

118. **RR research** has indicated that the majority of the around 24,000 in-scope organisations (over 97%) fall into the low and mid risk categories (49% and 48% respectively). Less than 3% of in-scope organisations could be considered high risk platforms and less than 0.1% are estimated to meet the Category 1 thresholds (additional requirements on the largest and highest risk businesses).

119. The table below outlines the percentage of in-scope businesses within each size category and risk tier (each figure is the percentage of all in-scope businesses, e.g. 40% of all in-scope businesses are low-risk micro businesses):

Table 5: Percentage of in-scope businesses in each size category and risk tier

	Low risk	Mid risk	High risk	Category 1
<b>Micro</b>	40%	40%	<1%	0%
<b>Small</b>	2%	2%	<1%	0%
<b>Medium</b>	5%	5%	2%	0%
<b>Large</b>	2%	<1%	<1%	<1%

## Duties and requirements on businesses

120. In order to calculate costs to business for each option, it is important to lay out how the duties will work for businesses in each risk tier.

121. As with the tiered categorisation of platforms, under **Option 1** and the preferred option, there will be a differentiated approach to harms which include:

- illegal content and activity;
- content and activity which is harmful or inappropriate for children; and
- legal but harmful content activity accessed by adults.

122. The below table outlines the responsibility of businesses under each option:

Table 6: Duties under different options

<b>Duty</b>	
<b>Option 1</b>	<ul style="list-style-type: none"> <li>• <b>All in-scope businesses</b> have duties regarding illegal harms and duties regarding the protection of children from all harms<sup>69</sup></li> </ul> <p>For illegal priority offences they have a duty to put in place systems and processes to identify, minimise and remove the presence of priority illegal content. For non-priority illegal offences, that they are made aware of by user reports or their own risk assessment, they have a duty to put in place systems and processes to identify and address - whether through removal or minimisation. Businesses must also consider whether they are likely to be accessed by children. If they are, they have a duty to assess the level of risk through a risk assessment and put in place systems and processes to protect children from harms. This duty relates not only to minimising the harmful content itself but also harmful behaviors such as grooming or bullying.</p>
<b>Option 2</b>	<ul style="list-style-type: none"> <li>• <b>As in Option 1, all in-scope businesses</b> have duties regarding illegal harms and duties regarding the protection of children from all harms.</li> </ul> <p>For illegal priority offences they have a duty to put in place systems and processes to identify, minimise and remove the presence of priority illegal content. For non-priority illegal offences, that they are made aware of by user reports or their own risk assessment, they have a duty to put in place systems and processes to identify and address - whether through removal or minimisation. Businesses must also consider whether they are likely to be accessed by children. If they are, they have a duty to assess the level of risk through a risk assessment and put in place systems and processes to protect children from harms. This duty relates not only to minimising the harmful content itself but also harmful behaviors such as grooming or bullying.</p> <ul style="list-style-type: none"> <li>• <b>Unlike Option 1, Category 1 platforms</b> have an additional duty to put in place systems and processes for legal but harmful material for adults</li> </ul> <p>Category 1 platforms are likely to be required to assess risks on their platforms and update and enforce their terms of service. There is also a requirement on Category 1 platforms to publish transparency reports.</p>
<b>Option 3</b>	<ul style="list-style-type: none"> <li>• <b>All in-scope businesses</b> have safety duties regarding illegal harms, legal but harmful harms and the safeguarding of children. The standards are set out in legislation and apply uniformly to in-scope businesses.</li> </ul> <p>For illegal priority offences they have a duty to put in place systems and processes to identify, minimise and remove the presence of priority illegal content. For non-priority illegal offences, they are made aware of by user reports or their own risk assessment, they have a duty to put in place systems and processes to identify and address - whether through removal or minimisation. Businesses must consider whether they are likely to be accessed by children. If they are, they have a duty to assess the level of risk through a risk assessment and put in place systems and processes to protect children from harms. This duty relates not only to minimising the harmful content itself but also harmful behaviors such as grooming or bullying. All in-scope businesses also have a duty to put in place systems and processes to identify and minimize legal but harmful harms accessed by adults.</p>

<sup>69</sup> When this IA discusses all in-scope platforms having to provide a higher level of protection for children - that is only insofar as assessing the likelihood of children accessing their services. They do not have to provide any protections for children (either from the child specific harms) or in other areas of the framework unless they are a service likely to be accessed by children.

## Transition costs

123. For appraisal purposes, it is assumed that legislation enters into force in 2023. The first year is assumed to be a transition year giving businesses time to prepare for compliance based on the specific details set out in codes of practice and secondary legislation. This IA assumes that businesses will incur transition costs in the first year but will not incur compliance costs until year two. The table below sets out the total transition costs across the policy options. Details on how these costs have been estimated is below.

Table 7: Options comparison - transition costs

	Option 1	Option 2	Option 3
Total transition costs	£36.3m	£36.3m	£57.1m

124. Costs to business have been assessed as either transition costs (actions businesses will take to ensure they are compliant with the regulation<sup>70</sup>) or compliance costs (ongoing costs incurred after the initial transition period to ensure businesses remain compliant throughout the appraisal period).

125. Businesses are expected to incur the following costs associated with transition:

- Reading and understanding the regulations (familiarisation);
- Ensuring a user reporting mechanism is in place; and
- Updating terms of service.

126. Transition costs will vary depending on the businesses' size and risk tier. The following table sets out which businesses are expected to incur costs for each of the actions:

Table 8: Transition actions by business size and risk

		Micro	Small	Medium	Large
Reading and understanding the regulations	Option 1	✓	✓	✓	✓
	Option 2	✓	✓	✓	✓
	Option 3	✓	✓	✓	✓
User reporting mechanism	Option 1	✓	✓	✓	✓
	Option 2	✓	✓	✓	✓
	Option 3	✓	✓	✓	✓
Updating terms of service	Option 1	✓	✓	✓	✓
	Option 2	✓	✓	✓	✓

<sup>70</sup> For ease, we have considered businesses making changes to their user reporting mechanisms and updating their terms of service as transition costs (as they normally would be) even though we estimate that these costs could be incurred throughout the appraisal period (as guidelines and codes of practice are revised).

		Micro	Small	Medium	Large
	Option 3	✓	✓	✓	✓

## Reading and understanding the regulations

### Option 2 (preferred option) - reading and understanding the regulations

127. While only in-scope businesses are required to familiarise themselves with the framework, some businesses who think they could potentially be in scope, i.e. those which offer online services with any features that could be considered in-scope<sup>71</sup> under a broad interpretation of the regulations, may have to read the legislation's explanatory notes - even if only to determine that they were out of scope. Based on analysis conducted by RR, this IA estimates that approximately 180,000 businesses could be considered potentially in scope. Of these, this IA estimates that only around 24,000 organisations are actually in-scope - or around 14%. For initial familiarisation, it is estimated that 25% of all businesses potentially in-scope (45,000) would read the regulations - this is approximately 20,000 out-of-scope organisations incurring costs of familiarisation. These platforms are likely to be on the margin where it isn't instantly clear whether they would come under the regulations, unlike for example, email service providers where it would be immediately obvious. For the initial familiarisation, one regulatory professional at an hourly wage of £20.66<sup>72</sup> is expected to read the regulations within each business. The explanatory notes are expected to be between 25,000 and 75,000 words<sup>73</sup> and would therefore take between 2-6 hours based on a reading speed of 200 words per minute<sup>74</sup>. This initial familiarisation cost is estimated to total **£4.0 million** in the first year. In the risks and sensitivity section, sensitivity analysis is conducted using an upper bound of 180,000 organisations or all potentially in-scope platforms incurring costs of initial familiarisation - the effect on the main metrics is negligible. The government will work closely with business organisations, including Tech UK and the Federation of Small Businesses to ensure that their members are quickly able to understand whether or not they are in scope.

128. Beyond the initial familiarisation, actual in-scope businesses are expected to spend more time reading the regulations and for medium and large in-scope businesses to disseminate the information throughout the business. For these (around 24,000), another member of staff in micro-businesses (rising to 2, 5, and 10 for small, medium and large businesses respectively) is expected to read the legislation's explanatory notes. Using the same methodology, this provides a further estimate of **£3.8 million** of familiarisation costs in the first year. Additionally, for medium and large in-scope businesses costs are expected to be incurred through disseminating the information across a proportion of the businesses' staff. While it is unclear what exact proportion of staff will need to be made aware of the regulations, this IA assumes that 10% of staff within in-scope medium and large businesses will spend 30 minutes familiarising themselves. This could be through a staff meeting or engaging with a summary email. Based on the average number of employees in medium and large businesses sourced from the Business Population Estimates and the number of in-scope medium and large businesses, this IA estimates an additional cost of dissemination of **£1.4 million**.

**129. Under the preferred option, the total cost to business of reading and understanding the regulations is estimated to be £9.2 million. This cost is incurred in the first year only.**

<sup>71</sup> These include posting, sharing, reacting to content, messaging, calling, commenting, tagging, discovering or seeing user generated content.

<sup>72</sup> All wages come from the [Annual Survey of Hours and Earnings](#) and are uplifted in the calculations by 22% to account for non-wage labour costs.

<sup>73</sup> Where ranges are used, the mid-point is taken for the central estimate - sensitivity analysis has been conducted on transition costs more widely in the sensitivity section.

<sup>74</sup> [Business Impact Target Appraisal Guidance](#) - BEIS

Table 9: **Option 2 (preferred option)** - Per business cost of reading and understanding the regulations (first year only)

	Low risk	Mid risk	High risk (including Category 1)
Micro	£177	£177	£177
Small	£266	£266	£266
Medium	£638	£638	£638
Large	£2,694	£2,694	£2,694

130. It should be noted that costs estimated above cover only familiarisation of the primary legislation. There will be additional costs to business incurred as a result of familiarising themselves with secondary legislation and necessary future codes of practice produced by OFCOM. At this stage, it is not possible to predict with any certainty how much material businesses will have to familiarise themselves with in order to comply and these costs are therefore, not included in the main metrics. However, by using OFCOM's Electronic Communications Code<sup>75</sup> as a proxy this IA can provide an indication of the likely scale of these impacts for one particular code.

131. OFCOM's Electronic Communications Code has three main sections and a consultation document. The main documents total 116 pages in length and the consultation document is 128 pages, totalling 244 pages. This means that businesses would have to read between 116-244 pages - on the assumption that all sections were relevant to that particular business. Based on the average number of words per page (500) and a reading speed of 200 words per minute, the time taken to read the code would range from just under 5 hours (reading only the main documents) to just over 10 hours (reading related guidance). Using the wage of a regulatory professional (as above), the per person cost of this familiarisation would be from £103 to £216. To illustrate what this could mean in the context of the Online Safety Bill, assuming one regulatory professional in each in-scope business were to undertake this familiarisation activity for a future code of practice, and all in-scope businesses would be required to familiarise themselves, the total cost could range from **£2.5 million to £5.2 million**. This calculation is solely to provide an indication of the likely scale of potential familiarisation costs associated with future codes of practice - both secondary legislation and future codes of practice will be subject to IAs.

### Option 1 - reading and understanding the regulations

132. It is unclear how or whether legislation supporting **Option 1** would be different to the preferred option in terms of resources required to familiarise. In the absence of evidence, it is assumed that the costs associated with reading and understanding the regulations are the same as under the preferred option. Given that **Option 1** does not have a requirement to address legal but harmful content accessed by adults, familiarisation costs with this option could potentially be lower however, this is uncertain.

### Option 3 - reading and understanding the regulations

<sup>75</sup> [Electronic Communications Code](#) - OFCOM

133. It is unclear how or whether legislation supporting **Option 3** would be different to the preferred option in terms of resources required to familiarise. In the absence of evidence, it is assumed that the costs associated with reading and understanding the regulations are the same as under the preferred option.

Table 10: Reading and understanding the regulations (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
Reading and understanding the regulations	£9.2m	£9.2m	£9.2m

## User reporting mechanism

### Option 2 (preferred option) - User reporting mechanism

134. Under the framework, businesses will be expected to accommodate user reporting of harms and provide an avenue for user redress (challenge of content removal). User reporting and redress mechanisms are expected to vary across businesses. For example, the smallest lowest-risk platforms may only be required to have an email address visible on their service (already a legal requirement under the Electronic Commerce Regulations 2002<sup>76</sup>) while high risk platforms may require reporting mechanisms which can handle and triage larger volumes of reporting.

135. Through interviews with a sample of in-scope businesses, **RR research** indicated that all high-risk platforms and the majority of mid risk platforms in the sample already had reporting functions and procedures for users who experienced or witnessed harms on their platforms. Many of these organisations also tailored the options in their reporting functions to represent the harms commonly reported on their sites, and to enable them to better triage reports to ensure they dealt with the high priority harms first. For instance, reports of sexual harassment or CSEA would be flagged as high priority, while ‘offensive’ content or spam would be ranked as lower priority. This was supported in DCMS’ **business stakeholder survey** where 100% of respondents already had reporting mechanisms in place and 88% had complaints handling processes (out of 8 that answered the question). In addition, this was further supported in the **AVMSD research**, with most platforms allowing users to flag content for review; however, small platforms were more likely to think that their reporting mechanisms required significant development to become effective - this highlights the need for the differentiated approach from the regulator.

136. Given that requirements will vary proportionately across risk levels and the fact that many businesses already have these mechanisms in place, businesses are not expected to have to undergo significant redesign of online services to comply with the reporting requirement principles set out in the duty of care. However, depending on the specific details set out in the user reporting code of practice, there will likely be some incremental costs to ensure compliance - this could be simply repositioning of the organisations’ email address for low risk businesses or minimally revising the triage functionality for higher risk businesses.

137. While the costs will be considered further once the code of practice has been developed, to provide an indication of the likely scale of the impacts at primary this IA assumes varying degrees of programmer time to make changes to the internal reporting mechanism:

- Low risk platforms: 1 hours of programmer time for micro businesses (rising to 2, 4 and 6 for small, medium and large businesses respectively).
- Mid risk platforms: 2 hours of programmer time for micro businesses (rising to 4, 6 and 8 for small medium and large businesses respectively)

<sup>76</sup> [The Electronic Commerce \(EC Directive\) Regulations 2002](#)

- High risk platforms: 8 hours of programmer time for micro businesses (rising to 12, 16 and 20 for small medium and large businesses respectively)

138. The calculations here assume that costs vary based on the size and risk tier of the business and these are reflected solely by varying degrees of programmer time. It may well be the case that the cost of programmer time itself varies between businesses, i.e. a micro-business may decide to outsource this activity whereas a large firm may be able to draw on internal resources - this cost differential is not reflected here. The government will seek to improve these estimates through consultation; however, given the size of this impact in comparison to the overall scale of the measure, this specific action is unlikely to have any material impact on the overall cost estimates.

139. In addition to programmer time, for each in-scope business, one hour of Chief Executive/Senior Official time is estimated for sign-off of the changes. Please note, this reflects only the implementation or revision of internal reporting mechanisms and not the moderating actions businesses will take on the basis of reports which in this IA, is considered a compliance cost. Given the uncertainty around the specific code of practice and therefore the costs businesses will incur above the baseline (what systems organisations already have in place), comprehensive sensitivity analysis is conducted on this assumption and the transition estimates as a whole.

140. Based on a programmer's hourly wage of £21.97 and a Chief Executive's hourly wage of £47.53, this IA estimates a total cost of implementing or revising user reporting mechanisms in the first year of **£2.3 million**. To reflect the possibility that organisations may need to make changes throughout the appraisal period to reflect decisions from the independent regulator, these costs are assumed to be incurred in each year but reduce by 50% from the second year.

141. **The total cost under Option 2 for ensuring a user reporting mechanism is in place and updating it throughout the appraisal period is estimated to be £12.4 million (total PV).**

Table 11: **Option 2** (preferred option) - Per business cost user reporting mechanism (first year only)

	Low risk	Mid risk	High risk (including Category 1)
Micro	£63	£85	£221
Small	£85	£131	£312
Medium	£131	£176	£402
Large	£176	£221	£493

### Option 1 - User reporting mechanism

142. The requirement for businesses to ensure that they have proportionate user reporting mechanisms is the same for **Option 1**. The government does not have evidence on whether costs associated with user reporting mechanisms would differ depending on the types of online harms an individual organisation had a duty to address. The government will engage with businesses on this point during pre-legislative scrutiny and through this IA.

### Option 3 - User reporting mechanism

143. Given that **Option 3** will set detailed safety duties, and requirements on businesses will be less differentiated, businesses are expected to incur greater costs under this approach. To reflect this,



it is estimated that all in scope businesses regardless of risk will incur the same costs as high risk platforms under the preferred option (this reflects that user reporting mechanisms for all organisations regardless of risk would have to be able to deal with larger volumes of reporting due to duties applying to legal but harmful too, including triaging functionality). The cost of ensuring appropriate user reporting mechanisms are in place under **Option 3** total **£6.2 million** in the first year, reducing by 50% from year 2 onwards.

Table 12: User reporting mechanisms (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
User reporting mechanisms	£12.4m	£12.4m	£33.2m

## Updating terms of service

### Option 2 - Updating terms of service

144. Under the preferred option all companies will be required to set terms of service for illegal content and, if relevant, protecting children. Category 1 organisations are required to set terms of service which explicitly state i) what categories of legal but harmful material they accept (and do not accept) on their service, and ii) what additional protections for journalistic and democratic content they will offer. Findings from DCMS' **business stakeholder survey** indicate that terms of service<sup>77</sup> are widespread amongst in-scope platforms (100% of 8 respondents that answered the question already had terms of service). In addition, the **AVMSD research** found that the most commonly implemented user-safety measure was 'acceptable use policies' which large and medium sized platforms in the sample<sup>78</sup> considered to be fully functional at addressing critical risks.

145. It could be argued that changes to terms of service should be considered a business as usual activity as the **AVMSD research** indicated that platforms regularly update these policies in response to their users. While most platforms will already have some form of terms of service which outline acceptable use, and these are potentially business as usual activities, all in-scope businesses are illustratively expected to incur some incremental costs in this IA associated with assessing their own terms of service and revising them to reflect the regulator's code of practice if necessary - businesses are not expected to incur significant costs above the current baseline level of activity. For Category 1 services, it should be noted that the legislation will not set what legal but harmful content is acceptable, or how journalistic and democratic content should be treated, only that these platforms set clear terms of service and enforce them.

146. Based on an assessment of 14 of the most popular online services' terms of service,<sup>79</sup> they range in length from 2,451 words to 15,260<sup>80</sup> with an average length of 5976. It is estimated that 1.5 hours will be spent initially on reading, assessing, and making the changes. For micro and small businesses, one member of staff proxied here as a senior official at a wage of £42.27 is expected to read and assess the current terms of service and make the necessary changes (30 minutes to read based on reading speeds as outlined above and 1 hour to make the changes). In addition, one hour of Chief Executive / Senior Official time for sign-off of the changes at an uplifted wage of £47.53 is included in the estimates. This IA therefore estimates a first year per business cost for micro and small businesses of £105.

<sup>77</sup> The survey asked specifically about 'terms and conditions' but these represent the same legal agreement.

<sup>78</sup> In the AVMSD research platform size was based on the number of unique users as opposed to employees; however, with the exception of two platforms, this mapped to size definitions based on employees.

<sup>79</sup> These include some of the most popular services such as Facebook Instagram, Twitter and TikTok.

<sup>80</sup> [Visualizing the Length of the Fine Print, for 14 Popular Apps](#) - visual capitalist (April 2020)

147. For medium and large businesses (including Category 1 platforms), one regulatory professional at a wage of £20.66 is expected to read and assess the current terms of service and make the necessary changes. These businesses are likely to require an additional 2 hours of legal advice (assumed to be given here by a legal professional at a wage of £39.48). In addition, one hour of Chief Executive/Senior Official time for sign-off at a wage of £47.53 is incorporated. This IA therefore estimates a first year per business cost for medium and large businesses of £153.

148. To reflect the potential need for ongoing updates, **this cost is expected to be incurred each year but reduces by 50% from the second year onwards for a total cost of £14.7 million (10-year PV).**

Table 13: **Option 2** - Per business cost of updating terms of service (first year only)

	Low risk	Mid risk	High risk
Micro	£105	£105	£105
Small	£105	£105	£105
Medium	£153	£153	£153
Large	£153	£153	£153

### Option 1 - Updating terms of service

149. It is not clear how or whether costs would differ for platforms updating their terms of service in relation to illegal harms only and platforms updating their terms of service additionally for legal but harmful content and activity. Under Option 1, if assessing and updating terms of service against illegal harms only requires less resources (in this case staff time), costs could be potentially lower under **Option 1**. However, in the absence of evidence, the costs for this activity under **Option 1** are assumed to be the same as under the preferred option. The government will seek further evidence in relation to the cost of this activity through pre-legislative scrutiny and further engagement with industry.

### Option 3 - Updating terms of service

150. As above, the costs for this activity under **Option 3** are assumed to be the same as under the preferred option. Given that all firms would be required to assess and update their terms of service in relation to all categories of harm under this option, it is reasonable to assume that per business costs could be greater than the preferred option; however, there is limited evidence to reflect this within the estimates at this stage and estimated costs are varied on business size only.

Table 14: Updating terms of service (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
Updating terms of service	£14.7m	£14.7m	£14.7m

**Consultation question 1:** Do you agree with the estimates for costs incurred in the transition period, including estimates for familiarisation, changes to user reporting mechanisms, and revising terms of service? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 2:** How (if at all) will the inclusion of user-generated fraud affect transition costs? The government welcomes any evidence you can provide to understand the impacts.

**Consultation question 3:** Are you able to identify any other costs businesses would incur during the transition period? The government welcomes any evidence you can provide.

## Compliance costs

151. Many organisations in scope will already be taking some action to reduce the risk of online harms on their services. For example, based on research carried out by ICF consultancy on behalf of the DCMS<sup>81</sup>, some organisations already carry out the following activities to varying degrees as part of their brand protection and existing legal requirements:

- Enforcing their own terms of service effectively and consistently;
- Supporting law enforcement investigations to bring criminals who break the law online to justice; and
- Regularly reviewing their efforts in tackling harm and adapting their internal processes where necessary.

152. Additional **RR research**<sup>82</sup> also found that some organisations<sup>82</sup> consider it unlikely that the regulation will result in significant incremental costs. This is because those organisations already take steps to mitigate against online harms on their services, for two main reasons. First, increasing user expectations over the safety of online communities and the services they access online means that organisations carry out both human and machine moderation. Second, businesses in scope face incentives from third parties to ensure the safety of users. This includes advertisers (who want to protect the reputations of the brands they represent), as well as payment service providers (who want to protect their own brand).

153. **RR research** also identified a small but important proportion of organisations that considered that they would incur material costs as a result of the implementation of the proposed regulation. In support of this, DCMS' **business stakeholder survey** found that only 20% of respondents had taken any actions since the publication of the White Paper, specifically to comply with the regime, and only 50% expected to incur any actions in the future to comply<sup>83</sup> (out of 10 that answered the question). Businesses may not expect to incur material costs as a result of regulation because they feel that their current user safety systems and processes are already sufficient to comply (based on their understanding of the policy position) or they expect to make the necessary changes to their platforms regardless of regulation.

154. All organisations in scope could potentially incur incremental ongoing costs as a result of the regulation. For example, these incremental costs may relate to the following activities:

- Carrying out a risk assessment;
- Undertaking additional content moderation (both proactively and responding to user reports)
- Age assurance technologies
- Transparency reporting
- Requirement to report online CSEA

---

<sup>81</sup> [Research into Online Platforms' operating models and management of online harms](#) (June 2019)

<sup>82</sup> Based on interviews conducted with 30 organisations in Jan/Feb 2020.

<sup>83</sup> An additional 10% didn't know whether they would have to take actions in the future based on the contents of the full government response.

155. The approach in this IA is to assume a set of plausible ongoing actions undertaken by businesses, based on the size of the business and its risk tier. Compliance costs are estimated to occur from the second year onwards. These actions are shown in the table below.

Table 15: Potential ongoing actions taken by businesses

		Micro	Small	Medium	Large
<b>Risk assessment</b>	<b>Option 1</b>	✓	✓	✓	✓
	<b>Option 2</b>	✓	✓	✓	✓
	<b>Option 3</b>	✓	✓	✓	✓
<b>Undertaking additional content moderation</b>	<b>Option 1</b>	High risk only	High risk only	High and mid risk only	High and mid risk (incl. Category 1)
	<b>Option 2</b>	High risk only	High risk only	High and mid risk only	High and mid risk (incl. Category 1)
	<b>Option 3</b>	✓	✓	✓	✓
<b>Age assurance technology<sup>84</sup></b>	<b>Option 1</b>	High risk only	High risk only	High risk only	High risk (incl. Category 1)
	<b>Option 2</b>	High risk only	High risk only	High risk only	High risk (incl. Category 1)
	<b>Option 3</b>	High risk only	High risk only	High risk only	High risk (incl. Category 1)
<b>Transparency reporting</b>	<b>Option 1</b>	Category 1 only	Category 1 only	Category 1 only	Category 1 only
	<b>Option 2</b>	Category 1 only	Category 1 only	Category 1 only	Category 1 only
	<b>Option 3</b>	✓	✓	✓	✓
<b>Requirement to report online CSEA</b>	<b>Option 1</b>	✓	✓	✓	✓
	<b>Option 2</b>	✓	✓	✓	✓
	<b>Option 3</b>	✓	✓	✓	✓

<sup>84</sup> We expect only a small percentage of the highest risk businesses that are likely to be accessed by children to be required to implement age verification systems.

## Producing a risk assessment

### Option 2 (preferred option) - Producing a risk assessment

156. Under the preferred option, all businesses within scope of the online safety framework will be required to produce a risk assessment. In line with this proposal's differentiated approach, businesses will be expected to assess risks corresponding to the type of content and activity a business is required to address. In practice, this means the vast majority of businesses will only be required to assess risks related to illegal content and activity, and content and activity which is harmful to children (if the service is likely to be accessed by children). Category 1 services will also be required to assess risks related to legal but harmful content and activity accessed by adults. While discussions with businesses have indicated that many (especially higher risk platforms) already conduct internal risks assessments, we expect all platforms to incur some incremental costs if only to align current practices to the requirements set out by the regulator.
157. While it is not yet clear on what level of evidence organisations will be required to provide in their assessment of the risks, this IA uses estimates from the Networks and Information Systems Regulations 2018 (NIS)<sup>85</sup> as a proxy for the cost of producing an online harms risk assessment (or revising an existing one). The government has limited evidence on the cost of producing a risk assessment in the context of online harms and considers estimates provided in NIS 2018 to be a reasonable proxy to illustrate the likely scale of the impact. Only one business interviewed in the **RR research** provided a cost for a risk assessment they already conduct, and this was between £2,500 and £3,500. In addition, one respondent to our **business stakeholder survey** provided a cost of £10,000 for a risk assessment that they currently produce but 75% of respondents already conducted these (as part of the *status quo*). Given that many organisations already produce these, we believe that this figure would overestimate the incremental cost; however, this IA conducts sensitivity analysis using this cost in the risks and sensitivities section.
158. In order to estimate the expected costs associated with producing risk assessments, the NIS assumed that reports are produced by IT professionals and that evidence and reports are reviewed and discussed by senior management and legal professionals. Estimates proxied here include 1.5 hours of time for a legal professional (at a wage of £39.48) and 2 hours for a senior manager (at a wage of £21) for micro and small businesses, rising to 5 and 7 for medium sized businesses and 10 and 14 for large businesses respectively.
159. Due to the differentiated requirements under the preferred option, it is likely that these costs will vary across risk categories with reduced costs for low risk platforms and increased costs for high risk platforms. In the absence of evidence, this IA does not vary costs across risk categories but does conduct sensitivity analysis around the estimates.
160. The cost of producing risks assessments are therefore estimated to be **£3.6 million** in the first year of compliance and **£31.0 million** in present value terms across the appraisal period.
161. It should be noted that in the absence of evidence this IA assumes that the cost of producing a risk assessment will be the same for all organisations (differentiated only by platforms' size). In reality, costs may be greater for Category 1 businesses (those which have duties relating to legal but harmful content activity accessed by adults) and a proportion of Category 2 businesses (those that are likely to be accessed by children and have to assess the nature and level of risk of their service specifically for children). It is unclear at this stage the proportion of Category 2 businesses likely to meet this criteria and the potential increase in risk assessment costs in assessing risk relating to legal but harmful content and activity. However, these costs would likely be captured within the upper bound of the sensitivity analysis. Additionally, this Bill gives the Secretary of State powers to set out the priority categories of legal but harmful material, including

<sup>85</sup> [The Network and Information Systems Regulation 2018](#) - DCMS (April 2018)

those impacting children, but they will be set out in secondary legislation. These estimates, while only illustrative at this stage provide an indication of the likely scale of impacts from this activity.

Table 16: Per business cost of carrying out a risk assessment (first year of compliance)

	Low risk	Mid risk	High risk
Micro	£104	£104	£104
Small	£104	£104	£104
Medium	£355	£355	£355
Large	£709	£709	£709

### Option 1 - Producing a risk assessment

162. The requirement to produce a risk assessment is the same as under the preferred option and costs for this activity under **Option 1** are expected to be the same. It is possible, given that no businesses will be expected to assess the risks of legal but harmful content accessed by adults, that costs could be lower than estimated under the preferred option; however, these differences are expected to be minimal.

### Option 3 - Producing a risk assessment

163. The requirement to produce a risk assessment is the same as under the preferred option and costs for this activity under **Option 3** are expected to be the same.

Table 17: Producing a risk assessment (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
Risk assessment	£31.0m	£31.0m	£31.0m

**Consultation question 4:** Do you agree with this assessment of the incremental cost of producing a risk assessment? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 5:** Would the cost of producing a risk assessment differ for Category 1 services (those expected to address legal but harmful content accessed by adults)? The government welcomes any evidence you can provide to refine the estimates.

## Undertaking additional content moderation

### Option 2 (preferred option) - Undertaking additional content moderation

164. Under the preferred option, the duty of care requires all in-scope organisations to:

- put in place systems and processes to identify, minimise and remove priority illegal content<sup>86</sup>
- put in place systems and processes to identify and minimise non-priority illegal content that they are made aware of from user reporting mechanisms; and
- put in place systems and processes to protect children from harms - both illegal and legal but harmful - only if the platform is likely to be accessed by children.

165. Under the preferred option, the duty of care requires only Category 1 organisations to:

- put in place systems and processes to minimise legal but harmful content likely to be accessed by adults as well as children (through updating and enforcing their own terms of service).
- In order to protect freedom of expression and privacy, these companies will also need to: a) undertake and publish an assessment on the impact of safety duties on users' rights to privacy and free speech; and b) put in place clear policies for protecting democratic content and journalism, and ensure these are enforced consistently and transparently.
- Note that b) builds on the existing policies that many platforms already have in place. For example, Facebook has a newsworthiness exemption which means that if someone makes a statement or shares a post which breaks their community standards, Facebook will allow it on their platform if they believe the "public interest" in seeing it outweighs the risk of harm. YouTube makes exceptions for content that is educational, documentary, scientific or artistic. These policies are often based on First Amendment principles that seek to protect freedom of speech and the press.

166. Evidence from DCMS' **REA of NetzDG** highlights that platforms cannot rely solely upon technology for content moderation. This has led to platforms such as Facebook hiring more human content moderators in order to comply with NetzDG, employing 125 additional people. Therefore it is expected that undertaking additional content moderation (through hiring additional content moderators or using automated moderation) will represent the largest compliance cost faced by in-scope businesses.

167. Many organisations in scope will already be taking some action to reduce the risk of online harms on their services. **RR research** found that some organisations consider it unlikely that the regulation will result in significant incremental costs. This is because those organisations already take steps to mitigate against online harms on their services, for a few main reasons. First, increasing user expectations over the safety of online communities and the services they access online means that organisations carry out both human and machine moderation to create a positive environment for users. Second, to meet the requirements of advertisers and third-party suppliers, such as payment providers who do not want to be associated with harmful platforms. Finally, to remain competitive in the industry.

168. As part of the **RR research**, in-scope businesses<sup>87</sup> were interviewed to determine: their current practices and processes to mitigate the risks of online harms occurring; where available, quantification of the associated resources and costs of practices and processes to identify and prevent harm; and how these costs and resources would change if a duty of care was enforced.

169. A strategic sample was used to select organisations to interview in this phase. Unlike the analysis conducted to determine the percentage of businesses in scope, this is not a representative sample of organisations in the UK but instead a strategic sample of organisations who may be affected by regulation, and from whom there is most to learn in interviews.

170. It was decided to skew interviews towards those who might be expected to do more under proposed regulation, as these organisations were more likely to have to make changes.

---

<sup>86</sup> The Bill gives ministers the power to set priority categories of offences in secondary legislation.

<sup>87</sup> Under the policy position as set out in the White Paper and not the subsequent exemptions.

Therefore, there were lower interview targets for low risk organisations and micro and small mid risk organisations, given that these types of organisations are less likely to have to change their business practices in any significant way.

171. A total of 118 organisations were contacted for interview, and 25% (or 30 organisations) agreed to and completed an interview. This sample included: social media (13 of the 16 most used social media sites in the UK); forums; review sites; blogs; gaming; retail; P2P marketplaces; volunteering; official fan sites; job searching; fan fiction; search engines; accommodation searching; adult entertainment; and dating sites. It should be noted that RR conducted this sampling and the interviews prior to the exemptions being determined and therefore, some of the organisations interviewed would no longer be within scope of the regulations. We consider insights gained from organisations no longer within scope to still be useful in understanding the potential incremental costs of regulation.
172. There are limitations as to how specific organisations can be about whether they will incur any additional costs, or what these costs might be, without knowing exactly what the regulation will require them to do (this will be set out by an independent regulator in future codes of practice and subject to an IA). Therefore, any estimates of additional cost are based on a reasonable interpretation of the policy as laid out in the OHWP, and current practice within the organisations.
173. To estimate the incremental cost of compliance, the analysis discounts organisations that already have sufficient content-moderating systems and processes in place and organisations that - due to being very low risk or smaller mid risk platforms - would likely not be expected to take additional actions in moderating content under the preferred option.
174. Based on findings from the interviews, the percentage of in-scope businesses requiring extra spend on content moderation is estimated to be 25% of high risk in-scope organisations and 10% of medium and large mid risk organisations. These estimates are highly uncertain given that it was not possible to interview a representative sample of in-scope businesses given the scope of the regulations. This IA conducts sensitivity analysis on these estimates and the government will seek to strengthen them through further engagement with businesses. The attempt to isolate incremental costs of additional moderation is supported by DCMS' **business stakeholder survey** in which 100% of respondents already employed human content moderators and 88% used automated content moderation (out of 8 respondents that answered the question).
175. Among interviewed organisations in the **RR research** that expected to require additional moderation, estimates for the incremental cost of regulation ranged from 1% of turnover<sup>88</sup> (the lowest estimate) to 15%<sup>89</sup> (the highest). These estimates were provided in the context of businesses' interpretation of the OHWP, i.e. the cost of additional content moderation for businesses required to address all harms in the OHWP including extra protections for children. This IA therefore takes the midpoint of this range (7.5% of turnover) to represent the cost of additional content moderation for Category 1 organisations (those expected to address all harms). Turnover estimates used come from average turnover by business size band in BEIS' Business Population Estimates.
176. For Category 2 - those not required to address legal but harmful content accessed by adults - costs are expected to be lower than those incurred by Category 1 businesses. To calculate the cost to these organisations, data from a number of large social media businesses' transparency reports on the volume of actioned content (content which was removed or minimised due to breaking the businesses' terms of service) are used as a proxy. In the transparency reports, actioned content is split into a number of broad harm categories which were assessed as either:

---

<sup>88</sup> The lowest estimate was actually 1% of operating costs which would likely be lower than 1% of turnover; however, for ease and given data availability, we proxy with turnover.

<sup>89</sup> The exact figure given in the interview was 14% of revenue which was rounded and due to data availability, turnover was used as a proxy.



- not applicable (categories such as ‘spam’ or ‘fake accounts’ which on the whole could not be considered an online harm),
- likely to be considered illegal, or
- likely to be considered legal but harmful.

177. Using the volume of actioned content in each category, an approximate percentage split of illegal vs legal but harmful actioned content was estimated. The proportion of actioned content likely to be considered illegal is expected to be a reasonable proxy for the proportion of moderating costs incurred by businesses not required to address legal but harmful content accessed by adults. However, this approach has the following limitations:

- It assumes that the cost of content moderation is linearly correlated with the volume of harmful content. For organisations that use automated moderation this may not be the case.
- It is difficult to determine whether content actioned under the broad categories in the transparency reports would be considered illegal or legal but harmful - the reports do not break the data down in this way. For example, Twitter uses a ‘hateful conduct’ category which - referring to Twitter’s policy on the topic - is likely to contain both illegal and harmful content. A judgement was made on each category one way or the other based on the text within the terms of service; however, there is likely to be within-category variation not reflected in this assessment.
- Four social media businesses’ reports were assessed<sup>90</sup> and there were various missing categories of harm across years and quarters which meant it wasn’t possible to compare the same quarter or full year across the four businesses. Instead, a year or quarter with the fewest gaps for each business was selected to provide a snapshot of the proportion of actioned content likely to be considered illegal on the platform - the assessment was always on 2019 data.
- It is not clear that the four social media businesses’ transparency reports are representative of the wider sample of in-scope organisations and are likely to be designated as Category 1 organisations. It may be the case that legal but harmful content represents a smaller proportion of overall harmful content on Category 2 platforms or vice versa.

178. The table below presents the results of this analysis:

Table 18: Proportion of actioned content illegal vs legal but harmful

<b>Platform</b>	<b>Time period</b>	<b>% of actioned content which we have categorised as likely illegal</b>	<b>% of actioned content which we have categorised as likely legal but harmful</b>
<b>Facebook</b>	Q3&4 2019	33%	67%
<b>Instagram</b>	Q4 2019	22%	78%
<b>Twitter</b>	Jul-Dec 2019	20%	80%
<b>Snapchat</b>	Jul-Dec 2019	15%	85%

179. The percentage of actioned content in categories assessed as being likely illegal ranged from 15% to 33%. To reflect the costs to Category 2 businesses (those which are not required to address harmful content accessed by adults), given the ranges above, this IA estimates that the

<sup>90</sup> Facebook, Instagram, Twitter, and Snapchat

relative costs to these businesses would be approximately 25%<sup>91</sup> of the costs to Category 1 businesses or 1.9% of turnover.

180. The two estimates for the incremental cost of content moderation (7.5% of revenue for Category 1 organisations and 1.9% for all other in-scope organisations) differentiates between organisations expected to address only illegal harms and those expected to address all harms including legal but harmful content and activity accessed by children and adults. There will be a proportion of Category 2 businesses expected to address legal but harmful content accessed by children in addition to illegal content if the service is likely to be accessed by children. While some of these organisations could have higher levels of protection for children already<sup>92</sup>, especially those designed specifically for children, those that do need to conduct additional moderation are likely to incur higher incremental costs than organisations only expected to address illegal content and activity, i.e. higher than 1.9% of revenue. The government does not have evidence on the proportion of Category 2 businesses likely to be accessed by children or the potential increase in costs for these organisations and this is therefore not directly captured in the illustrative central estimate. However, a platform likely to be accessed by children is automatically moved up a risk tier in **RR's research**, and sensitivity analysis is conducted on our assumptions for Category 2 content moderation costs in the risks and sensitivities section.

181. Additionally, the Age Appropriate Design Code has set out required standards for safeguarding children's personal data for services likely to be accessed by children, including ensuring that the best interests of the child are a primary consideration when designing and developing online services likely to be accessed by a child, that services uphold their published terms, policies and community standards (including but not limited to privacy policies, age restriction, behaviour rules and content policies), and that age assurance technology is used where appropriate. While provisions in the Age Appropriate Design Code relate to the use of personal data, many of the actions businesses will take to comply will be applicable in the context of content moderation under the OS Bill, and may reduce the overall cost to services likely to be accessed by children, required to deliver a higher level of protection for children.

182. The table below outlines the estimated per business cost of additional content moderation. The table displays the per business cost for those assumed to require additional moderation.

Table 19: Annual per business cost of undertaking additional content moderation (estimated number of businesses in brackets)

	Low risk	Mid risk	High risk	Category 1
Micro	n/a	n/a	£2,540 (14)	n/a
Small	n/a	n/a	£45,058 (14)	n/a
Medium	n/a	£255,662 (118)	£255,662 (118)	n/a
Large	n/a	£3.3m (12)	£3.3m (9)	£13.4m (5)

**183. Under the preferred option, the cost to business of additional content moderation is expected to be £1,700.2 million over the ten year appraisal period.**

184. Three businesses provided cost information as part of DCMS' **business stakeholder survey**, relating to the annual cost of user safety measures they currently undertake (i.e. not as a result of regulation). Costs provided came from the top two size categories (medium or large) and top two risk categories (mid or high) and were all below £1m per year. Estimates provided by platforms in

<sup>91</sup> The midpoint of the range is 24% and we rounded to 25% for ease.

<sup>92</sup> In the **business stakeholder survey**, 75% (3 out of 4 respondents) of those self-declared as likely to be accessed by children already employed age verification systems (compared to 38% for respondents as a whole). In addition, platforms in the **AVMSD research** reported having more effective measures already in place if they consider themselves as likely to be accessed by children.

the **AVMSD research** varied widely from hundreds of pounds for the smallest platforms to £1.5 billion for the largest video-sharing platform. With the exception of a handful of the largest and highest risk businesses, for those expected to undertake additional content moderation, the per business costs presented above would represent a doubling or more of current content moderation costs which is likely to be significantly conservative and potentially an overestimate. In addition, it is difficult to isolate the direct cost of the regulation given the general shift towards user-safety in the sector - for example, platforms within the **AVMSD research** highlighted that legislation is only one factor in their user-safety investments. The government hopes to strengthen these estimates through further consultation with business.

185. Given that the details of future codes of practice are unknown at this stage, the estimates presented here should be considered as providing an indication of the likely scale of the impact. In addition, this IA presents comprehensive sensitivity analysis around the main assumptions used, including here with the number of in-scope businesses and estimate for potential incremental spend (the two key assumptions for this cost).

186. The cost estimates presented here represent a general cost of content moderation and given that they were provided by a range of businesses varying in size, they are likely to be a mixture of both human and automated content moderation. As automated moderation improves, businesses are expected to rely less human content moderation and as a result, expect the ongoing cost of moderation to decrease<sup>93</sup>. Significant advances have already been made: for example, Facebook reports that the share of hateful content removed by AI systems “before users report it” rose from just 24% in late 2017 to 80% by 2019<sup>94</sup>. It is unclear at present exactly how advances in automated content moderation will reduce the cost of compliance and therefore, this is not incorporated in this assessment. The costs seen here are likely to be conservative as they are based on a continuation of the current levels of content moderation and do not reflect the potential increase in the use of automated systems.

187. During the Coronavirus pandemic there has been a rapid increase in the need to rely upon AI for content moderation due to human moderators being unable to come to work. However, the use of AI moderation during the pandemic has not always been particularly successful. Numerous reports suggest that there has been an increase in harmful content such as hate speech, resulting from the lack of human moderation.<sup>95</sup> During the pandemic YouTube has had to bring back human content moderators who had previously been ‘put offline’ after their AI systems had failed to match their accuracy.<sup>96</sup> This would imply that we cannot expect moderation to rely solely on AI in the near future and there will continue to be a role for human content moderation. Therefore, although human content moderation is expected to be significantly supplemented by automated moderation, it is unlikely that human content moderation will disappear entirely.

### Option 1 - Undertaking additional content moderation

188. Unlike the preferred option, under **Option 1**, no organisations will be required to address legal but harmful content and activity accessed by adults. Given that the scope of harms which the highest risk platforms are expected to address is smaller under **Option 1**, costs for these businesses are expected to be reduced in line with estimates for Category 2 businesses under the preferred option. Therefore, this IA estimates that all organisations expected to undertake additional content moderation (10% of larger mid risk platforms and 25% of high risk platforms - as in **Option 2**) will incur costs equivalent to 1.9% of turnover. This equates to a cost of **£149.1 million** in the first year of compliance and to total **£1,271.4 million** across the appraisal period (PV).

<sup>93</sup> While automated content moderation systems in some cases may require substantial upfront investment the ongoing costs are likely to be much lower than human content moderators.

<sup>94</sup> [Why content moderators should be key workers](#) - Turing Institute (April 2020)

<sup>95</sup> [What happened when humans stopped managing social media content](#) - Politico (October 2020)

<sup>96</sup> [YouTube brings back more human moderators after AI systems over-censor](#) - The Verge (September 2020)

### Option 3 - Undertaking additional content moderation

189. Under **Option 3**, detailed safety duties would be set out in legislation and they would apply uniformly to all in-scope businesses regardless of risk. In practice, this would likely mean that in addition to estimates under the preferred option, some low risk and smaller mid risk businesses would incur incremental costs of additional content moderation. Like larger mid risk platforms under the preferred option, an additional 10% of low risk businesses and smaller mid risk businesses are expected to incur incremental costs under this option. While the requirements under **Option 3** would apply uniformly, actual steps taken to comply would vary across risk categories given the lower prevalence of harms on low risk businesses. For example, the costs of current moderation activity among low risk and smaller mid risk organisations interviewed in the **RR research** were generally low, e.g. 'negligible' or '£700 per year'. Given the lower estimates provided by businesses in the lower risk categories and the lower prevalence of harms on these platforms, the lower bound or 1% of turnover is used - this is likely to be an overestimate.

190. Based on the methodology outlined above, costs in the first year of compliance under **Option 3** are expected to be **£703.0 million** and to total **£5,999.3 million** across the appraisal period (10 year PV).

Table 20: Undertaking additional content moderation (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
Additional content moderation	£1,271.4m	£1,700.2m	£5,999.3m

**Consultation question 6:** Do you agree with this assessment of the proportion of platforms that will require additional content moderation to ensure compliance? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 7:** Do you agree with the estimates for the incremental cost of additional content moderation? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 8:** How would the cost of additional content moderation differ for platforms required to address legal but harmful content? The government welcomes any evidence you can provide.

## Age assurance technology

### Option 2 (preferred option) - Age assurance technology

191. Whilst the duties to tackle illegal material would be on all in-scope platforms, and the duty to tackle legal but harmful material for adults would be on the highest risk and highest reach businesses (Category 1), there would be additional duties to protect children from legal but harmful material on platforms which are "likely to be accessed by children".

192. This approach of providing a higher level of protection for children has been established by the Information Commissioner's Office's 'Age Appropriate Design Code' with regards to protecting children's data. Using the same principle of a higher level of protection for children provides

consistency across digital regulation, reducing additional burdens on businesses, many of whom will already have taken steps to comply with the Code. It is also less prescriptive than setting a specific number or threshold of children accessing a service. The "likely to be accessed by children" test for the Age Appropriate Design Code is established in primary legislation (in section 123(1) of the Data Protection Act 2018: "The Commissioner must prepare a code of practice which contains such guidance as the Commissioner considers appropriate on standards of age-appropriate design of relevant information society services which are likely to be accessed by children."

193. Under the future online safety regime, a business likely to be accessed by children will be required to: assess the nature and level of risk of their service specifically for children; take reasonably practicable steps to improve safety and reduce risk of harm by identifying and mitigating known or foreseeable harms to children arising from content or activity on their services; and monitor these for effectiveness. Therefore, whilst legislation is technology neutral, a small number of high risk services which are likely to be accessed by children will be required to know the age of their users and therefore may choose to implement age assurance technologies<sup>97</sup> to do this. Age assurance technologies are important tools that help companies take steps to protect children from online harms, including both legal but harmful and illegal content and activity, e.g. protecting children from grooming.

194. Broadly, the concept of 'age assurance' is an umbrella term for methods and technologies that determine the age or age range of an online user. The Age Verification Providers Association (AVPA) describes 'age verification' as a subset of 'age assurance' providing a higher level of confidence of the age or age range of a user than is typically possible with other forms of age assurance<sup>98</sup>. For example, age verification currently relies on officially provided databases or 'hard identifiers', such as a passport, driving license or credit card. Age assurance technologies are relatively new and it is a fast developing space. Types of solutions, their accuracy and their availability are rapidly evolving. With this in mind, estimating the impact of regulation on their use and the cost for businesses is challenging and is likely to change over time.

195. It is unclear what percentage of businesses would be required to adopt age assurance measures or what kind of systems they would employ. However, of those self-declared as likely to be accessed by children in DCMS' **business stakeholder survey**, 75% said that they already employed age verification (3 out of 4). It should be noted that for this question the term 'age verification' was not defined and therefore it is possible that platforms selected 'age verification' when in reality they currently employ weaker forms of age assurance. While most businesses designated as Category 1 services are expected to already employ some type of process to attempt to determine the age or age range of their users, this could range from robust age verification controls to a simple self-declaration (which on its own would not be considered age assurance). In addition, the **AVMSD research** highlighted that coverage and perceived effectiveness of current age assurance measures among small and medium sized platforms was lower than larger platforms.

196. Given the uncertainties and the rapid development of solutions, it has not been possible at this stage to estimate the total cost to business of this potential requirement or provide a reasonably accurate estimate for the potential scale of the impact. This is because:

- The proportion of businesses required to employ age assurance controls and the type of controls required are unknown at this stage, this will be set out in future codes of practice.
- Different platforms will take different actions, for example, some larger platforms may develop in-house solutions while smaller platforms are likely to employ off-the-shelf solutions which are likely to be more costly over the long term.

---

<sup>97</sup> It should be noted that the codes of practice are unknown; however, at primary stage, it is reasonable to believe that some platforms may be required to introduce age assurance systems which could include age verification (if they do not operate them already) under this part of the duty of care.

<sup>98</sup> [Age Verification Providers Association](#)

- There are many providers of off-the-shelf age assurance technology (including age verification) in the UK and these providers use a range of techniques to determine the age or age range of users with differing levels of accuracy and varying price structures. A DCMS report<sup>99</sup> published in May 2020 identified 70 organisations registered within the UK dedicated to providing relevant Safety Tech products and services<sup>100</sup> and the government provides a directory of UK Safety Tech providers<sup>101</sup>, last updated in December 2020.

197. While it is difficult at this stage to provide an indication of the likely scale of impact, this IA can present the pricing structures of a number of well-known providers of off-the-shelf age assurance technology. These specific providers were chosen for no other reason than they are well-known in the sector and include pricing structures on their website. The methods and accuracy are likely to vary greatly between providers and therefore prices shown are not comparable:

Table 21: Example pricing structures for off-the-shelf age verification

Provider	Method	Price structure <sup>102</sup>
Yoti	Age estimation powered by AI: this system scans a customer's face to provide an estimate of the customer's age. Yoti estimates that the current accuracy rate is 2.35 years. Yoti digital ID: this includes live detection (capturing an image of a real person on camera), document check (scans and reads identification documentation, with involvement from a security team), biometric face matching (matching the initial live detection to the identification documents).  Age verification platform: A service which provides businesses with access to Yoti's age estimation and digital ID services.	Age estimation powered by AI = £0.25 per verification  Yoti digital ID = £0.25 per verification  Age verification platform = £90 per terminal per year
agechecked	Agechecked offers a range of solutions from light touch age assurance to full identity and ID scans.	Monthly service plans range from £0.28 to £0.35 per check. Pay as you go options are also available.
verifymyage	Verifymyage uses a range of methods, including database checks (third-party databases), mobile phone checks (verifying the phone is authorised for use by someone over 18), AI-powered age estimation, scanning official identification documents, e.g. passport or driving licence, credit card checks (verifying age using credit card records).	Per verification price varies according to the application or platform used, but for an ebay compatible application it is £0.45 per successful verification (unsuccessful

<sup>99</sup> [Safer technology, safer users: The UK as a world-leader in Safety Tech - DCMS \(2020\)](#)

<sup>100</sup> There was also an additional 42 organisations identified that were diversified (i.e. Safety Tech was part of what they provide) or are noncommercial in scope.

<sup>101</sup> [Directory of UK Safety Tech Providers - DCMS & DIT \(2020\)](#)

<sup>102</sup> Prices are accurate as of April 2021.



Provider	Method	Price structure <sup>102</sup>
		verifications are not charged).

198. The government will seek to improve these estimates for the final stage impact assessment.

Table 22: Cost of potential implementation of age assurance technology

	Option 1	Option 2	Option 3
Age assurance technology	Not monetised	Not monetised	Not monetised

## Transparency reporting

### Option 2 (*preferred option*) - Transparency reporting

199. The preferred option does not require all platforms to provide transparency reports, as this would be too onerous for small, low risk businesses. However, the highest risk and highest reach businesses (Category 1) will be required to produce these. This is in line with the **AVMSD research** which indicated that transparency reporting can be particularly challenging for some platforms who may not routinely collect this type of data. In line with the wider requirement placed on the regulator to act in a proportionate and risk-based manner, transparency reporting requirements will differ between the different types of businesses who are required to report. The specific information that these businesses will need to include, will be left to the regulator and will differ between businesses. Responses to DCMS' **business stakeholder survey** indicated that most of the large high risk business respondents (75%, or 3 out of 4) already produce transparency reports in some form (through NetzDG requirements for example).

200. To indicate the likely scale of the cost of this activity, this IA uses estimated costs from the transparency reporting requirements under Germany's NetzDG which were expected to be 50,000 EURO (approximately £45,000)<sup>103</sup>. Under NetzDG, social media businesses with more than 2 million registered users in Germany are obligated to report quarterly in German on their efforts to tackle illegal harms, including complaints and performance data. Estimates provided for NetzDG are a reasonable proxy for the transparency reporting requirements under the online safety framework.

201. Under the preferred option, only Category 1 businesses will be required to produce a transparency report and these will likely be required annually. The cost of this activity is likely to be front-loaded, especially for businesses without appropriate systems already in place - to reflect this, the cost of transparency reports is expected to reduce by 50% from year 2 onwards<sup>104</sup>. This will result in a cost of **£0.8 million** in the first year of compliance and a total cost of **£3.6 million** across the appraisal period (10 year PV).

202. The OS Bill does provide the Secretary of State with a power to expand the scope of transparency reporting beyond Category 1 businesses if necessary, without affecting the differentiated duties. However, when this power will be used and by how much the scope will be increased is unknown at this stage and will depend entirely on the situation at the time of consideration. The potential transparency reporting cost therefore, of an unknown proportion of Category 2 businesses is not included in the central estimate. However, this IA provides an

<sup>103</sup> [Act improving law enforcement on social networks \[Netzdurchführungsgesetz – NetzDG\] - European Commission \(2017\)](#)

<sup>104</sup> If the information required from businesses under the reporting requirements is changed frequently throughout the appraisal period, it is possible that costs could increase back to year 1 estimates.

indication of the likely scale of impacts stemming from this power. Given NetzDG applied only to large social media businesses, the original cost estimates are revised down for the other size bands. The government does not have online harms specific evidence of how costs for this activity would differ between business sizes and therefore, this IA uses differentiated compliance cost estimates from the Network Information Security Directive 2018 - which include reporting requirements - as a proxy. Compared to large businesses, the cost of this reporting activity was estimated to be 85.4% lower for small and micro businesses and 49.9% lower for medium sized businesses. Applying this scale to the estimates for Category 1 organisations (£45,000), it is estimated that transparency reporting requirements, if extended beyond Category 1 could result in per businesses costs of £45,000 for Category 2 large businesses, £22,454 for medium sized businesses and £6,570 for small and micro businesses.

### Option 1 - Transparency reporting

203. The requirements on transparency reporting are the same for **Option 1** as they are under the preferred option. In reality, given that organisations in **Option 1** would only be reporting on illegal content, the cost per report is likely to be lower than under the preferred option. However, the government does not have any evidence on the incremental costs of reporting legal but harmful content (in addition to reporting on illegal content) and therefore this IA assumes that the costs remain the same for **Option 1, £0.8 million** in the first year of compliance and a total cost of **£3.6 million** across the appraisal period (10 year PV).

### Option 3 - Transparency reporting

204. Under **Option 3** - a uniformly applied set of safety duties - all in-scope businesses would be required to produce transparency reports which would significantly increase the cost of this activity. As noted in the previous section, NetzDG applied only to large social media businesses. Using the estimates above, for **Option 3**, large businesses are expected to incur costs of £45,000, medium businesses £22,454, and small businesses and micro businesses £6,570. Using this per report costs as above but applied to all in-scope businesses rather than just Category 1 organisations, the cost of transparency reporting under **Option 3** is expected to be **£193.9 million** in the first year of compliance and total **£921.0 million** across the appraisal period (10 year PV).

Table 23: Transparency reporting (direct cost to business, monetised 10-year PV)

	Option 1	Option 2	Option 3
Producing transparency reports	£3.6m	£3.6m	£921.0m

**Consultation question 9:** Do you agree with the estimates for the cost of transparency reporting? The government welcomes any evidence you can provide to refine the estimates.

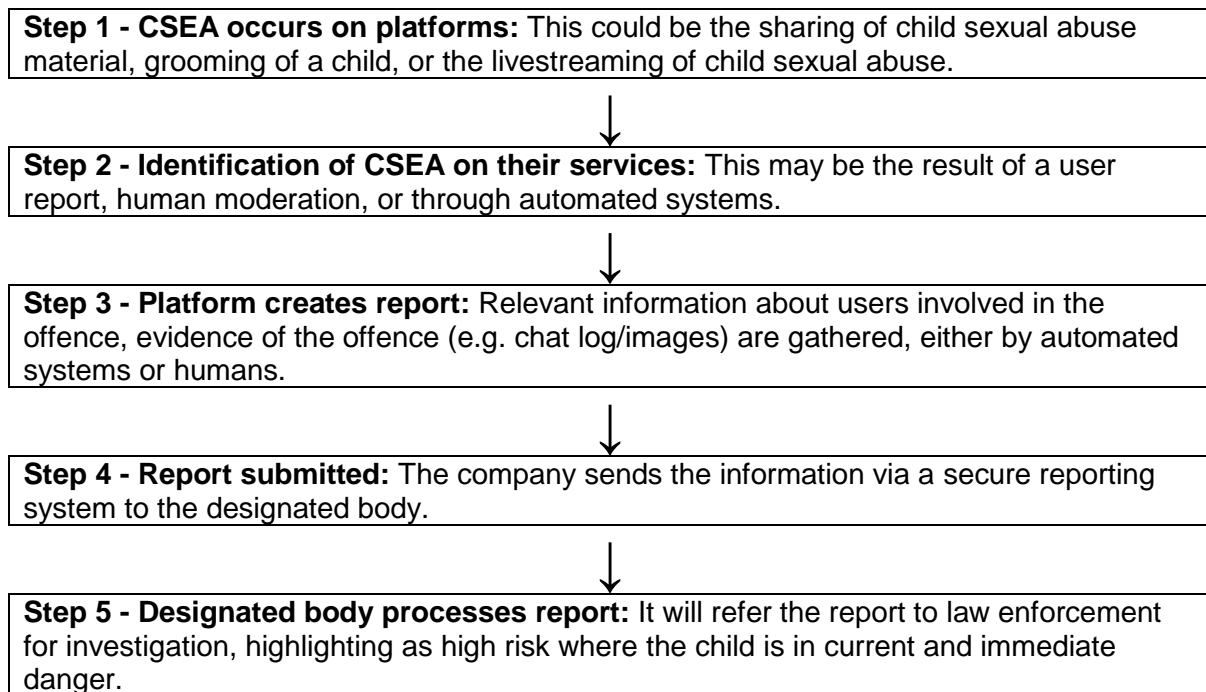
## Requirement to report online CSEA

205. As part of the OS Bill, the government plans to introduce a legal requirement on technology businesses to report online child sexual exploitation and abuse (CSEA). Introducing a CSEA reporting requirement on UK (and some non-UK) businesses will ensure that the businesses in scope are meeting best practice, which will help protect their users and provide law enforcement with the information they need to identify as many offenders and victims as possible.



206. This requirement will apply differently to businesses depending on where they are based, which is different from the approach being taken to the Online Safety regime generally, where duties will apply to all in-scope services that have UK users. UK businesses (those that provide services from within the UK) will be required to report all identified CSEA (all CSEA offences set out in the OSB) to a new designated body. Businesses providing services from outside of the UK will only have to report identified CSEA offences that are perpetrated by a UK user, and only if they do not already report CSEA. These services will be able to decide whether to report to the UK designated body or an equivalent entity or law enforcement agency in the country where they are based. This will ensure that businesses do not have to replicate their reporting efforts. For UK businesses, this will replace the current voluntary reporting regime within the UK. The below figure outlines the process:

**Figure 5: Mandatory reporting process**



### Option 2 (preferred option) - Requirement to report online CSEA

207. This requirement will have cost implications for the businesses that make these reports, and for the government through the establishment of a designated body to receive and process these reports. While this requirement will apply to UK and some non-UK businesses, this IA is only required to consider the cost to UK businesses. This section sets out the estimated costs to UK businesses and to the government.

208. As the UK's current reporting system is voluntary and reports are made to local police forces, it is difficult to provide accurate figures on the number of reports that are currently made, and how many additional reports would be made under a mandatory regime. However, figures from some businesses provide indications of the number of reports that may be made:

- **BT** – one of the largest UK businesses likely to report under this requirement (in terms of users and employee numbers), but they provide relatively low risk service types so are not necessarily representative of other large businesses with higher risk functionality. They

are in the process of establishing a reporting mechanism with the NCA and will be using the IWF hash list to proactively identify CSAM. From their consumer email and cloud storage services, they anticipate identifying less than 100 instances of CSEA per year;

- **Jagex** – a large (over 500 employees) UK-based gaming business that owns games including RuneScape (average of 100,000 players online at any given time) and War of Legends, with players around the world. Jagex is likely to be representative of other UK based gaming businesses as functionalities are likely to be similar, but this is one of the largest gaming businesses in the UK. From November 2019 to November 2020, they made 38 reports for child protection reasons.
- **MovieStarPlanet** – a Danish social game aimed at children. In the last 18 months they have made one report of CSEA (grooming) to UK law enforcement.

209. It is estimated that c.24,000 organisations are in scope of the online safety regime; however, this includes all UK registered businesses and many of these will have parent or subsidiary businesses based outside of the UK that already report CSEA. To avoid duplicate reports being made, these businesses will not be required to report again under the UK reporting regime. Therefore, the actual number of businesses required to report and the number of reports are expected to be manageable. For example, Facebook UK will not report as Facebook is HQ'd in the USA and already reports to NCMEC. It is unclear at this time how many of the c.24,000 businesses will be required to report under this new requirement, and how many reports these companies are likely to make.

210. Some countries, including the USA and Canada, have legal requirements on businesses to report online CSEA. USA businesses are required by law to report to the National Centre for Missing and Exploited Children (NCMEC). In 2019, NCMEC received 16.8 million referrals containing nearly 70 million images. Of these, 79,798 reports related to victims or offenders in the UK and were triaged and then sent to the UK's National Crime Agency.

211. The overall cost to the UK's technology industry on reporting CSEA offences is minimal and to some extent controllable by the organisation (e.g. whether they use automated reporting). The cost for identifying CSEA is already part of the business's costs within their identification and moderation process. These processes will vary by business, with some proactively identifying CSEA using automation while others relying on user reports and human moderation. The cost of reporting is the time it takes to send a report of the identified content or activity to the specified body, which translates to the cost of an employee's time, unless the process is automated. The impact of sending a manual report is estimated to be 5-10 minutes of an employee's time per report. When applied to the hourly cost of a regulatory professional uplifted to account for non-wage labour costs (£25.21) this provides an estimated cost per report of £2.10 - £4.20. SMEs are likely to use a manual reporting process and to make fewer reports compared with large tech businesses who will have the resources to implement automated reporting systems and will make a higher volume of reports.

Table 24: Estimated annual costs for BT (low-medium risk) & Jagex (medium-high risk), both large businesses, based on previous reporting.

Business	Reports	Annual cost
BT	100	£210 - £420
Jagex	37	£78 - £155

212. NCMEC's international reports from 2019 by electronic service providers<sup>105</sup> demonstrate that smaller technology businesses report less than the big tech businesses. Most large technology businesses, particularly social media sites where CSEA is most likely to occur, are based in the USA and already report to NCMEC. The government does not have sufficient evidence to fully monetise the cost of the requirement to report online CSEA; however, Table 24 above provides an indication of potential scale of the impact per business annually. The impact on UK businesses is minimal, and the cost to smaller organisations to report is low as they will have fewer reports of CSEA to share with the designated body.

### Option 1 - Requirement to report online CSEA

213. This requirement is uniform across all options and costs are therefore expected to be the same under **Option 1**.

### Option 3 - Requirement to report online CSEA

214. This requirement is uniform across all options and costs are therefore expected to be the same under **Option 3**.

Table 25: Cost of the requirement to report online CSEA

	Option 0	Option 1	Option 2	Option 3
Requirement to report online CSEA	-	Not monetised	Not monetised	Not monetised

215. The below table outlines the total compliance costs across the options.

Table 26: Total cost of compliance (direct cost to business, monetised 10 year PV)

	Option 0	Option 1	Option 2	Option 3
<b>Total cost of compliance</b>	-	£1,306.1m	£1,734.8m	£6,951.4m

**Consultation question 10:** Do you agree with the estimates of potential compliance costs? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 11:** How (if at all) will the inclusion of user-generated fraud affect compliance costs? The government welcomes any evidence you can provide to understand the impacts.

**Consultation question 12:** Are there any additional costs associated with compliance not considered in this IA? The government welcomes any evidence you can provide.

<sup>105</sup> [2019 Reports by Electronic Service Providers - NCMEC](#)

## Industry fees

### Option 2 (preferred option) - Industry fees

216. The primary means of funding the regulator would be through an annual industry fee. businesses with in-scope services, with revenues over the specified threshold would be required to notify the regulator and pay an annual fee. The aim is that the annual fee will be used to fund the regulator's operating costs. OFCOM will work closely with industry through consultations to ensure the regime is proportionate. Additionally, OFCOM will consult with industry and then document the methodology and principles for the charging regime in the Statement of Charging Principles. This will be published ahead of each financial year.
217. The annual industry fee will be tiered and informed by OFCOM's regulatory timesheet data. This means that fees would be informed by the total quantum of costs (both direct and indirect) incurred by the regulator in regulating the sector. The total costs would therefore be recharged to industry through the annual fee, with the fees flexed accordingly.
218. The mechanics of the fee and notification regime are as follows:
- The regulator will be required to set a threshold, based on qualifying worldwide revenue, at or above which a business would be required to notify the regulator and pay a fee. This threshold will be subject to the agreement of DCMS Secretary of State.
  - Businesses below this threshold would not be required to notify or pay a fee. This will minimise the administrative burden on smaller businesses and on the regulator for the collection of fees and the storing of data for all businesses in scope. An appropriate threshold will ensure all small and medium enterprises are exempt from the direct costs of paying a fee and the compliance costs of notification.
  - For businesses under the threshold, the regulator will use other regulatory tools to monitor and obtain information usually collected at notification.
  - It will be up to the regulator to determine the frequency of notification. Businesses may be required to update the regulator of any significant changes to their revenue on an annual basis. This allows the fee being paid to reflect the business's financial circumstances each year. For example, a business which may have been exempt from paying the fee in year 1, may have increased its global revenue and therefore be liable to pay a fee in year 2.
  - The regulator will be required to tier the annual fee based on the qualifying worldwide revenue of businesses in scope.
  - The regulator may also choose to use an additional second metric, based on business activity. An activity-based metric would help ensure large businesses without significant relevant online activities pay a proportionate fee. The second metric may be calculated using criteria such as the number of specific functions on a platform although the exact weighting and determination of the metric will be determined by the regulator.
  - As with the primary metric of qualifying worldwide revenue, the second metric should be objective and have the ability to be applied in a uniform manner.
219. While the industry fee will depend on the realised costs to the regulator of operating the online safety regime, DCMS has worked with OFCOM to estimate a reasonable and realistic ten-year profile of operating expenditure. This assessment estimates that the annual industry fee on average could equate to **£46.0 million** per year (PV) and total **£346.7 million** across the appraisal period (10 year PV).
220. Under Section 22(4)(a) of the Small Business, Enterprise and Employment Act 2015, taxes, duties, levies and other charges are excluded from the Business Impact Target. This cost therefore has not been included in the calculations of the illustrative equivalent annual net direct cost to business (although it is included in the illustrative net present value).

## Option 1 - Industry fee

221. The funding mechanism for the regulator would be the same under **Option 1** and therefore, the industry fee is estimated to be the same. In practice, the fee may be lower under **Option 1** given the reduced scope of the framework; however, this IA does not quantify any possible reduction.

## Option 3 - Industry fee

222. The funding mechanism for the regulator would also be the same under **Option 3**. Given that requirements under **Option 3** are uniformly applied it may be the case that the regulator's operating costs would be lower under this Option, i.e. key decisions would be made before legislation and there would be less scope for the regulator to make enforcement decisions. However, given the stricter requirements under **Option 3**, it may also be the case that the regulator would be required to take enforcement action more often. In the absence of **Option 3** specific costs, the industry fee is assumed to be the same across all options.

Table 27: Annual industry fee (10 year PV)

	Option 0	Option 1	Option 2	Option 3
Industry fee	-	£346.7m	£346.7m	£346.7m

## Regulator enforcement powers

223. OFCOM will have a suite of enforcement powers to take action against businesses that fail to meet their regulatory responsibilities. These are warnings and notices, fines of up to £18 million or 10% of qualifying worldwide revenue, and business disruption measures including restricting access to non-compliant services in the UK. Such enforcement powers are proposed to be applicable across different types of businesses, e.g. size, revenue, activity, overseas; and be used proportionately to potential or actual damage caused, and size and revenue of the business. OFCOM will be required to consult and produce guidance setting out it will use its enforcement powers.

224. OFCOM will be expected to have a sufficient evidence base to take enforcement action. All enforcement action and decisions are expected to take place following an investigation. In some cases this process may be expedited due to the nature of the failure and OFCOM will apply its principles of proportionality. Should an investigation show a problem with the platform, OFCOM will decide what the appropriate enforcement action will be.

## Warnings and notices

225. Warnings are not a formal enforcement sanction; they may be issued after a pre-investigation phase, or investigation as an informal action. The warning may set out where the platform is at risk of breaching standards identified and recommendations for how the platform could rectify this. If the breach is rectified, there would be no further enforcement action.

226. In addition to warnings, OFCOM can issue a formal notification of their enforcement action to the platforms, setting out:

- the failure identified
- clear next steps that must be taken to rectify,
- deadlines and consequences of no action for the latter
- any penalty for the failure identified

227. In line with the approach in the Communications Act 2003, there will be provisions for organisations to negotiate terms of notifications and/or settle early to secure a discount on the relevant penalty.

228. Warning and notices in and of themselves are unlikely to result in material costs to businesses, beyond administrative costs of engaging with the process, e.g. ensuring the warning or notice is handled appropriately within the organisation. As is standard practice in regulatory appraisal, compliance costs estimated above assume full compliance with the regime. Therefore, any costs to business from rectifying actions undertaken as a result of receiving a warning and/or notice is already captured in our assessment of compliance costs.

## Fines

229. Under the new regulatory framework, investigations conducted by OFCOM can end with an in-scope organisation being issued a monetary penalty for failing to comply with the requirements.

230. The approach to enforcement will aim to encourage compliance and drive positive cultural change. The regulator will support businesses to help them understand the expectations placed on them, and how the regulator's use of its enforcement powers will be proportionate. OFCOM will have the power to issue fines against businesses that fail to for example:

- fulfill the duty of care
- notify the regulator where required
- pay annual industry fee (if required to do so)
- provide transparency reports (if required to do so)
- respond to information requests
- cooperate with the regulator, or the skilled person, in relation to the regulator's use of its power to require a skilled person report
- to comply with a use of technology notice

231. These powers will be comparable to those already used by OFCOM and other UK regulators. OFCOM will use its enforcement powers in line with its duties and will ensure they are used proportionately, taking into account the level of harm and considering the impact on children.

232. The regulator's enforcement powers will include issuing directions for improvement and notices of non-compliance and issuing sanctions in the form of civil fines up to £18 million or 10% of annual global turnover, whichever is higher. The fine limit is in line with the limits for those currently issued by OFCOM, the Financial Conduct Authority and the Competition and Markets Authority. In cases of repeated or particularly egregious non-compliance, OFCOM will be able to take measures to disrupt a business's activities in the UK, including restricting access to the non-compliant service in the most serious circumstances. If a business fails to meet its regulatory responsibilities, the regulator may be able to pursue enforcement action against a parent business that has sufficient control over the non-compliant business. The regulator may also be able to pursue enforcement action against subsidiaries of parent businesses, where they are involved in the breach of a group business.

233. It is not possible at this stage to provide accurate estimates of fines resulting from enforcement of the OS regime as the level of fines (up to the limit set out in the OS Bill) and the frequency with which they are levied on businesses will depend on both the specific details of the codes of

practice and the level of non-compliance. While, fines and penalties are excluded from the Business Impact Target under administrative exclusion G for the current Parliament (interim guidance), for illustrative purposes, details of fines issued by the ICO for non-compliance with the requirements it enforces can provide an indication of the likely scale of impact.

234. In 2019/20, the ICO took regulatory action 236 times which resulted in fifteen fines issued. Fines ranged from £15,000 to £400,000 with an average fine of around £120,000 and a total of **£1.8 million**<sup>106</sup>. As noted, these costs are illustrative only and not included in either the net present social value or equivalent net direct cost to business.

Table 28: Fines for non-compliance

	Option 0	Option 1	Option 2	Option 3
<b>Fines for non-compliance</b>	-	Not monetised	Not monetised	Not monetised

## Business Disruption Measures

235. In the most egregious instances of non-compliance, OFCOM will have the power to initiate business disruption measures, to be used as a last resort where other interventions have failed to tackle the harm occurring on a service.

236. Under Level 1 business disruption measures, OFCOM will have the power to seek a court to require third parties to take measures that make it less commercially viable for a non-compliant business to provide services to UK users by requiring those service providers to withdraw access to key services. This includes services that facilitate the provision of the non-compliant business, or display content relating to such a business. Examples of such services include payment services which enable funds to be transferred to a non-compliant business, search engines which display content relating to a non-compliant business, social media services which make content relating to a non-compliant business available to users and services which facilitate the display of advertising on a non-compliant business.

237. Under Level 2 business disruption measures, OFCOM will have the power to seek a court order requiring third parties to take measures that restrict access to a non-compliant business's services in the UK, by requiring the withdrawal of key services by internet infrastructure providers (e.g. browsers, web-hosting businesses, app stores, online security providers or Internet Service Providers).

238. As noted, these measures will only be used as a last resort and this type of regulator action is expected to be rare. The frequency with which these measures are used (and therefore the potential costs to businesses such as payment service providers and Internet Service Providers (ISPs)) depend on future codes of practice, the level of compliance, and the effectiveness of preceding regulator action on in-scope organisations, e.g. warning and notices and fines - all of which are unknown at this stage. The impacts of these are therefore not included in the illustrative NPSV or EANDCB.

239. While this IA is not able to accurately estimate these costs, to provide an indication of the likely scale of potential impacts, estimates from the IA for 'Age verification for pornographic material

<sup>106</sup> [ICO Enforcement Action \(2021\)](#)



online<sup>107</sup> which similarly involved notifying payment service providers and ISPs of sites in breach, enabling them to withdraw their services and/or initiate blocking are presented. It was estimated here that the cost to payment service providers of working with the regulator and processing requests would be approximately **£0.5 million** per year and the cost to ISPs - based on a domain name system approach to blocking - was estimated to be between **£0.1 million to £0.5 million** per year. In the context of this Bill, access restriction is expected to be very rare and a last resort in instances of egregious harm and non-engagement. The Government is engaging with relevant third party services to attain cost estimates specifically in the context of the online safety framework but they are expected to be lower than those estimated in the IA for 'Age verification for pornographic material online'.

Table 29: Business disruption measures

	Option 0	Option 1	Option 2	Option 3
<b>Business disruption measures</b>	-	Not monetised	Not monetised	Not monetised

## Costs to individuals

### Option 2 (preferred option) - Costs to individuals

240. Under the preferred option, some individuals (who own an online platform) might be in scope of the new regulation if they meet all of the criteria. If this were the case, these individuals would incur the same costs as any business in scope. However, given the low risk functionality exemption, the vast majority (if not all) individuals are likely to be out of scope. This is to futureproof the regulations as technologies develop which lower the bar to entry; and to prevent a loophole under which bad actors could make individuals (rather than companies) the service provider to evade regulation. The government does not have any evidence of individuals who could be in scope of the regulation and given the proportionate and risk-based design of the regime, these are expected to be extremely rare and any costs minimal.

### Option 1 - Costs to individuals

241. As under the preferred option, under **Option 1**, there may be individuals within scope of the regulations; however, these are expected to be rare and the costs to be negligible.

### Option 3 - Costs to individuals

242. Under **Option 3**, uniformly applied safety duties do not take account of the risk of online harms on the regulated service. While it would still be rare for individuals to be in scope of the regulations, any that are would face the full costs of compliance. This would likely lead to the individual shutting the service down as the costs could be too prohibitive.

Table 30: Costs to individuals of compliance

	Option 0	Option 1	Option 2	Option 3

<sup>107</sup> [Age Verification for Pornographic Material Online, Impact Assessment - DCMS \(2018\)](#)



Costs to individuals	-	Negligible	Negligible	Negligible
----------------------	---	------------	------------	------------

**Consultation question 13:** Do you agree with the assessment that the costs to individuals (i.e. not businesses or civil society organisations) will be negligible? The government welcomes any evidence you can provide to refine the estimates.

## Costs to government

### Justice impacts

243. The regulator will be able to take enforcement action against any in-scope business that fails to meet its regulatory responsibilities. The independent regulator's core enforcement powers will be to issue warnings, notices and civil fines to businesses. Businesses and the public would be able to seek judicial review of the regulator's actions and decisions through the High Court. The enforcement decisions of the Online Safety regulator can also be appealed via the Upper Tier Tribunal: Administrative Appeals Chamber. OFCOM decisions to either allocate a business as a Category 1 provider, or require a business to provide transparency reports, may also be appealed here by affected service providers. Complaints about OFCOM can also be made to the Parliamentary and Health Service Ombudsman.

244. The primary objective of the proposal is for more effective action by businesses to respond to harmful content/activity, reducing the damage done by unacceptable or illegal content/activity and providing a safer online environment. The proposal is not primarily intended as a criminal justice solution, but to encourage more effective preventative and remedial action and effective assistance from businesses to law enforcement (where appropriate).

245. This IA estimates a potential impact on the criminal justice system in four areas:

- **The introduction of (i) a new criminal offence** for entities that fail to comply with information requests **(ii) a potential new criminal offence** for named senior managers who fail to comply with information requests **(iii) potential new criminal liability of corporate officers** where an entity's failure to comply with an information request is committed with the consent or connivance of a director or senior official
- **Court orders** required to carry out an investigation, business disruption measures or to enforce requirements in the regulator's confirmation decisions.
- **A new appeals process, via the Upper Tier Tribunal: Administrative Appeals Chamber** providing an accessible and affordable alternative route to appeals for businesses affected by the regulator's decisions and parties with sufficient interest in the decisions of the regulator, where relevant. Appeals - which are expected to amount only to low single figures - will be heard using judicial review (JR) principles.
- An **impact on the number of incidences of online illegal activity and content reported** to law enforcement and/or other authorities. The government cannot be specific on future numbers, but has conducted a Justice Impact Test which contains a comprehensive analysis of the current numbers, problems with accounting data, likely scenarios, and the existing obligations on social media platforms, and those in scope, to report illegal harms. In the longer term, the new duty of care is expected to decrease the need for referrals of illegal and harmful online content to appropriate authorities.

246. Current estimates indicate that the only costs occurring will be for the appeals body, with an estimate from the Ministry of Justice of £42,000 for the first year made up of:

Start-up cost : £7,000  
 Running cost : £35,000 (cost per case : £3,500)

247. For the purposes of the IA, the appeals body cost estimates are rounded up to £50,000. Ongoing costs for future years may be lower or greater and would be dependent on the number of cases being heard. Given the uncertainty around the number of future cases, this IA assumes justice impacts estimated here are constant across the appraisal period.

248. Justice impacts have been assessed only for the preferred option. In practice, they are likely to occur under all the policy options but they could potentially differ under **Option 1** and **3**. For ease, this IA assumes justice impacts are the same across the options (this assumption - given the scale of the impacts - does not materially affect the main metrics of the IA).

Table 31: Justice impacts (10 year PV)

	<b>Option 0</b>	<b>Option 1</b>	<b>Option 2</b>	<b>Option 3</b>
<b>Justice impacts</b>	-	£0.4m	£0.4m	£0.4m

## Requirement to report online CSEA

249. This section sets out estimated costs to the government of establishing a body that will be responsible for receiving and processing CSEA reports. The costs will vary significantly depending on the number of reports that are made, and which body takes on this role.

250. In 2019, electronic service providers in the US made 16.8m reports. Excluding reports from Facebook, Google and Microsoft, all other US service providers made 1.28m reports. Even if Facebook, Google and Microsoft are excluded, the US's tech sector is substantially larger than that of the UK or any other country. The government therefore, expects far fewer than 1.28m reports to be made by UK businesses.

251. Based on engagement with UK technology companies and law enforcement and knowledge of the UK technology sector, the government expects that the UK reporting body will receive a maximum of 10,000 reports annually. UK companies currently report to UK police forces on a voluntary basis, and of the 7 police forces engaged this far, only two have received any industry reports at all, with the Met Police receiving the most reports in a 12-month period (92). Even large UK based in-scope services make very few reports of CSEA. For example, the UK gaming company Jagex (which owns Runescape and other popular games) has over 500 members of staff and 100,000 players online at any given moment. From November 2019 to November 2020, they made just 38 reports for child protection reasons. However, the nature of the technology sector means that the number of reports made may rapidly change if a new app, game or trend quickly becomes popular, or new offender behaviours may result in a sudden increase in online CSEA. The designated body will therefore need systems and infrastructure that have the capacity for receiving much greater numbers of reports.

252. Three bodies that could take on the role of the UK reporting body have been identified and further work is ongoing to determine which of these bodies will be designated. Table 32 below sets out some of the key costs across the system for these three bodies. The primary cost will be setting up technological systems and infrastructure to enable the designated body to securely receive, process and store industry reports. There will be further costs relating to the analysis and onward referral of these reports to law enforcement agencies, and the necessary resources for investigations by law enforcement. The costs below reflect these requirements and will enable

the designated body to effectively meet their requirement and allow law enforcement to act on the reported cases. The government is engaging with the IWF and NCMEC to attain more accurate figures.

Table 32: Key costs expected to be incurred by government from the requirement to report CSEA

Organisation	Setting up the system	Analysis	Investigations by law enforcement	Total
NCMEC	NCMEC is exploring this and it will likely be low cost as technology and processes already exist. Legal and data protection implications require further assessment given that they are a US based organisation.	\$100k (£80k) for one additional analyst.  Each actionable report passed onto the NCA costs the NCA £140 to process.	Approximately 10% (200) of reports the NCA receive will be passed to law enforcement.  £2,100 - £5,700 for each referral to law enforcement (staff costs only)	Legal costs in addition to £140 per report referred to the NCA plus £2,100-£5,700 for the 10% of reports that result in a law enforcement referral
NCA Referrals Bureau	£8m investment in technology. The NCA does not have processes in place to receive reports.	£140 per actionable report received.	Approximately 10% (200) of reports received will be passed to law enforcement.  £2,100 - £5,700 for each referral to law enforcement (staff costs only)	£8m set up costs plus £140 per report plus £2,100-5,700 for the 10% of reports that result in a law enforcement referral
IWF	Further engagement with the IWF needed	Further engagement with the IWF needed	Approximately 10% (200) of reports received will be passed to law enforcement.  £2,100 - £5,700 for each referral to law enforcement (staff costs only)	Further engagement with the IWF needed

# Benefits

## Treatment of benefits

253. The calculation of benefits of each option is challenging: it is not possible to develop a precise estimate of the reduction in online harms that will be achieved by the policy options. This is due to:

- limited longitudinal data on the impact of internet use given the way in which the nature of the internet and its uses have evolved over time;
- the novelty of the proposed policy measures, which means there is a lack of relevant precedent in other sectors or countries;
- the scale of the internet and the way in which it is used, which means that it is not possible to run trials or experiments in a way that can be robustly scaled up;
- the rate of change in the sector and the way people use technology; and,
- ultimately, the regime will be implemented and operated by OFCOM and, therefore, government has limited control; there is also uncertainty as to how platforms will change their behaviour in response to new regulation

254. A key objective of the OS Bill is to reduce the prevalence of online harms. The effectiveness of the policy will be assessed against this objective and others through a robust monitoring and evaluation programme (see Monitoring and Evaluation section). However, at this stage, and given the reasons stated above, the benefits estimated in this IA are illustrative only and have not been included in the net present value of the policy.

255. As with other similar uncertain policies, to address the problems with benefit estimations, this IA presents break-even analysis: estimating the reduction in online harms<sup>108</sup> required to exactly match the economic costs.<sup>109</sup> The methodology used to estimate this and the results are in the sections below.

## Methodology

256. This section outlines the analytical approach taken to understanding and estimating the size of the benefits of intervention, including how these potential benefits are weighed against the costs of the preferred way forward, and how uncertainty around some of the monetised impacts and effectiveness of intervention is accounted for.

257. In the simplest terms, in the context of online harms, this IA considers the anticipated illustrative benefits to be any online harms avoided as a result of the preferred policy. At a very simplified level, this can be expressed as:

***Anticipated benefits = Total monetised impact of harms over the forecast period absent policy intervention \* % reduction in harms expected to result from the policy intervention<sup>110</sup>***

258. Based on the formula outlined above, there are two core elements to benefits evaluation:

- The total monetised impact of harms; and
- The expected reduction in harms (in % terms) due to the policy intervention.

---

<sup>108</sup> Harms which we were able to quantify and are therefore included in the estimated benefits, are: cyberbullying, cyberstalking, intimidation of public figures, child sexual abuse and exploitation, modern slavery, hate crime, drugs facilitated online.

<sup>109</sup> These costs comprise all monetised costs within this impact assessment.

<sup>110</sup> The percentage reduction in harms can be considered to be the effectiveness of the policy.

259. The first element outlined involves understanding the prevalence, impact and subsequent cost of a range of harms, both illegal and legal but harmful, encountered online including many of those set out in the OHWP<sup>111</sup>. The monetised impacts that have been estimated are focussed on the subset of harms where the available data is the most robust and therefore greatly understate the societal impact of online harms as a whole.

260. It is assumed that the quantified harms grow at 5% per year. This is in line with growth in the amount of hours spent online (an average of 5.4% per year over the last four years).<sup>112</sup> While the rate of internet use (as a percentage of the UK population) is broadly constant, growth in the UK population means the number of UK internet users is increasing, meaning the rate of growth may be faster than this. It should be noted that this IA is not able to model any impact that changes in demographics may have on the distribution of different harms - the estimates assume that the current distribution grows at 5% per year. This assumed growth rate is also in line with the 5% annual growth in the total number of videos viewed online since 2017 as indicated in the DCMS commissioned **AVMSD research**.

261. There are a wide range of different harms, both illegal and legal but harmful. Of these, a total of seven defined harms have been quantified, at least partially, based on available evidence. These include five illegal harms:

- Child Sexual Exploitation and Abuse (CSEA);
- Modern slavery;
- Hate crime;
- Illegal sales of drugs; and
- Cyberstalking.

262. Two further harms that are legal but harmful have also been costed:

- Cyberbullying
- Intimidation of public figures.

263. Quantitative evidence is provided to demonstrate the scale of the problem as well as more qualitative assessments based on expert judgement. These calculations rely on a number of uncertain assumptions, proxies and experimental data. They do not reflect a government view of the impacts, rather, they represent simplified, indicative estimates designed to enable analysis of online harms.

264. One of the challenges of estimating the online element of illegal harms is that the way in which harms occur varies, with some harms being purely online (e.g. viewing indecent images of children) and others taking place offline but being facilitated through online activity (e.g. grooming children online prior to a physical offence).

265. With regard to the illegal harms that have been quantified, this IA uses data on the prevalence of crime from the Office for National Statistics (ONS) which includes experimental data based on the online crime flag. In April 2015, it became compulsory for police forces to flag whether crimes are committed online (in full or in part). This does not provide information on the extent of the online component, that is, whether it was a significant or a minor part of the offence. It also does not provide information on whether, in the absence of the online component, the offence would still have taken place via alternative means.

266. In addition, many of these harms can involve both online and offline elements, which are often closely linked (e.g. traditional bullying and cyberbullying). It can therefore be difficult to completely disaggregate the impacts.

---

<sup>111</sup> [Online Harms White Paper, April 2019](#), see pg. 31 for a comprehensive list of the types of harms under consideration.

<sup>112</sup> Based on OFCOM's Adults' Media use and attitudes report, 2016-2019

267. As well as issues relating to the use of the flag by police forces,<sup>113</sup> an additional limitation is that not all crime is reported and recorded by the police. Therefore, the Crime Survey for England and Wales (CSEW) is generally preferred as a source of data to establish the prevalence of crime<sup>114</sup> since it allows for the measuring of “hidden” crime (that is, crime that is not reported and therefore that law enforcement does not come across). Consistent with other Home Office analysis, this IA uses a multiplier approach to uplift the ONS data to take account of actual levels of crime rather than just reported crime.

268. All quantified illegal harms below contain the cost to the Criminal Justice System (CJS), aside from Modern Slavery<sup>115</sup>. The CJS costs for Cyberstalking are set out explicitly in the related cost table. For all other quantified harms, the full methodology, including the costs covered, is set out in the associated Home Office statistics publication, each of which can be found in the footnote for each harm.

269. For some harms there is inadequate quantitative evidence to enable the government to develop a rough estimate. This is because the true prevalence of harmful content or activity may be unknown, and because of the shortcomings of data that is available (for example, screen time does not reflect what that time was used for). In some cases, it was not possible to establish a causal link between online activity and the harm.

## Quantifying online harms in the baseline

### Physical child sexual abuse and exploitation

270. The following categories of physical child sexual abuse and exploitation that were flagged as having an online component were considered:

- Rape; and
- Other physical sexual offences.

271. All quantified benefits in this section are presented in 2019 prices and the majority of police recorded crime figures, especially those based on Home Office analysis of crime data - are 2019. For both the break-even analysis and illustrative benefit scenarios, the estimated prevalence of quantified harms in the baseline was uplifted at an annual growth rate of 5% to 2023 (the date of implementation) and assumed to continue to grow in line with hours spent online. As noted previously, this assumed growth rate is tested with sensitivity analysis.

272. The table below summarises the data and calculations used to estimate the impact of physical child sexual abuse and exploitation with an online element in 2019, based on Home Office analysis of police recorded crime data.

Table 33: Annual impact of physical child sexual abuse and exploitation (with online element)

---

<sup>113</sup> This is defined as “An offence where the reporting office believes that on the balance of probability, the offence was committed, in full or in part, through a computer, computer network or other computer-enabled device”. Source: Counting Rules Crime Flags, Home Office, July 2019. We note that a potential limitation of this approach is that the use of the online flag is a manual process and inherently relies on an element of subjective judgement. There is evidence of inconsistent use of the flag across police forces, with forces typically tending to underuse it given that it is not a mandatory requirement and it has little operational impact compared to other flags. While the data shows an increasing volume of offences with an online flag, this is likely largely due to increased use of the flag rather than an increasing online component of crime.

<sup>114</sup> ONS (July 2019). Crime Survey for England and Wales,

<sup>115</sup> It has not been possible to estimate the cost to the CJS for a number of reasons. Modern slavery offences that go through the CJS are long and complex and can often take up to two years to complete. This is reflected in the proceedings data for these offences. The cost model that the Ministry of Justice used to estimate the cost of other crime types relies on a full set of data to profile the cost through the courts for a given year. Because of the lags from a criminal proceeding being commenced to its disposal, the data for all modern slavery offences produces results that are not reliable.

	Category	Prevalence	Unit cost	Annual cost
Physical child sexual abuse and exploitation	Rape	442	£54,516	£24.1m
	Other Physical sexual offences	124,856	£14,801	£1,848.0m
<b>Total £1,872.1m</b>				

273. The calculations and data used to estimate the impact of child sexual abuse and exploitation can be further broken down as follows:

- Prevalence is calculated using Home Office analysis of police recorded crime data in 2018/19 multiplied by a 3.4 multiplier as a central estimate.<sup>116</sup> An alternative estimate of prevalence uses a multiplier of 8.0 based on an external estimate that is specific to CSEA.<sup>117</sup>
- The economic and social unit cost of rape can be broken down into total emotional harm (£53,019), total physical harm (£131), and health services costs which includes both emotional and physical (£1,088 and £278 respectively).<sup>118</sup> This gives a total unit cost of £54,516. These are internal estimates provided by the Home Office based on analysis on the safeguarding of children.

274. The calculations and data used to estimate the impact of other sexual offences can be further broken down as follows:

- Prevalence is calculated using Home Office analysis of police recorded crime data in 2018/19 \* 16.5 multiplier. An alternative estimate of prevalence uses a multiplier of 8 based on an external estimate that is specific to CSEA.<sup>119</sup>
- The economic and social unit cost of this crime can be broken down into total emotional harm (£14,262), total physical harm (£52), and health services costs - both emotional and physical (£362 and £125 respectively). This gives a total unit cost of £14,801. These are internal estimates provided by the Home Office based on analysis on the safeguarding of children.

## Hate crime

275. Hate crime relates to a range of offences including violence against the person (VATP), public order offences, criminal damage and arson offences, and other notifiable offenses. The quantification of harms is focussed on violence against the person as this represents the majority of hate crime offences that are flagged as online, and cost data is not available for the other offences. Therefore, this cost estimate is likely to underestimate the total cost of online hate crime.

276. The table below summarises the data and calculations used to estimate the impact of hate crime with an online component.

Table 34: Annual impact of hate crime (with online element)

	Category	Prevalence	Unit cost	Annual cost
--	----------	------------	-----------	-------------

<sup>116</sup> This multiplier is calculated using CSEW data for adult sexual offences and rape offences as proxies in the absence of specific CSAE multipliers.

<sup>117</sup> [Protecting children from harm - Children's Commissioner \(November 2015\)](#).

<sup>118</sup> Source for cost and multiplier: Home Office (July 2018). Economic and Social Costs of Crime. Second edition, pp.11

<sup>119</sup> [Protecting children from harm - Children's Commissioner \(November 2015\)](#).

Hate crime	Violence against the person	1,926	£6,301	£12.1m
<b>Total £12.1m</b>				

277. The prevalence of crime relating to violence against the person with an online element is calculated using Home Office data on reported crimes (1,284).<sup>120</sup> A multiplier of 1.5 is applied to reflect unreported crime. This is based on data on the closest proxy which is “violence without injury offences” since the vast majority of VATP hate crime offences relates to malicious communications (86%) and these offences appear unlikely to lead directly to physical injury. This is likely to be a conservative approach since other types of online hate crime are not covered. These include public order offences, criminal damage and arson offences and other notifiable offences. This is due to cost data for these offences not being available.

278. Applying an alternative multiplier allows us to consider a sensitivity analysis. In particular the Home Office hate crime statistical bulletin indicates that only 53% of hate crime incidents came to the attention of the police in 2017/18, suggesting a multiplier of 1.9 may be appropriate (producing an alternative estimate of prevalence of 2,423).

279. Since there is no data available specifically for the cost of online hate crime offences, data on the cost of “violence without injury” is used as a proxy, inflated to 2018/19 prices. It is unclear whether this will tend to under or over-estimate costs for the following reasons:

- Physical violence could result in greater harm than online VATP offences;
- The volume data is likely to be understated as it is based on data from only 30 police forces; and
- The multiplier applied is relatively low compared to alternative proxies and therefore represents a conservative approach.

## Illegal sale of drugs

280. The available police recorded crime data contains figures for drugs offences flagged as having an online component. In 2018/19, 85 drugs offences were recorded as being online. These recorded figures are very low and there is no comprehensive set of reasons as to why this is, it is likely to be to do with the way the online flag is applied to drug cases. This could be because it is difficult to prove there is an online component, and that a lot of drug trafficking is from overseas and so may not be recorded in the online data.

281. Unit cost data is unavailable for this harm and so a top down approach has been taken instead. The total social and economic cost of organised drugs supply is estimated to be £20 billion<sup>121</sup>. Inflating this figure from 2015/16 prices to 2018/19 provides a total estimate of £21.25 billion. The proportion of recorded drugs offences flagged as online was around 0.1% in 2018/19. Combining these proportion and cost estimates provides an indicative estimate of the cost of drugs offences flagged as having an online component of around **£14.9 million**.

Table 35: Annual impact of illegal sale of drugs (with online element)

	Category	Prevalence	Unit cost	Annual cost
Illegal sale of drugs	Drug offences with an online component	85	n/a	£14.9m

<sup>120</sup> [Hate Crime, England and Wales, 2017/18 - Home Office \(October 2018\)](#).

<sup>121</sup> [Understanding organised crime 2015/16 Estimating the scale and the social and economic costs. Second edition. - Home Office \(February 2019\)](#).



**Total £14.9m**

## Modern slavery

282. This section considers the economic and social cost of physical modern slavery offences with an online element.<sup>122</sup>

Table 36: Annual impact of modern slavery (with online element)

	Category	Prevalence	Unit cost	Annual cost
Modern slavery	Modern slavery cases with an online element	100	£341,499	£34.2m
<b>Total £34.2m</b>				

283. The unit cost of modern slavery is given as £328,720 in the Home Office report “The economic and social costs of modern slavery”<sup>123</sup>. It covers the costs of physical & emotional harm, the cost of lost output & time, costs to health services, costs to victim services and law enforcement costs.

284. This unit cost is given in 2016/17 prices. Inflating the estimate to 2018/19 prices, provides an estimate of £341,499. This cost relates to physical modern slavery offences. It is assumed, for the purposes of this analysis, that modern slavery offences do not take place purely online. There could theoretically be a scenario where the definition of modern slavery could be met with an entirely online situation, but that would be highly unusual and infrequent. The facilitator needs to somehow benefit which could be difficult virtually.

285. This unit cost can then be applied to an estimate of modern slavery offences with an online component, to provide an estimate of the impact of these offences. As with CSEA, this estimate does not involve any judgement as to the extent of the online component, or what would happen in the absence of the online component. It simply reflects an estimate of the cost associated with modern slavery offences flagged as having an online component.

286. There were estimated to be a total of 10,000 to 13,000 modern slavery offences per year<sup>124</sup>. Taking the central point of this range as the central estimate for the number of cases, gives 11,500 cases per year. Multiplying this number of cases by the proportion flagged as online (0.87%) gives 100 cases with an online component. This is then multiplied by the unit cost outlined above (£341,499) to give a central estimated cost of **£34.2 million** for modern slavery with an online component.

## Cyberstalking

<sup>122</sup> While there could theoretically be a situation where the definition of modern slavery could be met with an entirely online situation, this would be highly unusual and infrequent, particularly as it would mean that the facilitator would be able to benefit virtually.

<sup>123</sup> [The economic and social costs of modern slavery - Home Office \(July 2018\)](#).

<sup>124</sup> Modern Slavery: an application of Multiple Systems Estimation - Professor Bernard Silverman for Home Office (2014).

287. There is no single definition of cyberstalking, however it is widely considered to refer to the repeated use of online communications tools to stalk, harass or frighten a victim. Online services may be used for a range of purposes in this regard, for example:

- to locate personal information about a victim;
- as a means of surveillance of the victim;
- to directly harass the victim;
- to carry out identity theft;
- to carry out electronic sabotage such as spamming; and
- to damage the reputation of the victim.

288. The table below sets out the data and calculations used to estimate the impact of cyberstalking.

Table 37: Annual impact of cyberstalking

	Category	Prevalence	Unit cost	Annual cost
Cyberstalking	Impact on victims	68,832	£30,253	£2,082.4m
	Cost in police time	68,832	£293	£20.2m
	Cost in criminal justice system (CJS) time	13,766 offences proceeded against <sup>5</sup>	£1,070 <sup>6</sup> (assuming 20% of offences lead to charges)	£73.6m
<b>Total £2,176.2m</b>				

289. The estimated number of incidents is based on the number of police recorded stalking and harassment offences in England and Wales – 458,881 in the period from July 2018 to June 2019. This is then multiplied by the proportion of stalking and harassment offences flagged as online crime by the police (15%) to give an estimate of cyberstalking offences.<sup>125</sup>

290. This figure is likely to underestimate the true prevalence of cyberstalking in the UK for two reasons. Firstly, some incidences of cyberstalking will go unreported or will not result in a recorded offence. Secondly, the majority of stalking cases - including those not flagged as an online crime - are still likely to involve an online element (such as the use of social networks). This estimate is therefore likely to be conservative.

291. The unit cost of a cyberstalking episode is based on the cost to the victim of a stalking episode from a 2019 Home Office report.<sup>126</sup> According to Paladin, the national stalking advocacy service, cyberstalking inflicts the same amount of psychological damage as offline stalking.<sup>127</sup>

292. The unit cost comprises three elements: emotional cost to the victims (£21,920), cost to health services (£1,210) and cost in lost productivity (£6,560). The total has been uplifted to 2019 prices. The number of police investigations is assumed to be the same as the number of recorded offences. Each investigation is assumed to take 7.5 hours of a police officer's time (rank sergeant

<sup>125</sup> The [online flag](#) is defined as "an offence where the reporting officer believes that on the balance of probability, the offence was committed, in full or in part, through a computer, computer network or other computer-enabled device."

<sup>126</sup> [The economic and social costs of domestic abuse. Research Report 107 - Home Office \(2019\).](#)

<sup>127</sup> [Stalking and Harassment: a Shorthand Guide - Paladin \(2014\).](#)

or below), based on mid-point estimate used in the IA for Strengthening the Law on Domestic Abuse (2014)<sup>128</sup>. The cost per investigation has been uplifted to 2019 prices<sup>129</sup>.

293. This IA assumes a 20% charge rate per offence, based on police data for the charge rates of domestic abuse offences. Each prosecution is assumed to cost the CJS an average of £5,000 per defendant, based on costs used in the IA for Strengthening the Law on Domestic Abuse (2014). Uplifting to 2019 prices equates to an average cost to CJS of £1,070 per offence.

## Cyberbullying

294. Cyberbullying is defined as bullying which takes place over digital devices, such as mobile phones, tablets and computers. Cyberbullying can be both public and private, acting on public forums or through private messaging.<sup>130</sup> Cyberbullying can take the form of many behaviours including: harmful messages; impersonating another person online; sharing private messages; uploading photographs or videos of another person that leads to shame and embarrassment; creating hate websites/social media pages; and excluding people from online groups.

295. Whilst the lines between cyberbullying and traditional bullying can sometimes be blurred, online bullying does have a number of elements that make it different from traditional bullying. In particular, cyberbullying can occur 24/7 and may be seen and shared by a much wider audience. A cyberbullying incident may also have a much longer lasting impact. Further, anonymity can make cyberbullying incidents more intimidating, and the degree of separation between bully and victim can make it hard for perpetrators to appreciate the impact of their behaviour.<sup>131</sup>

296. Given the majority of academic research available focuses on the impact of cyberbullying on young people, the estimates used in this analysis focus on the impacts on those aged 10 to 15 years old (based on the age range typically used in cyberbullying studies). Therefore, the estimate will underestimate the impact of cyberbullying on the UK as a whole. The table below outlines the core unit costs for the central estimate of the economic costs of cyberbullying.

Table 38: Annual impact of cyberbullying

	Category	Prevalence	Unit cost	Annual cost
Cyberbullying	Direct impact on victim	777,177 (17% of 10-15 year olds)	£640	£497.1m
	Cost to health services of treating related depression	72,464 children accessing specialist mental health treatment	£354*	£25.7m
	Cost of treating cyberbullying related self-harm	1,943 children	£838*	£1.6m
	Lifelong impact	Estimated 14,000 16 year olds each year	Women: £2,335 Men: £7,031	£184.3m
	<b>Total</b>	<b>777,177</b>	<b>£673</b>	<b>£522.7m<sup>132</sup></b>

\*Assumption of one incident in a given year

<sup>128</sup> [Strengthening the Law on Domestic Abuse, Impact Assessment - Home Office \(2014\)](#).

<sup>129</sup> Uplifted using HMT's 'GDP deflators at market prices, and money GDP'.

<sup>130</sup> [What is Cyberbullying? - StopBullying.gov \(2018\)](#)

<sup>131</sup> [Bringing an end to online bullying: Whose job is it anyway? - Anti-Bullying Alliance \(2019\)](#).

<sup>132</sup> This figure does not include the cost of treating cyberbullying related self harm or the lifelong impact - these are shown illustratively only given the minimal evidence base.

297. The below sections will outline the methodology for each of the estimates in the rows of the table above.

298. There are an estimated 4.57 million children in the UK aged 10-15.<sup>133</sup> This number, combined with figures on the prevalence of cyberbullying amongst children (estimated at 17% based on a number of studies from 2017)<sup>134</sup><sup>135</sup>, is used as the basis for our estimates. This central prevalence estimate of 17% equates to 777,177 children victims of cyberbullying in the UK in a given year. Based on the range of prevalence estimates in the studies observed, sensitivity is conducted using 8% and 23% as upper and lower bounds which results in 320,000 and 1,051,000 cyber bullied children..

### Costs to the individual (QALY-based methodology)

299. The costs to the victim of a cyberbullying incident include the impact on the victim's mental health and wellbeing, which may result in a depressive episode. This impact is estimated using quality-adjusted life years (QALYs), which enables quantification (in monetary terms) of the impact of various health conditions on a person's quality of life.

300. To estimate the cost of a minor/moderate depressive episode required information includes:

- the likelihood of sustaining depression (LIKE);
- the percentage reduction in quality of life (REDUCEQL);
- the duration of the depressive episode (DUR) as a fraction of a total year; and
- The value of a year of life at full health (VOLY).<sup>136</sup>

301. These are multiplied together to give an estimate of the average cost associated with the crime. The formula is as follows:

- ***LIKE \* REDUCEQL \* DUR \* VOLY = Average physical and/or emotional cost***

302. On this basis, the depression associated with non-violent crime, which is used here as the closest available proxy for the impact of cyberbullying, has a QALY loss (REDUCEQL) of 14.5%<sup>137</sup>. The duration (DUR) is estimated at 0.167 years (or 2 months) and a value of a life year (VOLY) of £71,385 (uplifted to 2019 prices). Therefore, the unit cost is  $0.145 * 0.167 * £71,385 = £1,728$ <sup>138</sup>. This £1,728 unit cost can then be multiplied by the probability of harm occurring (LIKE) – that is, what proportion of victims of cyberbullying suffer depression as a result. An annual bullying survey in 2017 found that 37% of those who were victims of cyberbullying went on to suffer from depression.<sup>139</sup> As such:

- ***Unit cost of episode of depression: £1,728 \* 0.37 = £640***

---

<sup>133</sup> [ONS Population Estimates \(2018\)](#)

<sup>134</sup> This figure is based on averaging prevalence estimates from different government and academic studies on cyberbullying in the UK for 2017: [Annual Bullying Survey 2017, Ditch the Label](#); [Mental Health of Children and Young People in England, 2017, NHS Digital](#); [The Suffolk Cybersurvey 2017](#); and [Bullying in England, April 2013 to March 2018 Analysis on 10 to 15 year olds from the Crime Survey for England & Wales](#)

<sup>135</sup> For sensitivity, an estimate was also produced looking at a wider range of studies between the years 2013 and 2017 which also produced an average prevalence of 17%.

<sup>136</sup> Valued at £60,000 by the Department of Health (DfE) and referenced in [HMT Green Book](#) (page 72) in 2012 prices. Uplifted to 2019 prices, giving a value of £71,385.

<sup>137</sup> This represents the estimated impact of a mild episode of a depressive disorder - see below footnote for further information.

<sup>138</sup> As used in [The Economic and Social Costs of Crime](#), Home Office, July 2018. The estimate for 'REDUCEQL' originally comes from '[Disability weights for the Global Burden of Disease 2013 study](#)', Lancet Global Health 2015, e712-23. The duration (DUR) (0.167 years or two months) is an average originally derived from Wasserman and Ellis (2007) '[Impact of crime on victims](#)'. Chapter 6 in National Victim Assistance Academy Track 1: Foundation-Level Training.

<sup>139</sup> [Annual Bullying Survey 2017 - Ditch the Label](#)

303. This can then be multiplied by the total number of cases to give an estimate of the personal cost (in terms of quality of life reduction) to the individual:

- **£640 (Unit cost) \* 777,177 (number of cyberbullying cases (ages 10-15)) = £497,065,000 per year**

### Cost to the NHS of treating depression

304. As outlined above, it is estimated that 37% of cyberbullying victims go on to suffer depression as a result based on Ditch the Label's Annual Bullying Survey<sup>140</sup>. This gives a central estimate of 338,000 children per year suffering from depression as a result of cyberbullying.

305. Currently, NHS digital research has found that only 1 in 4 children (25.2%) who report having mental health problems access specialist mental health services<sup>141</sup>. It is assumed this is also the proportion of cyberbullied children who have developed depression that access mental health services. The National Institute for Health and Care Excellence (NICE) estimates the following costs for the treatment of depression<sup>142</sup>:

- A referral for psychological treatment: £14.50;
- Of the referrals, 67% accept the psychological treatment;
- 60% of these are low-intensity interventions at a cost of £45; and
- 40% are high-intensity at a cost of £1,125.

306. This gives an average cost from referral through to treatment for all patients (including those who are referred but don't subsequently take up full treatment) of £334.09 per person for a single treatment:

- Once uplifted to 2019 prices using a GDP deflator, this is £354.37 per patient on average; and
- This IA also assumes each individual accesses an intervention once per year - it is quite likely that a proportion of those seeking treatment may be treated multiple times and so this particular assumption is conservative.

307. Assuming those children who have suffered depression due to cyberbullying access care in similar proportions to all children with mental health problems (25.2%), this would give an annual cost of cyberbullying to health services of **£25.7 million**.

### Cost to NHS of treating cyberbullying-related self-harm

308. The 2017 Annual Bullying Survey found that 25% of cyberbullying victims surveyed went on to self-harm. This implies that 194,294 children per year self-harm as a result of cyberbullying. A large proportion of self-harm incidents will go unnoticed or treated (there is a three-fold difference in prevalence of self-harm as reported by young people and by their parents, suggesting that many acts of self-harm in the young do not come to the attention of their families). As such, information on many of these children formally seek help or attend hospital as a result is uncertain.

309. Based on a study in the Lancet<sup>143</sup>, the average cost to UK hospitals of treatment of self-harm is £809 per incident. Uplifting this to 2019 prices yields a cost per incident of £838. Given the uncertainty above, this IA assumes a conservative proportion of those self-harming due to

---

<sup>140</sup> [Annual Bullying Survey 2017 - Ditch the Label](#)

<sup>141</sup> [Mental Health of Children and Young People in England, 2017- NHS Digital \(2018\)](#)

<sup>142</sup> [Resource impact statement: Depression and anxiety disorder - National Institute for Health and Care Excellence \(2015\)](#)

<sup>143</sup> [General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis - Tsiachristas et al \(2017\)](#)

cyberbullying require hospital treatment (1% or 1,943 cases), this would result in an annual cost to the NHS of £1,629,000.

310. Given the difficulty in ascertaining exactly how many of those who self-harm due to cyberbullying would go on to require NHS treatment, this cost is only included as an illustrative upper estimate of cyberbullying, and is not included within the total estimated cost in the table above.

## Lifelong impact

311. The estimate for the lifelong impact draws upon a 2017 study which explores the long-term economic impact associated with childhood bullying.<sup>144</sup> These include mental health costs and employment-related costs (due to being unemployed or economically inactive). This study finds that the impact to society of childhood bullying is £90 per year per bullied woman and £271 per bullied man.<sup>145</sup> Based on the assumption that these costs start to apply at age 16 and apply for the duration of a person's working life, a net present social value (NPSV) was computed using a social discount rate of 3.5%. This resulted in lifetime costs of £2,335 and £7,031 for women and men respectively.

312. Note that the study also found former bullying victims had, at age 50, lower weekly earnings than peers who were not bullied, and were less likely to own a property or to have significant savings. These results suggest that the actual costs to society (for example, demand for social care, tax revenues foregone) could be substantially higher.

313. The prevalence rates (17% as in other cyberbullying estimates) were then applied to the population of 15-year old boys and girls in the UK (ONS data<sup>146</sup>). This gives an estimate of the 'cyberbullied' population of 15-year olds. The estimated lifetime costs for women and men were then applied to this population to estimate the total annual cost (each year, the new cohort of 15-year olds will also accrue similar lifetime costs, assuming cyberbullying prevalence remains constant). The total cost per year for each new cohort of 15 year olds in terms of the lifelong impacts of being cyberbullied is £184,302,000.

314. Given that this estimate is based on one academic study, and is therefore a more novel and less well-tested approach, it is only included as an illustrative upper estimate of cyberbullying, and is not included within the total estimated cost in the table above.

## Intimidation of public figures

315. Figures in the public eye, such as MPs, campaigners, and judges, frequently receive online abuse and threats. This is not only harmful to the individual concerned - it may sway them into making decisions against their better judgement. The fear of abuse and threats may also dissuade citizens (and certain groups in particular) from entering public life, for example by standing for election.

316. This analysis focuses specifically on the impact on MPs due to the availability of data and pre-existing research. Other figures in the public eye also receive online abuse and threats, but it is not possible to quantify this with any certainty.

Table 39: Annual impact of online intimidation of public figures

---

<sup>144</sup> [Long Term Economic Impact Associated with Childhood Bullying Victimization - Brimblecombe et al. \(2018\)](#)

<sup>145</sup> The societal costs are higher for men as there were found to be statistically significant employment-related costs for men but not for women. The costs to society affecting women were in mental health services, and this effect was not found to be statistically significant for men. See the above referenced paper for further details.

<sup>146</sup> [Population Estimates - ONS \(2019\)](#)



	Category	Prevalence	Unit cost	Annual cost
Intimidation	Cost of MPs' security measures	N/A	N/A	£4.2m
	Cost in police time investigating online threats	Over 6,000 threatening tweets sent to MPs	£37	£0.24m
	Impact on MPs and candidates' mental health	132 MPs (55% of survey respondents) experienced behaviour that made them fearful <sup>5</sup>	£1,333	£0.18m
	Impact on diversity of parliamentary candidates and any elected or public office official	N/A	Not quantified	Not quantified
	Impact on democratic/legal processes	N/A	Not quantified	Not quantified
	Impact on other public figures	N/A	Not quantified	Not quantified
<b>Total (quantifiable) £4.6m</b>				

317. Online threats increase the cost of security measures for MPs. According to the Institute for Government (IFG), £4.2m was spent on security measures for MPs in 2017/18, up from £171k in 2015/16.<sup>147</sup> The Committee on Standards in Public Life believes that *“the widespread use of social media has been the most significant factor accelerating and enabling intimidatory behaviour in recent years”*, as it creates *“an intensely hostile online environment,”* making it likely to be a key driver in this expenditure. The expenditure excludes the cost of police protection, which is kept confidential for security reasons.

318. According to research from Demos<sup>148</sup> (based on data from 2016) an average of 16,500 relevant tweets (English language, geo-located to UK, not containing links to external sites, and not duplicated) a day are sent to MPs on Twitter. Research from the University of Central Lancashire indicates that threatening tweets comprise around 0.1% of all tweets sent to MPs (equating to 16.5 threatening tweets per day, 6,022 per year). We have assumed that investigating and responding to a threatening tweet takes on average, one hour of police time (£37).

319. Surveys of MPs have found that many MPs feel fearful.<sup>149</sup> Surveys generally ask about intrusive and aggressive behaviour and threats in general so this may overestimate the impact of online abuse. However, it is also assumed that non-responders to the survey experienced no harm

<sup>147</sup> [Parliamentary Monitor Snapshot - IFG \(2019\).](#)

<sup>148</sup> [Can Technology Provide a Window into the New World of Digital Politics in the UK? - Demos \(2017\)](#) - see page 19. An average of 36,000 tweets are sent a day, which once accounting for duplicates is reduced to 16,500.

<sup>149</sup> [For example, The Personal Security of Individuals in British Public Life - Demos \(2018\)](#)

whatsoever. It is assumed that ‘fearful’ equates to a moderate anxiety disorder, which has a QALY (quality-adjusted life year) impact of 0.133 years.<sup>150</sup>

320. As discussed previously, there are other impacts which are not quantifiable but have potentially high costs. It is possible that online abuse may distort democratic and legal processes. Research from Amnesty International<sup>151</sup> found that women – particularly black, Asian and minority ethnic women – experience more targeted abuse (such as gendered insults and greater incidence of threats).

## Break-even analysis

321. As outlined in the above sections, this IA quantifies the annual social cost - under baseline - of a subset of online harms, including both illegal and legal but harmful harms. Harms are assumed to grow at 5% per year (in line with internet use) and the table below outlines the estimated cost in the first year of the appraisal period.

Table 40: Annual social cost of online harms in the first year of the appraisal period (2023)

Harm	Annual cost to society*
Child sexual abuse and exploitation	£2,052.4m
Hate crime	£13.3m
Illegal sale of drugs	£16.3m
Modern slavery	£37.5m
Cyber stalking	£2,385.8m
Cyber bullying	£573.0m
Intimidation of public figures	£5.0m
<b>Total annual</b>	<b>£5,083.4m</b>
<b>Total across the appraisal period (10-year PV)</b>	<b>£54,280.7m</b>

\* Please note these cost figures do not align exactly with the analysis above of the individual harms. Individual harm analysis was assessed using 2019 data (and therefore 2019 prevalence where available). These have been uplifted in line with the expected growth in online harms to 2023 - the first year of the appraisal period and converted to 2019 prices and 2020 present value base year.

322. These estimates are likely to underestimate the full extent of online harms for several reasons.

- It has only been possible to quantify the cost of a subset of all online harms in scope: there are a number of harms that are encountered by a significant number of adults and children in the UK, but for which there is no evidence on which to make an estimate of their cost. These include encouraging terrorism and radicalisation online, which 5% of adults and 6% of children in the UK have encountered, and encouraging self-harm, which 5% of adults and 10% of children have encountered.<sup>152</sup>
- For those harms that have been quantified, a conservative approach has been undertaken. For example, for illegal harms analysis is based on the number of recorded offences with an online element, which is likely to understate the true prevalence (as some crimes will go unreported - although this is adjusted in part by the use of multipliers where appropriate).

<sup>150</sup> [The Economic and Social Costs of Domestic Abuse - Home Office \(2019\)](#)

<sup>151</sup> [Black and Asian Women MPs Abuse More Online - Amnesty International UK \(2017\)](#)

<sup>152</sup> [‘Internet users’ concerns about and experience of potential online harms’ - OFCOM and ICO \(2019\)](#)



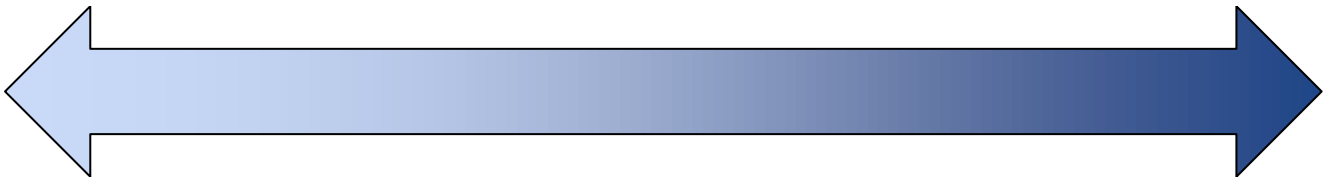
- Crimes may feature an online element but not be flagged as online: currently, whether a crime is recorded as having an online element is reliant upon police recording practices and how police forces apply the online flag. This, again, will reduce the reported prevalence of a given harm, and lead to an underestimate of its cost.

323. Given the difficulty in providing an evidenced estimate for a percentage reduction in online harms resulting from this proposal, the benefits remain purely illustrative and are not considered in the calculation of the NPSV.

324. The illustrative benefit of each policy option is the value of a reduction in online harms. Given the data limitations described above, this IA has only been able to quantify estimated benefits for a reduction in the subset of online harms outlined above. It is assumed that, once enacted, a policy will start to reduce online harms in the second year of the appraisal period.<sup>153</sup>

325. Some options will be more effective in reducing online harms than others. The table below sets out a qualitative analysis of the relative effectiveness of each of the options in reducing harms.

Table 41: Efficacy of policy options in reducing online harms

Smallest reduction in online harms expected		Largest reduction in online harms expected
		
<p><b>Option 1</b></p> <p>Only businesses likely to be accessed by children will be required to address legal but harmful content</p>	<p><b>Option 2</b></p> <p>In addition to the requirements under <b>Option 1</b>, the preferred option places an additional duty on high risk businesses to address legal but harmful content and activity accessed by adults.</p> <p>The majority of high risk platforms are larger organisations who are best-placed (due to resources and technological capability) to tackle harms. The additional requirement on these organisations is expected to result in greater harm reduction than under <b>Option 1</b></p>	<p><b>Option 3</b></p> <p><b>Option 3</b> is likely to result in the greatest reduction in online harms due to the requirement on all organisations to address legal but harmful content and activity accessed by adults.</p> <p>However, evidence is limited on the actual distribution of online harms, given that risk categories are based on number of features and functionality of a service, and the majority of harms occur on high-risk platforms - expanding the requirement to address legal but harmful content to all organisations therefore, would only have a marginal effect on the reduction in harms. For this reason, <b>Option 3</b> is not expected to result in significantly higher benefits than <b>Option 2</b> but it will result in significantly more burdensome costs to business.</p>

<sup>153</sup> The reduction is relative to the estimates of harms under BAU and is not applied cumulatively. It is also noted that the year in which reductions would start will depend on the year in which regulation is enacted.

326. Although in principle, **Option 3** will reduce harms to a greater extent than the preferred option, this is not expected to be a large enough reduction to account for the additional burden placed on businesses and the regulator under **Option 3**.

327. Evidence on the likelihood of benefits occurring is limited. Similar regulations abroad are either planned and not yet implemented or have not been fully assessed, as is the case for the German NetzDG. Additionally, it is difficult to highlight specific incidences of harm that have occurred in the past but would not have done so had the OSB been in place. This is due to the complex nature of online harms, especially in relation to how they lead to realised impact. For example, hate speech aimed at an individual impacts both the direct victim but also other users who may see it. The level of harm mitigation achieved from user safety measures will depend on the type of harm and the point at which it is addressed, this makes it difficult to determine the precise likelihood of a reduction in online harms resulting from platforms’ responses to the OSB. However, the Bill is expected to lead to a reduction in online harms compared to a *do nothing* baseline through the following mechanisms (this is not an exhaustive list):

Table 42: Qualitative assessment of why the OSB is expected to result in reduced harms

Outcome	Harm reduction
Content moderation	In 2020 for example, Facebook took action on 35.9 million pieces of content relating to child nudity and sexual exploitation of children, around 99% of which was found and flagged before users reported it <sup>154</sup> . This highlights how systems and processes to moderate content can mitigate the impact of online harms. The Bill is expected to lead to some platforms conducting additional content moderation to address online harms. This could be through bolstering existing content moderation processes or implementing new ones for platforms that do not currently moderate content.
User reporting	In 2020, nearly 1.4 million YouTube videos were removed as a result of user reporting mechanisms on the platform <sup>155156</sup> . This means that nearly 1.4 million potentially harmful videos (or videos that did not comply with YouTube’s community guidelines) were removed from the platform which is likely to have mitigated their impact. Under the Bill, platforms will be expected to accommodate user reporting of harms and therefore, some platforms without these systems will be required to implement them and those that do have them may be required to make improvements.
Transparency and user behaviour	Category 1 services will be expected to publish transparency reports under the Bill and OFCOM will have a range of information gathering powers as well as a responsibility to conduct research into online harms. In addition, alongside the Bill the government is undertaking a number of projects and initiatives aimed at improving media literacy. Giving users more information about the risks and prevalence of online harms on platforms and the government’s initiatives related to media literacy are both expected to increase user safety online

<sup>154</sup> [Community Standards Enforcement Report - Facebook \(2021\)](#)

<sup>155</sup> User reporting was the first source of detection

<sup>156</sup> [YouTube Community Guidelines Enforcement - YouTube \(2020\)](#)

Outcome	Harm reduction
	and mitigate some of the impacts associated with online harms.
Risk assessments	The Bill requires platforms to undertake risk assessments to assess risks corresponding to the type of content and activity a business is required to address. Many platforms already conduct risk assessments; however, there will be some that do not and these assessments could result in more or better targeted content moderation leading to a more efficient allocation of resources and greater harm mitigation.

328. Given the uncertainty around the reduction in online harms that could be achieved under each option (as described above), this IA estimates the reduction in the sub set of quantified online harms required to exactly match the costs of each policy option, that is, the scale of the reduction of harm required to deliver a benefit-cost ratio of precisely 1. The results are shown in the table below.

Table 43: Break even analysis

	Option 0	Option 1	Option 2	Option 3
Average annual reduction in harms needed	-	3.1%	3.9%	13.5%

329. To further inform the analysis, this section considers how the benefit-cost ratio would change if different illustrative assumptions were made about the effectiveness of the proposed policy measures in reducing harms.

Table 44: Illustrative scenarios

	Option 1	Option 2	Option 3
Low reduction scenario	1%	2.5%	4%
Mid reduction scenario	3%	5%	7%
High reduction scenario	5%	7.5%	9%

330. Based on these scenarios, the table below compares the costs and benefits of each policy option.

Table 45: Benefit cost ratios (BCR) under illustrative scenarios

	Option 1	Option 2	Option 3
<b><i>Low reduction scenario</i></b>			
Benefits	£436m	£1,230m	£1,967m
Implied BCR	0.3	0.6	0.3
<b><i>Mid reduction scenario</i></b>			

	Option 1	Option 2	Option 3
Benefits	£1,308m	£2,459m	£3,443m
Implied BCR	0.8	1.2	0.5
<b>High reduction scenario</b>			
Benefits	£2,180m	£3,689m	£4,427m
Implied BCR	1.3	1.7	0.6

## Indirect Costs and Benefits

331. This IA is not able to monetise any of the expected indirect impacts resulting from the policy options. While all impacts depend on future codes of practice and the actions of OFCOM, indirect impacts are even more reliant on these two unknowns. Where possible, this section provides a qualitative discussion of the impacts under the preferred option. Unless stated otherwise in this section, it is assumed that indirect impacts are equal across the policy options (given that these are not monetised this does not impact on the quantitative options comparison).

## Freedom of expression

332. If the regulatory framework (under all options) was to result in overblocking by platforms this could have an impact on freedom of expression. However, the chosen model under the preferred option has built in appropriate safeguards to ensure protections of freedom of expression. The regulator will have an obligation to protect freedom of expression, for which it can be held to account. Additionally:

- To protect freedom of expression the regulation will treat illegal and legal but harmful content for in-scope services differently.
- Journalistic and editorial content on news publishers' own sites will be excluded from scope as well as 'below the line' comments on directly published media content. In addition, news publisher content shared on in-scope services will be exempt.
- Category 1 providers will also provide additional protections for journalistic and democratic content. Providers will have a duty to set policies for protecting such content which they must enforce consistently and transparently.
- Effective transparency reporting will help ensure content removal is well-founded, as the decisions platforms make on content removal and user appeals on content removal will have greater visibility.
- Escalating enforcement sanctions will avoid incentivising content takedown, with judicial oversight to safeguard the most severe sanctions like access restriction.
- User redress mechanisms will enable users to challenge content that restricts their freedom of expression and to more effectively appeal content removal.
- Super-complaints will allow organisations to lodge concerns on behalf of users, which can include concerns about limits on freedom of expression.

333. Under the status quo, businesses already remove content according to their terms of service, and where they assess content is likely to be illegal. This process lacks transparency and stakeholders have complained that content may be removed or its visibility reduced with little explanation and it is very hard to get content reinstated. The regulatory model's focus on transparency and user reporting and redress should therefore lead to some improvements in users' ability to appeal content removal and get this reinstated, with a positive impact on freedom of expression.

## Privacy impacts

334. As set out in the full government response to the OHWP, the regulatory framework will apply to public communication channels and services where users expect a greater degree of privacy - for example online instant messaging services and closed social media groups.
335. The regulator will set out how businesses can fulfil their duty of care in codes of practice, including what measures are likely to be appropriate in the context of private communications. This could include steps to make services safer by design, such as limiting the ability for anonymous adults to contact children. Businesses in scope will need to consider the impact on users' privacy and ensure users understand how business systems and processes affect user privacy.
336. Given the severity of the threat, the legislation will also enable OFCOM to require businesses to use technology that is highly accurate to identify and remove tightly defined categories of illegal material relating to child sexual exploitation and abuse on public and, where proportionate, private channels.
337. Recognising the potential risk of an impact to users' privacy, the government will ensure this is only used as a last resort where alternative measures are not working and will be subject to stringent safeguards to protect users' rights.
338. The regulator will advise the government on the accuracy of tools and make operational decisions regarding whether or not a specific business should be required to use them. Before the regulator can use these powers it will need to seek approval from Ministers on the basis that sufficiently accurate tools exist. The regulator will also be able to require businesses to use highly accurate technology to identify illegal terrorist content, also subject to stringent safeguards but on public channels only.

**Consultation question 14:** Are there any additional indirect costs or benefits not discussed in this IA? The government welcomes any evidence you can provide.

## Summary of impacts

339. All impacts are assessed over a ten year appraisal period starting from the date of implementation (expected to be 2023). For present value costs and benefits, a discount rate of 3.5% has been applied in line with Green Book guidance. All costs are presented in 2019 prices with 2020 as the present value base year.
340. As outlined in the first section, given the uncertainties around future codes of practice, the EANDCB and NPSV are illustrative only at this stage and do not include all impacts. However, this IA attempts to calculate these metrics based on the government's best-estimate of the likely business requirements. The table below summarises the estimated impacts.

Table 46: Summary of impacts (10 year PV)

Impact	Option 1	Option 2	Option 3
Reading and understanding the regulations (cost to business)	£9.2m	£9.2m	£9.2m
Ensuring a user reporting	£12.4m	£12.4m	£33.2m

Impact	Option 1	Option 2	Option 3
mechanism is in place and updated (cost to business)			
Updates to terms of service (cost to business)	£14.7m	£14.7m	£14.7m
Producing a risk assessment (cost to business)	£31.0m	£31.0m	£31.0m
Content moderation (cost to business)	£1,271.5m	£1,700.2m	£5,999.3m
Transparency reporting (cost to business)	£3.6m	£3.6m	£921.0m
Requirement to report online CSEA (cost to business)	Not monetised	Not monetised	Not monetised
Industry fee (cost to business)	£346.7m	£346.7m	£346.7m
Industry fines (cost to business)	Not monetised	Not monetised	Not monetised
Business disruption measures (cost to business)	Not monetised	Not monetised	Not monetised
Justice impacts (cost to government)	£0.4m	£0.4m	£0.4m
Freedom of expression & privacy implications (cost to society)	Not monetised	Not monetised	Not monetised
Innovation and competition impacts (cost to society)	Not monetised	Not monetised	Not monetised
Reduced prevalence of online harms (benefit to society)	Break even - 3.1% (average annual reduction in harms required)	Break even - 3.9% (average annual reduction in harms required)	Break even - 13.5% (average annual reduction in harms required)

## Business Impact Target Calculations

341. It is not possible to predict with certainty the actions of OFCOM or the steps businesses may take to ensure they are compliant with the regulation at this stage. While this Bill sets out a duty of care for businesses, the specific requirements on businesses and the actions they can take to comply will be set out in codes of practice laid by the independent regulator. The content of the codes of practice are unknown at this stage and it will be for OFCOM to determine the content. The regulatory burden placed on each individual business will be proportionate to a business's size and risk and the impact on business will be carefully considered as part of the code development process.

342. OFCOM currently carries out IAs across all of its functions including broadcasting, radio, telecoms and post under Section 7 of the Communications Act. At present, OFCOM is required to carry out and publish IAs where they consider a proposal to be important. OFCOM has discretion as to the form and matters covered by an IA. They may alternatively publish a statement setting out their reasons for why carrying out an IA is unnecessary.

343. The OH Bill extends OFCOM’s existing duties under Section 7 of the Communications Act to all “important proposals” related to online harms and there will, therefore, be a duty on OFCOM to produce an IA for and consult on all new codes of practice and for revisions to existing codes of practice, unless the change was considered negligible whereby OFCOM would be able to publish a statement setting out why an IA was not necessary.

344. Furthermore, OFCOM’s current legislative requirements do not specifically require any assessment of the impact of proposals on small and micro businesses. Provisions under Section 7 of the Communications Act and the SBEE Act 2015 may not be sufficient to guarantee that economic impacts on small and micro businesses are taken into account. This Bill, therefore, will require OFCOM to explicitly consider the impacts on small and micro businesses in its appraisal of all important proposals. This will reduce the risk of online safety regulations disproportionately affecting small and micro businesses and ensure that a risk-based and proportionate approach is maintained throughout the online safety framework.

345. Given that specific business requirements are unknown at this stage, the EANDCB calculated here is largely illustrative and aims to indicate the potential scale or nature of impacts of the whole policy (scenario 2 in the RPC’s primary legislation guidance<sup>157</sup>).

## Calculations

346. Under requirements set out in the Better Regulation Framework, this IA calculates an illustrative overarching EANDCB covering the whole policy, including best estimates for requirements resulting from future codes of practice. The illustrative EANDCB includes all monetised direct costs to business.

Table 47: Calculation of illustrative whole policy EANDCB under the preferred option

	Option 1	Option 2	Option 3
<b>Net Present Social Value (NPSV)</b>	-£1,689m	-£2,118m	-£7,355m
<b>Equivalent annual net direct cost to business (EANDCB)</b>	£156m	£206m	£814m

347. The NPSV for the Bill (and future codes of practice) under the preferred option is estimated to be **-£2,118 million** with an EANDCB of **£206 million**. This EANDCB is illustrative only and is based on our best estimate of likely business requirements stemming from future codes of practice. It will be for OFCOM to determine specific requirements and is required in the legislation to conduct consultations and produce IAs.

348. The government hopes to strengthen these estimates through information provided by businesses (and wider society) as part of the pre-legislative scrutiny process.

<sup>157</sup> [RPC Case Histories: assessment and scoring of primary legislation measures \(2019\)](#)

# Risks and Assumptions

## Policy Risks

349. The following section outlines the core risks identified with the preferred option as a whole with regards to its impacts on both individuals and businesses. Alongside each of the identified policy risks is an outline of the necessary mitigations built into the policy to help address these risks.

Table 48: Strategic risks of the policy programme and mitigations

Risk	Mitigation
<p>Regulation disproportionately impacts on freedom of expression, by incentivising or requiring content takedown.</p>	<p>The approach has built in appropriate safeguards to ensure protections for freedom of expression, including:</p> <ul style="list-style-type: none"> <li>• Differentiated approach of legal/illegal content, e.g. not requiring takedown of legal but harmful content</li> <li>• Safeguards for journalistic content</li> <li>• Effective transparency reporting</li> <li>• Proportionate enforcement sanctions to avoid incentivising takedowns</li> <li>• User redress mechanisms will enable challenge to takedown</li> <li>• Super-complaints will allow organisations to lodge complaints where they may be concerned about disproportionate impacts</li> <li>• Regulator has a duty to consider freedom of expression</li> </ul>
<p>Non-UK based businesses are impacted less than UK-based ones. Risk to competition due to uneven enforcement.</p>	<p>To ensure the effective implementation of the regime, the regulator will have a suite of robust enforcement powers, to be able to take appropriate proportionate action against organisations that fail to fulfil the duty of care. These are in line with existing regulatory enforcement powers and will apply to all in-scope businesses. The regulator will be able to take enforcement action against any in-scope business worldwide that provides services to UK users. This will help to ensure a level playing field between businesses that have a legal presence in the UK, and those who operate entirely from overseas.</p> <p>Powers include:  Warnings and notice of non-compliance  Fines up to £18 million or 10% of annual global turnover, whichever is higher.  Business disruption measures ( including access restriction for egregious breaches) which we consider could be more easily leveraged against non-UK based organisations.</p> <p>Incentives to comply include:  - Avoiding sanctions for failure to comply.  - Maintaining reputation as good corporate actors and meeting public appetite for better online safety measures.</p>
<p>Increase in regulation make the UK a less attractive place to set up and run a digital business</p>	<p>A clear, proportionate and effective regulatory framework for online harms will ensure the UK remains an attractive place for digital businesses. The regime will provide legal clarity and predictability, which are important for businesses.</p>



Risk	Mitigation
	<p>As other countries worldwide are also introducing legislation in this space, e.g. Germany and the EU, there will not be a marked difference in operating costs between similar jurisdictions. While all businesses in-scope may incur incremental costs, the risk-based proportionate approach seeks to maintain UK competitiveness while delivering a higher level of online safety.</p> <p>The majority of in-scope businesses will only be required to respond to illegal content and put in place measures to protect children (including from online content/activity which may be legal for adults, e.g. pornographic). Partial exemptions (applied to online product and service reviews as well as ‘below the line’ comments) will mitigate the impact on many SMBs. The regulator will also carry out IAs before issuing codes of practice.</p> <p>The programme is carefully designed to encourage compliance and minimise its cost. The business support measures will be critical to providing businesses with the clarity they need to meet the requirements of the regulation:</p> <ul style="list-style-type: none"> <li>- Safety by design guidance to set out what ‘good’ online safety looks like, including child protection measures</li> <li>- Promotion of a vibrant safety tech market that can deliver competitively priced technology solutions for businesses</li> </ul> <p>The full list of mitigations can be found in the SaMBA.</p>
<p>Possibility that businesses will pass on the costs of regulation to its users</p>	<p>Although there is potential for businesses to pass the costs incurred through regulation on to their users, this is expected to be unlikely given this has not been done for NetzDG. Also, this has not been reported by businesses likely to incur the greatest costs. These businesses are already shifting towards greater user safety, a cost which is currently not being passed on to users.</p>
<p>Risk of non-compliance by businesses that are ideologically opposed to regulation in this area</p>	<p>OFCOM has a strong track record of engagement. Its annual report<sup>158</sup> details how it seeks to understand consumers’ and citizens’ interests and behaviours, and how it engages with industry and government. The regulator’s information gathering powers will play a crucial role in supporting its various regulatory functions. These powers will help the regulator build an in-depth understanding of the online harms landscape, prioritise its activity and oversee companies’ compliance with the regulatory framework. The approach to enforcement will aim to encourage compliance and drive positive cultural change. The regulator will support businesses to help them understand the expectations placed on them, and how the regulator’s use of its enforcement powers will be proportionate. OFCOM will have a suite of enforcement powers available to use against companies who fail to fulfil the duty of care, or fail to put in place appropriate measures after being alerted to an issue. These powers will be comparable to those already used by OFCOM and other UK regulators. OFCOM will use its enforcement powers in line with its duties, including being proportionate, taking into account the level of harm and considering the impact on children.</p>
<p>Possibility that harms could be</p>	<p>The regulation sets a broad scope for regulations and from that</p>

<sup>158</sup> [The Office of Communications Annual Report and Accounts \(19/20\)](#)

Risk	Mitigation
displaced onto out-of-scope platforms	basis then includes service specific exemptions and a low risk functionality exemption. The likelihood of harms occurring on platforms that are exempted under the low risk functionality exemption is low to begin with given the link between harmful content and functionality that facilitates it. The risk of displacement to platforms currently exempted is therefore expected to be minimal. However, should harmful activity increase on such platforms the Secretary of State would retain the right to rescind the exemption, although this action would be subject to the affirmative procedure in order to provide democratic accountability to an expansion in online harms scope.

## Analytical risks and assumptions

350. The below table sets out the analytical risk and assumptions, alongside the evidence to support each assumption and how these risks have been addressed through further sensitivity analysis. The table splits these risks out into core categories of scope, transition costs and compliance costs. Full sensitivity analysis is then included in a further section below, addressing each identified risk.

351. It should be noted across the below table that the estimation of costs (both in terms of cost per organisation and the number of organisations incurring those costs) is indicative, as it will be the role of the independent regulator to determine exactly how regulation is specified.

Table 49: Assumptions and Risks in Analytical Approach

Assumption	Evidence	Risk
<b>Scope</b>		
Number of businesses in scope of the regulation	Based on a random stratified sample of 500 businesses from the IDBR. <sup>159</sup>	Central estimate has been produced but there remains a degree of uncertainty around the estimate
A breakdown of businesses in scope by size of business	Based on an IDBR random sample divided into 5 strata (100 businesses in each strata) with each strata accounting for different sizes of business (see 'number of businesses in scope' section).	Whilst the sample size is sufficiently large to give a robust estimate of the number of in-scope businesses, given the breadth of businesses covered in the IDBR, certain types of business within certain categories may not be captured within the sample which could affect final results.
The number of civil society organisations in scope of regulation	Civil Society organisations are covered by the IDBR, though this coverage is not fully complete (88,240 non-profit (or mutual	A number of potentially in-scope civil society organisations may not be captured by the analysed

<sup>159</sup> Inter-Departmental Business Register - 2019

Assumption	Evidence	Risk
	association) organisations) <sup>160</sup> , whilst there are 166,592 voluntary organisation in total in the UK <sup>161</sup> .	sample.
<b>Transition costs</b>		
Number of businesses out of scope of the regulation but that will undertake some degree of familiarisation costs to establish this (we have conservatively assumed 25% of businesses potentially in scope would incur costs).	Based on an IDBR sample of 500 - there are 180,000 businesses that enable either: P2P interaction, i.e. allowing users to interact with other users in any way; or access to UGC.	There is a degree of uncertainty to exactly how many businesses would undertake initial familiarisation with the regulation due to the potential for them to be in scope.
Percentage of staff in medium and large businesses that would spend time familiarising themselves with the regulation during dissemination of the information (current estimate 10%)	This is a reasonable assumption based on a relatively large segment of the business needing to know, e.g. staff involved in user safety. From interviews with businesses, this is broadly in line with the percentage of employees working in this area within each business (although this does vary).	It may be the case that a smaller or larger percentage of staff take time to familiarise themselves with the regulations. Sensitivity analysis is conducted on this.
Cost of implementing a user reporting mechanism. This IA assumes varying amounts of programmer time to update these to reflect future codes of practice.	Interviews with businesses indicated that all businesses that had significant amounts of UGC and P2P interaction already had reporting mechanisms. This cost is likely to be minimal and broadly in line with the estimate presented here.	For businesses in scope which do not currently have user reporting mechanisms (and would be required to have them under the framework) would likely incur greater costs than estimated here.
Cost of revising terms of service (assumed varying amounts of staff time, including legal advice for larger businesses)	This is based on the average length of services terms of service.	It is possible, especially for large high risk businesses that they may require greater legal input than this IA assumes.
<b>Compliance costs</b>		
Cost per business of producing a risk assessment	Based on cost of risk assessments as outlined on NIS Regulation 2018 <sup>162</sup>	These costs are proxies and therefore, whilst likely to be similar, are not directly applicable.

<sup>160</sup> Inter-Departmental Business Register - 2019

<sup>161</sup> [UK Civil Society Almanac 2020 - NVCO](#)

<sup>162</sup> [The Network and Information Systems Regulation 2018, Impact Assessment](#)

Assumption	Evidence	Risk
<p>Cost per business of additional content moderation is 7.5% of turnover on average per year for Category 1 businesses and 1.9% for Category 2 businesses</p>	<p>Based on the midpoint of estimates provided by in-scope businesses during the interview phase of RR research project. For Category 2 businesses, this is based on an assessment of the ratio of illegal to harmful content actioned by some of the largest social media businesses.</p>	<p>Estimates varied greatly with some small low risk businesses saying they were ‘negligible’ and other larger businesses providing figures of 7-10% of revenue or costs.</p>
<p>Proportion of mid and high risk businesses requiring additional content moderation (25% of high risk in-scope businesses and 10% of medium/large mid risk)</p>	<p>Based on interviews with a strategic sample of 30 in-scope businesses (out of 118 contacted).</p>	<p>Costs provided are broad averages based on risk tier and size, and are based on responses from a subset of 30 of the strategic sample of 118 businesses (25%).</p>
<p>Cost of transparency reporting per business</p>	<p>Based on NetzDG transparency reporting costs (£45,000)</p> <p>Other figures used in previous IAs with a transparency reporting requirement have been used to sense check.</p>	<p>Proxy figures in NetzDG used are not directly applicable to Online Harms transparency requirement and so only provide an initial indicative estimate. These could represent a conservative estimate given the German language demands on NetzDG transparency reports. Conversely, NetzDG only covers large businesses, so for <b>Option 3</b>, the estimates were revised down for smaller organisations. It may be the case that the cost would be equal across business sizes and therefore the estimates would understate the true cost of this activity.</p>
<p><b>Illustrative benefits</b></p>		
<p>Assumed reductions on online harm prevalence (illustrative scenarios)</p>	<p>While all policy options are expected to reduce the prevalence of online harms to varying degrees, evidence of the effectiveness of similar policies is limited (or non-existent). There has been no published government-run economic review of Germany’s NetzDG which is to some extent similar to some of the requirements under the online safety framework.</p>	<p>Benefit estimation in this IA is only illustrative so this will not affect the main metrics presented. However, this will affect the implied benefit cost ratios for the three options.</p>

Assumption	Evidence	Risk
Assumed growth rate of online harms across the appraisal period	This is in line with growth in the amount of hours spent online (an average of 5.4% per year over the last four years).	While Covid has increased internet use, working from home and communication on digital devices, the true extent of this or whether any change is long term is currently unknown. If under the baseline, harms grow at a slower rate across the appraisal period this IA will have overestimated the potential benefits of the policy. If they grow faster, the IA will have underestimated. Sensitivity has been conducted in the table below.
Prevalence of cyberbullying	Based on the average of a number of published studies on cyberbullying.	If the prevalence is in fact greater under the baseline, this IA will have underestimated the benefits of the policy. Likewise, if the prevalence is lower under the baseline, the IA will have overestimated. While this is true for all of the quantified harms, the majority are based on official crime data whereas for cyberbullying, a survey of academic literature has been used.
Prevalence of harms based on crime data	Based on experimental data where crimes are recorded as having an online element	The data does not provide information on the extent of the online component, that is, whether it was a significant or a minor part of the offence. It also does not provide information on whether, in the absence of the online component, the offence would still have taken place via alternative means. In addition, many of these harms can involve both online and offline elements, which are often closely linked (e.g. traditional bullying and cyberbullying). It can therefore be difficult to completely disaggregate the impacts.

**Consultation question 15:** Do you agree with the assumptions used in this assessment of the costs and benefits of the policy? The government welcomes any evidence you can provide to refine the assumptions.

# Sensitivity Analysis

352. Unlike scenario analysis, sensitivity is used to illustrate how sensitive the key metrics presented in this consultation stage IA are to any important assumptions made in the analysis. By varying only one assumption at a time and assessing the percentage change in the key metrics, this section is able to highlight the most important assumptions. The table below indicates that the most important assumptions (in terms of their effect on the key metrics) are the number of businesses in scope, the incremental cost of additional content moderation, and the percentage of in-scope organisations expected to require additional content moderation.

353. Two methodologies were considered when attempting to determine the number of businesses in scope. The first was a Standard Industrial Classification (SIC) code or ‘activity-based’ approach where whole groups of businesses were considered in scope due to the classification of their main economic activity, e.g. dating sites. The second - which has been used in this IA through research conducted by RR - was to manually assess the functionality of a stratified sample of businesses extrapolated to the UK business population. Given that whether an organisation is in scope of the regulations is determined solely by the functionality present on a service, it was determined that the stratified sample approach would be more reliable. For example, it is likely that some businesses would be in scope and some would not within the same SIC code. However, as noted in previous sections, once the stratified sample had been assessed the number of in-scope platforms were supplemented with known types of businesses. Given the overall difficulty in attaining a reliable estimate for the number of in-scope businesses, the IA presents sensitivity analysis below.

354. The central estimate for Category 1 organisations of the incremental cost of content moderation was the midpoint of the range provided directly by businesses in interview. The businesses were being asked to estimate these costs based on a broad interpretation of the OHWP and therefore these estimates cover the cost of addressing both illegal and harmful content. For Category 2 organisations (all other businesses expected to incur additional costs), it was assumed that the incremental cost of content moderation was proportionate to the volume of legal vs harmful content actioned by some of the largest social media businesses. While these estimates are inherently uncertain, they are reasonable assumptions and at this stage do provide an indication of the likely scale of impact of the regulations.

355. The percentage of in-scope businesses expected to incur any additional content moderation costs is based on research conducted by RR. Given that RR was able to interview 13 of the 16 largest social media businesses, we believe that the assumption of 25% of high risk businesses - while conservative - is reasonable. The percentage of mid risk businesses is more uncertain and we vary this extensively in the below sensitivity analysis.

Table 50: Overview of sensitivity analysis

<b>Number of businesses within scope of the regulations</b>		
Central estimate <b>24,311</b>	EANDCB <b>£205.8m</b>	NPSV <b>-£2,118m</b>
Lower <b>18,872</b>  <i>The lower bound represents the number of organisations identified by RR prior to supplementing with additional known types of organisations likely to be in scope. The lower bound also still includes the estimated number of CSOs in scope.</i>	EANDCB <b>£138.2m</b>  33% change	NPSV <b>-£1,536m</b>  27% change
Upper	EANDCB	NPSV

<b>135,327</b>	<b>£259.9m</b>	<b>-£2,583m</b>
<i>The upper bound represents the highest possible estimate accounting for RR's margins of error at the 95% confidence level. The risk categorisation for these businesses is based on the same methodology as under the central estimate. This estimate is the most conservative and highly unlikely.</i>	<i>26% change</i>	<i>22% change</i>
<b>Number of businesses expected to incur initial familiarisation</b>		
Central estimate <b>45,000</b>	EANDCB <b>£205.8m</b>	NPSV <b>-£2,118m</b>
Lower <b>24,311</b>	EANDCB <b>£205.5m</b>	NPSV <b>-£2,116m</b>
<i>Lower estimate assumes only in-scope businesses will incur the initial cost of familiarisation.</i>	<i>&lt;1% change</i>	<i>&lt;1% change</i>
Upper <b>180,000</b>	EANDCB <b>£207.2m</b>	NPSV <b>-£2,130m</b>
<i>180,000 businesses (all businesses that could potentially be considered in scope under a broad interpretation of the Online Harms White Paper) is the highest reasonable estimate.</i>	<i>&lt;1% change</i>	<i>&lt;1% change</i>
<b>Percentage of staff involved in dissemination of information (in medium and large)</b>		
Central estimate <b>10%</b>	EANDCB <b>£205.8m</b>	NPSV <b>-£2,118m</b>
Lower <b>5%</b>	EANDCB <b>£205.7m</b>	NPSV <b>-£2,117m</b>
	<i>&lt;1% change</i>	<i>&lt;1% change</i>
Upper <b>25%</b>	EANDCB <b>£206.1m</b>	NPSV <b>&lt;£2,120m</b>
	<i>&lt;1% change</i>	<i>&lt;1% change</i>
<b>Cost of user reporting mechanism</b>		
Central estimate <b>Differentiated by size and risk</b>	EANDCB <b>£205.8m</b>	NPSV <b>-£2,118m</b>
Lower <b>Differentiated only by size</b>	EANDCB <b>£205.5m</b>	NPSV <b>-£2,116m</b>
<i>For the lower bound, costs are assumed to only vary between business sizes and all businesses incur the costs estimated for low risk organisations under the preferred option. This reflects a scenario in which only minor changes are required to existing user reporting</i>	<i>&lt;1% change</i>	<i>&lt;1% change</i>

<i>mechanisms.</i>		
<p style="text-align: center;"><b>Upper Differentiated only by size</b></p> <p><i>For the upper bound, costs are assumed to vary between business sizes and all businesses incur the costs estimated for high risk organisations under the preferred option. This reflects a scenario in which more substantial changes are required to existing user reporting mechanisms.</i></p>	<p style="text-align: center;"><b>EANDCB £208.2m</b></p> <p style="text-align: center;"><i>1% change</i></p>	<p style="text-align: center;"><b>NPSV -£2,139m</b></p> <p style="text-align: center;"><i>&lt;1% change</i></p>
<b>Cost to Category 1 organisations of additional content moderation</b>		
<p style="text-align: center;"><b>Central estimate 7.5% of turnover</b></p>	<p style="text-align: center;"><b>EANDCB £205.8m</b></p>	<p style="text-align: center;"><b>NPSV -£2,118m</b></p>
<p style="text-align: center;"><b>Lower 1% of turnover</b></p> <p><i>The lower bound reflects the lowest estimate provided by businesses in interviews.</i></p>	<p style="text-align: center;"><b>EANDCB £148.2m</b></p> <p style="text-align: center;"><i>28% change</i></p>	<p style="text-align: center;"><b>NPSV -£1,623m</b></p> <p style="text-align: center;"><i>23% change</i></p>
<p style="text-align: center;"><b>Upper 15% of turnover</b></p> <p><i>The upper bound reflects the highest estimate provided by businesses in interviews.</i></p>	<p style="text-align: center;"><b>EANDCB £272.2m</b></p> <p style="text-align: center;"><i>32% change</i></p>	<p style="text-align: center;"><b>NPSV -£2,690m</b></p> <p style="text-align: center;"><i>27% change</i></p>
<b>Cost to Category 2 organisations of additional content moderation</b>		
<p style="text-align: center;"><b>Central estimate 1.9% of turnover</b></p>	<p style="text-align: center;"><b>EANDCB £205.8m</b></p>	<p style="text-align: center;"><b>NPSV -£2,118m</b></p>
<p style="text-align: center;"><b>Lower 0.25% of turnover</b></p> <p><i>The central estimate of 1.9% used above is 25% of the midpoint of estimates provided by businesses in interviews (7.5% of turnover). This reflects the proxied volume of illegal vs harmful content actioned by social media businesses (25% illegal content). This lower bound uses the same methodology but on the lowest estimate provided by businesses in interview, 1% of turnover.</i></p>	<p style="text-align: center;"><b>EANDCB £92.1m</b></p> <p style="text-align: center;"><i>55% change</i></p>	<p style="text-align: center;"><b>NPSV -£1,140m</b></p> <p style="text-align: center;"><i>46% change</i></p>
<p style="text-align: center;"><b>Upper 3.75% of turnover</b></p> <p><i>Using the same methodology as above, the upper bound is 25% of the highest estimate provided by businesses in interviews (15% of turnover).</i></p>	<p style="text-align: center;"><b>EANDCB £336.9m</b></p> <p style="text-align: center;"><i>64% change</i></p>	<p style="text-align: center;"><b>NPSV -£3,247m</b></p> <p style="text-align: center;"><i>53% change</i></p>
<b>Percentage of mid and high risk businesses requiring additional content moderation</b>		
<p style="text-align: center;"><b>Central estimate 10% - mid risk</b></p>	<p style="text-align: center;"><b>EANDCB £205.8m</b></p>	<p style="text-align: center;"><b>NPSV -£2,118m</b></p>



<p align="center"><b>&amp; 25% - high risk</b></p>		
<p align="center">Lower <b>5% - mid risk</b> <b>&amp;</b> <b>10% - high risk</b></p> <p><i>The lower bound reflects the potential that only a small number of businesses would incur additional content moderation. This would be the case if this IA underestimated the current level (and capacity) of content moderation under the baseline.</i></p>	<p align="center">EANDCB <b>£94.2m</b></p> <p align="center"><i>54% change</i></p>	<p align="center">NPSV <b>-£1,158m</b></p> <p align="center"><i>45% change</i></p>
<p align="center">Upper <b>25% - mid risk</b> <b>&amp;</b> <b>25% - high risk</b></p> <p><i>High risk businesses are relatively well represented in the research and interview sample and we believe that our assumption of 25% only requiring additional content moderation is the most conservative estimate while remaining reasonable. For the upper bound therefore, only the percentage of mid risk businesses expected to incur costs is varied.</i></p>	<p align="center">EANDCB <b>£310.5m</b></p> <p align="center"><i>51% change</i></p>	<p align="center">NPSV <b>-£3,020m</b></p> <p align="center"><i>43% change</i></p>
<p align="center"><b>Cost per high risk business of transparency reporting</b></p>		
<p align="center">Central estimate <b>£45,000</b></p>	<p align="center">EANDCB <b>£205.8m</b></p>	<p align="center">NPSV <b>-£2,118m</b></p>
<p align="center">Lower <b>£22,500</b></p> <p><i>The lower bound reflects the possibility that the cost of transparency reporting (currently proxied from Germany's Network Enforcement Act) is half as costly. This would be a reasonable assumption if for example, all businesses designated as Category 1 already produced transparency reports and were able to revise to reflect OFCOM's requirements easily.</i></p>	<p align="center">EANDCB <b>£205.5m</b></p> <p align="center"><i>&lt;1% change</i></p>	<p align="center">NPSV <b>-£2,116m</b></p> <p align="center"><i>&lt;1% change</i></p>
<p align="center">Upper <b>£90,000</b></p> <p><i>The upper bound reflects the possibility that the cost of transparency reporting (currently proxied from Germany's Network Enforcement Act) is twice as costly. This would be a reasonable assumption if for example, the requirements set by OFCOM were such that Category 1 organisations had to implement large changes to their data gathering and reporting processes.</i></p>	<p align="center">EANDCB <b>£206.2m</b></p> <p align="center"><i>&lt;1% change</i></p>	<p align="center">NPSV <b>-£2,122m</b></p> <p align="center"><i>&lt;1% change</i></p>
<p align="center"><b>Growth rate of online harms under the baseline</b></p>		
<p align="center">Central estimate</p>	<p align="center">Break even</p>	<p align="center">Benefits (5%)</p>

<b>5%</b>	<b>3.9%</b> <i>average annual reduction</i>	illustrative reduction compared to baseline) <b>£2,459m</b>
Lower <b>2%</b>	Break even <b>4.6%</b> <i>average annual reduction</i>	Benefits (5% illustrative reduction compared to baseline) <b>£2,067m</b>
Upper <b>7.5%</b>	Break even <b>3.4%</b> <i>average annual reduction</i>	Benefits (5% illustrative reduction compared to baseline) <b>£2,844m</b>

## Small and Micro Business Assessment

### Justification for non-exemption of SMBs under the preferred option

356. As explained in guidance from the RPC, the default position is to exempt SMBs fully from the requirements of new regulatory measures.<sup>163</sup> However, the evidence suggests that the objectives of the regulations would be compromised by exempting SMBs as business size is not always a good proxy for risk in the context of online harm. As described in further detail in the rest of this section, the policy cost of exempting SMBs is significant for three main reasons.

357. First, there is evidence of harms occurring on smaller platforms. In particular, law enforcement and NGOs regularly see child sexual exploitation and abuse offenders active on small chat forums, live streaming apps and file sharing/hosting services. The IWF notes that online harms exist ‘in vast quantities’ on smaller platforms.<sup>164</sup> 87% of the content the IWF removes from the internet is from small and medium size sites including file sharing sites, image hosting boards and cyberlockers. The inclusion of SMBs in the proposed regulation is therefore welcomed by the foundation.<sup>165</sup>

358. In addition, terrorist actors have sought to ‘exploit an overlapping ecosystem of services’, taking advantage of the fact that smaller businesses ‘don’t have the scale or resources to handle the challenge on their own’. The Tech against Terrorism project indicated that Daesh supporters use larger, well-known platforms (e.g. Twitter) to share links to smaller, less well-resourced platforms, where it is easier to exchange terrorist content.<sup>166</sup> Second, there is a limited relationship between the size of an organisation in terms of turnover and employees and the reach and impact of a given organisation. Therefore, traditional measures of size of organisation would not allow the correct identification of the platforms with the highest risk of harms.

<sup>163</sup> [Small and Micro Business Assessments: guidance for departments, with case history examples - RPC \(2019\).](#)

<sup>164</sup> [IWF Online Harms White Paper Response \(2021\)](#)

<sup>165</sup> Ibid.

<sup>166</sup> [UK launch of tech against terrorism at Chatham House - Tech Against Terrorism \(2017\).](#)

359. Third, given the fluidity of the online space, it would be possible for individuals to migrate from large to small platforms in a short time frame. This most likely to occur when public perceptions of the apps change quickly, such as in the event of a data breach by a business, a high profile trial, or action taken by the authorities.

## Impacts on SMBs

360. Based on **RR research**, it is estimated that around 21,000 SMBs will be within scope of the online safety framework. The in-scope SMBs are estimated to fall within the following risk categories:

Table 51: Estimated number of SMBs in each risk tier (rounded to the nearest ten)

	Low risk	Mid risk	High risk	Category 1
<b>Micro</b>	9,800	9,800	60	0
<b>Small</b>	560	560	60	0

361. The tables below outline the costs that SMBs are expected to incur as a result of the regulations (with medium and large businesses included for comparison):

Table 52: Per business transition cost to SMBs (first year)

	Low risk	Mid risk	High risk
<b>Micro</b>	£345	£368	£504
<b>Small</b>	£457	£502	£683
<b>Medium</b>	£922	£967	£1,193
<b>Large</b>	£3,023	£3,068	£3,340

362. As the table above illustrates, the largest per business transition costs are expected to fall on medium and large businesses who are better placed to absorb them.

Table 53: Annual per business cost to SMBs of compliance

	Low risk	Mid risk	High risk	Category 1
<b>Micro</b>	£104	£104	£2,645	n/a
<b>Small</b>	£104	£104	£45,163	n/a
<b>Medium</b>	£355	£256,017	£256,017	n/a
<b>Large</b>	£709	£3.4m	£3.4m	£13.4m

363. It should be noted, that the per business costs for medium and large businesses above are for businesses in these categories expected to incur costs for all compliance actions. Most larger mid risk businesses and high risk businesses will incur lower costs than noted in the above table as we do not expect them to incur additional content moderation costs (90% of medium and large mid risk businesses and 75% of high risk businesses).

364. Given the proportionate and risk based design of the regulations, the vast majority of costs fall on medium and large businesses. Based on the cost distribution across size bands in the table

below (and the per business cost in the table above), costs are not expected to fall disproportionately on SMBs.

Table 54: Total costs for each size band

	<b>Total costs (10 year PV)*</b>	<b>Number of businesses (to nearest ten)</b>	<b>Percentage of in-scope businesses</b>
<b>Micro</b>	£40.3m	19,650	81%
<b>Small</b>	£8.2m	1,180	5%
<b>Medium</b>	£528.9m	2,830	12%
<b>Large</b>	£1,191.9m	660	3%

\*Please note: These costs do not include the industry fee as it is not clear which businesses are likely to contribute; however, given the revenue threshold aspect of the fee, the majority are expected to fall on medium and large businesses.

365. For potential business actions where cost evidence is limited, this assessment attempts to reflect differences in costs faced by small and micro businesses. For example, familiarisation costs are expected to vary between different sizes of businesses, reflected in this assessment through the number of staff that read the regulations and the additional activity of disseminating the information throughout medium and large businesses. It may also be the case that in reality this should further be reflected in the wage of the member of staff familiarising themselves with the information. In this IA, the wage of a 'regulatory professional' is used as a proxy for the member of staff's wages; however, small and micro businesses are unlikely to employ these types of employees and it may in fact be the business owner that familiarises themselves. For the final stage, the government will engage further with SMBs to better understand differences in costs and whether they will face any additional costs not incurred by larger businesses.

## Mitigations for SMBs

366. This section sets out the ways in which the preferred option has been designed to ensure that overall costs of compliance are not disproportionate (i.e. that it does not represent a high proportion of total costs for SMBs compared to medium and large businesses). This is done by addressing how the potential mitigations for SMBs identified by the RPC have been considered<sup>167</sup> (see table below). This has been used to develop a "journey" for SMBs to identify how policy measures will adequately respond to challenges and barriers faced by SMBs and the points along this journey at which these will be addressed.

367. As set out in the table below, across the regulatory framework, SMBs will be protected by the risk-based and proportionate approach of the regulatory framework. Proportionality is further embedded through the following policy decisions:

- Reviews and comments on products, services and content directly published by an online service provider will be out of scope of regulation. In practice this will exclude low-risk businesses with limited functionality, many of which are likely to be SMBs.
- The vast majority of businesses will likely only be expected to respond to illegal harms, and not legal but harmful content and behaviour accessed by adults (only Category 1 organisations).

<sup>167</sup> [Checklist tool for a high-quality SaMBA - RPC \(August 2019\)](#)

- Furthermore, the regulator, in developing its codes of practice, will be required by law to pay due regard to innovation, to consult with a range of stakeholders, and to carry out IAs which consider explicitly the impact on SMBs.

Table 55: SMB mitigations

Potential mitigations (as suggested by the RPC)	How they have been considered in the Bill
Differentiated regulatory approach and requirements, which will likely apply to the majority of small businesses	<p>The majority of in-scope businesses will only be required to respond to illegal content and put in place measures to protect children (including from online content/activity which may be legal for adults, e.g. pornographic). A narrower range of service providers (Category 1) will be additionally required to respond to both illegal and legal but harmful content and behaviour on their services. This will form a broader duty of care for the safety of <i>all</i> users. Additionally, only Category 1 businesses will be required to publish transparency reports. We expect a small number of only large businesses to be designated as Category 1.</p>
Partial exemptions - use of derogations and de minimis measures (e.g. use of warnings to businesses rather than applying sanctions where non-compliance is identified)	<p>Exemptions will apply to online product and service reviews as well as 'below the line' comments. This will reduce the regulatory burden on many low risk businesses who have a low degree of user interactions and user generated content. Many of these will be SMBs.</p> <p>Businesses will be involved during the investigation process, allowing the opportunity to make representations about draft judgements and decisions the regulator makes. This, for example, would allow for greater understanding of how a business has responded proportionately to the regulatory framework.</p> <p>Enforcement measures will begin with warnings and notices ahead of any sanctions being issued. The regulator will have the discretion to set the level of fines which will take into account the size of the business (revenue, users, staff) alongside the actual or potential harm caused.</p> <p>The proposed appeals process would create a statutory appeal body which would allow for a more accessible and affordable route to appeal.</p>
More discretion for smaller businesses to meet regulatory requirements* (e.g. extended transition period or temporary exemption)	<p>This was not considered separately as the duty of care approach already builds in significant discretion for businesses to decide how to meet regulatory requirements. businesses will not face prescriptive requirements, but will be expected to assess their level of risk and put in place proportionate measures to address this.</p>
Simpler and clearer guidance on how to comply. More compliance support for small	<p>The regulator's codes of practice will play a critical role in ensuring that businesses of all sizes understand what is required of them under the duty of care. As well as the requirement to be consistent with the</p>

Potential mitigations (as suggested by the RPC)	How they have been considered in the Bill
businesses from the government and regulators	<p>principle of risk-based and proportionate action, the regulator will also be required to have regard to the need to:</p> <ul style="list-style-type: none"> <li>● <b>ensure all businesses are able to understand and fulfil their responsibilities</b> under the “duty of care” (for example through information about assessing and responding to risk; providing sufficient certainty about what is required and ensuring steps set out in codes of practice are feasible) and</li> <li>● <b>cater for all businesses whatever their risk level and capacity</b> (for example by providing support to start-ups and SMBs, drawing on best practice in other sectors).</li> </ul> <p>Businesses will not be obliged to comply directly with all the contents of the codes of practice; they may implement alternative approaches provided they can demonstrate that these are as effective or are more effective.</p> <p>The government is also developing a Safety by Design framework targeted at SMBs that will support businesses in adopting a “Safety by Design” approach, helping them design in user-safety to their online services and products. This work will produce practical online guidance tailored to SMBs. The framework will support SMBs to prepare for the introduction of the duty of care.</p> <p>The government will also build into the regulator a practical compliance support function for SMBs, following the model adopted by the Financial Conduct Authority (FCA).</p> <p>In addition, DCMS is undertaking a number of measures to stimulate and grow the UK commercial market in products and services supporting online safety, so that businesses in scope of the duty of care have a greater choice of tools they need to monitor online behaviour or protect users, at appropriate price levels. In particular, DCMS will explore opportunities to encourage suppliers to integrate safety tech into businesses’ existing product suites (e.g. their current moderation, cybersecurity or filtering solutions) rather than expecting businesses to purchase new standalone technology.</p>
Stronger culture of transparency and learning*	<p>The independent regulator will be a centralised body with a clear remit and responsibility to lead efforts to share learning and encourage collaboration between businesses and between sectors and to promote innovation and best practice. As per the envisaged set-up of the regulator, it will have a dedicated digital, data and innovation function to lead these efforts.</p> <p>The regulator will be one fit for the digital age, with a culture of proactive monitoring, evaluation and improvement, working with a range of stakeholders including industry, civil society and users to be continuously improving, refining and innovating. For example, a rigorous approach to understanding business impact based on on-the-ground research would help it to understand what’s working well and where businesses might need more support. It will also focus on</p>

Potential mitigations (as suggested by the RPC)	How they have been considered in the Bill
	<p>collaborative methods for policy and implementation, and focus on inclusion of a broad range of stakeholders.</p> <p>In addition, OFCOM will be required to conduct IAs on all new (or revised) codes of practice with further requirements to specifically assess the impacts on SMBs and innovation - this goes beyond normal regulator requirements as set out under the SBEE Act 2015.</p>
Different requirements for different sizes of businesses	<p>As mentioned above, not all businesses will be expected to respond to all harms: many, and most SMBs, will only be required to respond to illegal harms and to protect children online. Furthermore, the regulator's codes of practices will set out proportionate requirements. For example, the legislative requirement to have effective and accessible mechanisms for user redress will vary between businesses; the smallest and lowest-risk businesses might only be expected to have an email address for contact (which is already a legal requirement under the Electronic Commerce Regulations 2002).</p> <p>SMBs will unlikely be required to pay the annual fee or notify the regulator as they will fall under the notification threshold set by the regulator.</p> <p>For those businesses over the specified notification and fee threshold, in order to ensure the fee being paid reflects the business's financial circumstances each year, businesses will be required to update their business data relating to revenue so that the regulator can reassess the applicable fee. Therefore, a business which may have paid a fee and notified in year 1 could be exempt from paying a fee and notifying in year 2 should its global revenue decrease upon updating the regulator with their financial circumstances.</p>
Financial aid (e.g. reimbursement of compliance costs)	<p>Whilst there may not be reimbursement of payments from businesses to the regulator, there are mechanisms in place to ensure that any non-enforcement related payments from businesses are not disproportionate. All funding options for the regulator have been evaluated against the criteria of proportionality and affordability and one specific example is the annual industry fee. The fee will be tiered and informed by the regulator's regulatory timesheet data. The annual fees charged to industry will therefore be informed by the total quantum of costs incurred by the regulator in running the online safety regime, therefore the fee is proportionate.</p> <p>The regulator should not be in a position to reimburse businesses or not be able to cover any regulatory costs.</p>
Opt-in and voluntary solutions	Voluntary approaches have been tested in the sector but have not been successful (see rationale for intervention).

**Consultation question 16:** Do you agree with the assessment of the impacts on small and micro businesses? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 17:** Is there anything additional the government can do to support small and micro businesses in the implementation of this regime?

## Wider impacts

### Trade impacts

**Does this measure have potential impacts on [the value of] imports or exports of a specific good or service, or groups of goods or services?**

368. The online harms regulations will apply to any in-scope service provided to UK users regardless of where the service is based. The scope of the framework is functionality based, i.e. it is both good/service and sector agnostic.

369. It is difficult in the context of online platforms and online harms in particular to apply the import/export framework to assess potential impacts. For example the UGC/P2P interaction functionality offered by an online platform could be the service itself - in which case a normal trade in services framework would apply - or it could just be a minor part of the online presence of a business which attains revenue from an entirely unrelated good or service.

#### Where UGC/P2P interaction is the main offering

370. The UK is an important market for many of the most affected types of organisations. Using social media businesses as an example (whose main offering to users and the advertisers who spend money is the UGC and P2P interaction<sup>168</sup>), the UK is the 13th largest market in terms of user base for Facebook<sup>169</sup>, 5th for Twitter<sup>170</sup>, 8th for Instagram<sup>171</sup>, 3rd for Pinterest, and 4th for Snapchat<sup>172</sup>. It is therefore unlikely that the online safety framework would lead to a reduction in services offered to UK users (or UK advertisers). Additionally, similar regulations such as Germany's NetzDG and the EU's Digital Services Act have been/will be implemented. Other countries are also expected to follow suit in updating regulatory policy for online harms. Platforms offering UGC and P2P services to UK users will therefore not be at a significant disadvantage from those that operate elsewhere as the regulatory landscape for online platforms is evolving internationally.

371. Unlike a business which manufactures goods, production (or in this case the ability to provide the service) for businesses whose main offering is online UGC and P2P interaction is not finite. In other words, if the cost of regulatory compliance for a business producing goods becomes excessive in one country, given the business's finite productive capacity, it would be worthwhile instead selling the goods elsewhere where regulatory burdens are lower. This is not the case for businesses whose main offering is UGC and P2P interaction where the choice is determined

---

<sup>168</sup> It could be argued that the main offering to advertisers is the user base (rather than UGC and P2P interaction specifically); however, the ability of users to react, like, discuss and share is what sets social media advertising apart from traditional forms.

<sup>169</sup> [Leading Countries Based on Facebook Audience Size as of January 2021 - Statista](#)

<sup>170</sup> [Leading Countries Based on Number of Twitter Users as of January 2021 - Statista](#)

<sup>171</sup> [Instagram Demographic Statistics: How many people use Instagram in 2021? - Brian Dean](#)

<sup>172</sup> [Leading Countries Based on Snapchat Audience Size as of January 2021 - Statista](#)



solely by whether the benefits from providing the service in that country, e.g. ad revenue or similar, exceed the cost of compliance. This IA estimates a relatively modest per business cost of compliance which is proportionate to business risk.

372. For services currently offered to UK users only, who may in the future, look to enter other markets, we do not expect compliance costs to put them at a competitive disadvantage. The cost of complying with the regulation will increase business costs; however, businesses will be in a more favourable position to compete on user safety. Over half of respondents to an OFCOM survey have spontaneous (not prompted by the interview question) concerns about interaction with other people/content online<sup>173</sup>. The same survey indicated that 9 of the top ten most cited concerns of adults relating to children's use of the internet were on interactions with other people/content (including bullying/abusive behaviour/threats, sexual/pornographic content, and violent and disturbing content - as the top three). Given the general public's concerns about internet safety, compliance with the online safety framework could be considered to be a competitive advantage for UK providers<sup>174</sup> on the international stage.

### **Where UGC/P2P interactions are secondary**

373. Some businesses - that could not be considered traditional digital businesses - will be within scope of the regulations solely due to offering UGC or P2P interaction functionality on their website. For example, a business which sells a traditional good or service (retailers, legal services etc) but that offers a forum function on its website could be in scope. As noted earlier, compliance with the online safety framework will increase the cost of doing business for these organisations. However, given the risk-based design of the framework, any compliance costs are expected to be proportionate. Further, the introduction of the 'low risk' functionality exemption has removed a large proportion of these types of businesses from scope, e.g. small hospitality, beauty and health businesses and most retailers, where they simply have a comment function for reviews on their products.

374. At the margins, some of these businesses - still in scope after all the exemptions - may remove some functionalities from their websites instead of incurring compliance costs.

### **Does this measure include different requirements for domestic and foreign businesses?**

375. The framework will apply to any in-scope business worldwide that provides services to UK users. There are no differing requirements for domestic and foreign businesses. Applying this policy to all businesses providing services in the UK will help to ensure a level playing field between businesses that have a legal presence in the UK, and those who operate entirely from overseas. The UK is pathing the way in this regulatory landscape, although, other countries worldwide are also introducing legislation in this space. There may consequently not be a marked difference in operating costs between similar jurisdictions as other countries look to align.

### **Does this measure have potential impacts on [the flow or value of] investment into and out of the UK?**

376. There is a risk that the regulation could dissuade foreign investment and/or encourage UK based organisations to disinvest in the UK if the compliance costs are too high. The arguments presented above on trade apply equally for investment in so far as businesses are not expected to stop providing services to UK users and compliance costs are not expected to stop platforms who provide services to UK users to be able to provide services to non-UK users.

377. There is evidence to suggest that, in the short- to medium-term, there will not be a large net outflow of investment, especially from digital sectors. The largest businesses have large and

---

<sup>173</sup> [Internet Users' Experience of Potential Online Harms: Summary of Survey Research - OFCOM \(2020\)](#)

<sup>174</sup> UK providers here refers to platforms providing services to UK users only.

sticky investments in the UK market. Each of the GAFAM (Google, Apple, Facebook, Amazon, Microsoft) businesses<sup>175</sup> has a UK HQ. They also have large investment in value-add employment (i.e. not just selling to UK customers but services that can be exported): the UK hosts the largest Facebook engineering base outside of the US, and Apple has a large R&D centre in Cambridge. Large businesses are already taking measures to combat online harms, we would therefore expect there to be a minimal impact upon their investment and business activity within the UK. For these reasons, investment flows are not expected to be significantly affected.

## WTO notification

378. The WTO requires members to “promptly or at least annually issue notifications of new or amended legislation that will ‘significantly affect’ international trade in services under the GATS”. On advice from the Department for International Trade the government will not be required to notify the WTO about this legislation.

379. **The forthcoming regulatory framework will not target specific service sectors.** The legislation will focus on businesses who host user generated content, or facilitate interactions between users.

380. For those businesses who fall within scope, the framework will impose a duty of care to ensure that they have robust systems and processes in place to keep their users safe. They will need to take action with regard to illegal content on activity, and will need to assess the likelihood of children accessing their services and, if so, provide additional protections for children using them.

381. **The Online Safety Bill will not ‘significantly affect’ the ability of UK businesses to trade or provide services overseas, or the ability of overseas businesses to export to the UK or provide services in the UK.**

382. As set out above, all businesses will be required to take action with regard to relevant illegal content and activity and assess the likelihood of children accessing their site. Only a small number of ‘Category 1’ businesses providing high-risk, high-reach services will have a legal obligation with regard to legal but harmful content and activity accessed by adults on their services. These businesses will be required to undertake regular risk assessments to identify legal but harmful material on their services, and enforce these terms of service consistently and transparently. Therefore even for this small number of high-risk businesses, the requirements will not significantly impact the ability to provide their services to the UK.

383. Where an organisation fails to fulfil their duty of care, the regulator will be able to issue fines and take business disruption measures against them. This may include removing access to payment facilities, advertising revenue, or reducing their visibility in search results. Only in the most serious and egregious of failures, the regulator will be able to block access to the services from the UK. This is expected to be a very small number of cases.

384. **This legislation will impact UK businesses in the same way as foreign businesses operating in the UK.**

## Innovation Test

385. While sector agnostic in its design, the online safety framework is risk-based and therefore, the majority of requirements will fall on businesses with websites offering high levels of UGC and

---

<sup>175</sup> Google, Apple, Facebook, Amazon and Microsoft.

P2P interaction functionalities, e.g. social media and other digital technology businesses. These types of businesses are often innovative and high-growth businesses. For example, in 2019 Facebook and Snapchat spent \$13.6 billion (around £10.4 billion) and \$884 million (around £678 million) respectively on research and development<sup>176</sup> with Twitter spending \$682 million (around £522 million) in the same year<sup>177</sup> - all expected to be in scope of the regulations. The compliance requirements of this framework will therefore disproportionately fall on highly innovative sectors. However, these are also the types of businesses that are already investing substantially in user safety and it is therefore assumed that they do not necessarily see a trade-off between user safety and innovation. Furthermore, DCMS did not find any evidence in the [REA on NetzDG](#) that the German legislation impacted innovation.

386. The aim of this proposal is to minimise any indirect impacts of regulatory compliance on wider innovation while encouraging innovation in safety technology markets. As noted in the first section of this IA, alongside the legislation the government is separately funding a number of business and user support measures to both support the online safety framework implementation and invest in the safety technology sector.

387. Protecting and encouraging innovation is a key consideration for the framework. The policy has been designed from the start with innovation at the forefront:

- By implementing through primary legislation and codes of practice, it gives the regulator flexibility to lay and revise codes of practice as new technologies emerge
- There is a specific requirement on the regulator to produce IAs for all new and revised codes of practice and to ensure within these, that the impact on innovation is considered.
- The framework is principles-based and businesses are given the freedom to meet high-level requirements in the most efficient way allowing them to undertake alternative measures that prove to be sufficiently effective.
- Options analysis considered the adaptability to future technological changes as a key criteria and impact on innovation.
- Implementation of the policy will be risk-based so the regulator can focus resources on the most serious online harms (even if that changes).
- The approach taken will be technology neutral and therefore encompass future changes to how the architecture of the internet functions.
- Development of business and user support measures, separate to OH regulation, which focus on researching emerging harms and the working safety technology sector to encourage innovative solutions to the problems.
- Proportionate system (e.g. smaller and less risky businesses have to do less), this will minimise the disincentive effects of the regulation and minimise the impact on new entrants.
- Partial exemptions will be implemented to reduce the regulatory burden on many low risk businesses who have a low degree of user interactions and user generated content. Many of these will be SMBs.

388. The impact on smaller businesses and start-ups will depend on the degree to which proportionality is built into the system, and the ways in which the independent regulator is able to reduce the burden on SMBs. The SaMBA above outlined a number of potential mitigations for SMBs - these include: partial exemptions; proportionate enforcement; a duty of care with significant discretion for businesses to decide how to meet the requirements; clear and tailored guidance for SMBs, including in advance of legislation, a voluntary Safety by Design framework targeted at SMBs; a practical compliance support function for SMBs built into the regulator; and a proportionate fee structure which considers business size.

---

<sup>176</sup> [Research and Development Expenditure of Leading Internet Companies 2014-2019 - Statista](#)

<sup>177</sup> [Twitter Fiscal Year 2019 Annual Report - Twitter](#)

389. Consideration of innovation has been at the forefront of policy design and will continue to be during its implementation. For the reasons noted above, indirect impacts on innovation are expected to be negligible.

390. Finally, the monitoring and evaluation (M&E) section outlines a detailed plan which will consider the policy's impact on innovation and any unintended effects in this area.

**Consultation question 18:** Do you agree with the assessment of the impacts on trade and innovation? The government welcomes any evidence you can provide to refine the estimates.

## Equalities Impact Assessment

Statutory Equalities Duties	Completed
<p>Alongside this consultation impact assessment, DCMS has produced an equalities impact assessment, the contents of which is summarised in this table and further in the below equalities section.</p> <p>Proposals set out in the Online Safety Bill to make the internet a safe place for all users are expected to have an overall positive impact on individuals with protected characteristics. The government is not aware of any possible direct discrimination, in relation to the Bill, and when considering indirect discrimination various elements of framework are expected to positively impact users with protected characteristics. These elements include a higher level of protections for children, requirements to assess risks to users, requirements for major platforms to clearly state what content is considered acceptable in their terms of service and to enforce these consistently and transparently, further promotion of media literacy, the establishment of a super-complaint function, and the requirement for all services to have easily accessible user redress mechanisms. <b>Overall, the proposed framework will help advance the protections of the Equality Act 2010 online and make the internet a safer place for all, including those with protected characteristics.</b></p>	<p>Yes</p>

391. DCMS as a public body has a legal obligation to consider the effects of policies on those with protected characteristics<sup>178</sup> under the Public Sector Equality Duty set out in the Equality Act 2010.

392. Overall, these proposals are expected to have a positive impact on users with protected characteristics under the Equality Act 2010. This is incorporated in the overarching aim of the policy; to make the internet a safe place for all users. Reducing online harms is particularly important for those with protected characteristics, many of whom are disproportionately more likely to be victims of online abuse and discrimination, for example:

- From a survey of 700 LGBT+ people, 8 in 10 respondents had experienced anti-LGBT+ hate crime and hate speech online in the last 5 years<sup>179</sup>.
- Community Security Trust logged 697 instances of online antisemitism in 2019, comprising 39% of the annual total and a rise of 50% from the 466 online incidents reported in 2018 (28% of that year's total)<sup>180</sup>. And similarly, interim figures from Tell

<sup>178</sup> Age, disability, sex, gender reassignment, pregnancy and maternity, race, religion or belief and sexual orientation

<sup>179</sup> [Online Hate Crime Report - Galop \(2020\)](#)

<sup>180</sup> [Antisemitic Incidents Report - Community Security Trust \(2019\)](#)

MAMA (Measuring Anti-Muslim Attacks) showed that in 2019 alone they received 155 reports of online hate crime and incidents.<sup>181</sup>

- Users with disabilities have been forced to leave social media as a result of the abuse they had experienced online.<sup>182</sup>
- Women tend to be disproportionately affected by online offences like harassment, stalking, revenge pornography.

393. This section will highlight some of the key ways in which the regulation will address current issues faced online by users with protected characteristics.

394. The assessment of prospective equality impacts that online harms proposals may have on those with protected characteristics was considered in regards to both direct and indirect discrimination:

- Direct discrimination occurs when a person (A) treats person (B) less favourably than A treats or would treat others because of a protected characteristic possessed by B (s13, Equality Act 2010).
- Indirect discrimination occurs when a person (A) applies to person (B) a provision, criterion or practice which is discriminatory in relation to a relevant protected characteristic of B's (s19, Equality Act 2010).

395. At present, the government is not aware of any possible direct discrimination, in relation to each of the protected characteristics, which will result from this policy and will amend the Equalities IA if any become apparent.

396. Additionally, when considering indirect discrimination, various elements of the regulatory framework indicate ways in which the policy will positively impact users with protected characteristics. These include:

- Requirement for Category 1 organisations to have clear terms of service and to enforce them effectively and transparently. Platforms will be required to have clear guidance in their terms of service about what is acceptable behaviour on their platform. These may contain explicit guidance about unacceptable behaviours relating to people with protected characteristics, for example, abusive language toward disabled users.
- Improving media literacy for all users. Some individuals from protected characteristic groups, for example children, the elderly or in some cases disabled people, have been identified as more vulnerable to online harms. The media literacy efforts incorporated in this policy may therefore be particularly important to enable these users to be able to critically and independently manage their own risks online.
- Super-complaints. This function would be open to organisations, who meet a set eligibility criteria, wishing to report systemic failures to comply with the duty of care across two or more services (or in exceptional circumstances one or more services). This would allow organisations representing users from protected characteristic groups to report systematic issues affecting those groups.
- Requesting that redress mechanisms are easily accessible for all users. This would ensure that report functions are clear and accessible to all users, including those with protected characteristics who may be otherwise less likely to navigate and pursue them. At present, complaint and reporting functions tend to be disproportionately used by those from higher-socioeconomic groups and with higher educational qualifications, men, and older people.

397. The government does not expect this policy to impact negatively on people with protected characteristics. However, it is possible that in response to regulation companies may adopt a content takedown focussed approach which could potentially impact people with protected characteristics disproportionately. Both DCMS and OFCOM will be monitoring this post-

---

<sup>181</sup> [The Impact of the Christchurch Terrorist Attack, Tell MAMA Interim 2019 Report](#)

<sup>182</sup> [House of Commons Petitions Committee report \(2018\)](#)

commencement. However, the focus of the framework on systems and processes, as opposed to content, is intended to avoid this. Legislation also includes further safeguards, such as a duty on companies to consider impacts on users' rights, including freedom of expression. Additionally the super-complaints function and user advocacy mechanism will help the regulator to understand whether there are systematic failures affecting protected groups, and respond according to its own Public Sector Equality Duty obligations and further obligations to vulnerable users.

398. Overall, the proposed framework will help advance the protections of the Equality Act 2010 online and make the internet a safer place for all, including those with protected characteristics.

**Consultation question 19:** Do you agree with the points made in the equalities IA? The government welcomes any evidence you can provide.

## Competition

399. The **REA on NetzDG** did not present any evidence that the German policy had any impact on market competition; however, the proposals in the OS Bill could potentially impact competition in the market if compliance costs:

- create – or are viewed by potential new entrants as - a barrier to entry; or
- fall disproportionately on SMBs, i.e. they are not able to absorb the costs (in unit terms) as easily as larger businesses; or
- dissuade foreign investment and/or encourage UK based businesses to disinvest in the UK.

### Will the measure indirectly or directly limit the range or number of suppliers?

400. The proposals could indirectly limit the number of suppliers if for example, compliance costs are seen by potential entrants to the market as barriers to entry or realised costs of compliance force some providers out. Given the differentiated requirements on businesses (of size and risk) and the proportionate enforcement expected of the regulator, these impacts are expected to be minimal. Beyond familiarising themselves with the regulations, a low risk in-scope micro business may only be required to produce a risk assessment, ensure it has an email address for potential user reporting and conduct no or minimal additional content moderation (one small low risk organisation interviewed for example, noted that moderating was already a part of business as usual and 'negligible'). Given this approach, the proposal does not directly limit the number of suppliers and even indirectly, businesses would not be expected to exit the market due to the proposed regulation.

### Will the measure limit the ability of suppliers to compete?

401. For platforms where UGC and P2P interaction is secondary to the good or service being sold, this measure is not expected to limit their ability to compete given the main areas of competition (price and quality) are largely unrelated to that aspect of their website. These businesses may find that the cost of compliance is not worth the benefits of having this functionality on their site and they may remove it. However, for platforms where UGC and P2P interaction is the service, this proposal may reduce smaller businesses' ability to compete. For example, size is not a perfect proxy for risk of online harms (although there is a link) and therefore, a business like Facebook may be in the same risk tier as a much smaller (in terms of employees and revenue generation) social media business. Businesses in the same risk category are bound by the same duty of care and given that Facebook (in our example) will find it much easier to absorb compliance costs than the smaller social media platforms there may be distortionary effects. To limit this, there will be differentiated requirements within duties - for example, while all Category 1 businesses will have to report on transparency, the information they are required to collect and publish may vary proportionately. Additionally, based on the intention of the policy, small or micro



businesses are not expected to be designated as Category 1. Every effort will be made to minimise any impacts on competition both in implementation and enforcement of the regime.

### **Will the measure limit the suppliers' incentives to compete vigorously?**

402. Regulation of online harms will have a minimal impact upon the suppliers' incentive to compete. There is a risk that the regulation could inadvertently encourage collusion (e.g. sharing data, forming research groups and sharing technology), however, this risk is expected to be negligible. By introducing a minimal level of online harm action this proposal could potentially limit businesses' ability to compete on that aspect of their services, i.e. user safety.
403. The policy will encourage competition in the safety tech market as businesses will be looking for efficient and cost effective compliance solutions, leading to the potential for growth of the UK safety tech sector<sup>183</sup>.

### **Will the measure limit the choices or info available to consumers?**

404. The policy will increase information available to consumers through bridging the information gap between businesses and consumers through increased transparency, as detailed in the Rationale for Intervention. This will allow consumers to make informed decisions about their use of online platforms, driving greater competition between businesses to implement measures meeting regulatory and consumer demands for increased safety on online platforms.

**Consultation question 20:** Do you agree with the assessment of the impacts on competition in the market? The government welcomes any evidence you can provide.

405. For a full list of the consultation questions, please see [Annex F](#).

## **Devolution Test**

406. Internet law and regulation is a reserved policy area under all three devolution settlements. The online safety regime will apply across the whole of the UK.
407. The online safety legislation is considered to be reserved, however, there are a number of areas within the regime where there is possible interaction with devolved competencies, and so government is working closely with the Territorial Offices (TOs) and Devolved Administrations (DAs) to ensure that such issues are taken into account. This includes issues such as harms in scope and media literacy.
408. While some of the harms relate to offences in Scottish or Northern Irish Law, and therefore involve devolved competences, the legislation is not seeking to change the law in relation to these offences. Instead, our proposals seek to clarify the responsibility of businesses to tackle this activity on their services.
409. DCMS has engaged regularly with the DAs, TOs, and OFCOM's offices in the devolved nations as proposals have been developed, and it will continue to engage throughout the legislative process.

---

<sup>183</sup> Under the Better Regulation Framework (and RPC's guidance) that this would be considered resources used to comply with regulation.

# Monitoring and Evaluation

## Evaluation plans

### Review clause

410. The OS Bill will contain a statutory review clause and a post-implementation review (PIR) will be conducted within 5 years of implementation. At this stage, it would not be wise to provide a more explicit timeline for the review given the fast moving nature of the policy area and the iterative process of producing codes of practice. It will be for the Secretary of State to determine the specific point at which a review is necessary but this is expected to be between 2-5 years of implementation (and within 5 years) unless there is a clear and obvious reason for delaying or expediting the review. Any review will take a holistic approach and will evaluate the entirety of the online safety framework through consultation with relevant organisations and affected parties.

411. There are broadly three areas of evaluation:

- Evidence from implementation of individual codes of practice;
- A review of the wider online safety framework; and
- Reviews of the government's separate but related activity of investment into business and user support measures, including media literacy projects, investment in the safety technology sector, safety by design, and adult and child safety initiatives.

412. This section focuses mainly on the review of the wider online safety framework and it will be that which is the subject of a post-implementation review. While it will be for OFCOM to monitor the effectiveness of and evaluate individual codes of practice, where possible, these will be used in the production of the overarching review.

### Who will conduct the review?

413. The review will be led by the DCMS Secretary of State, however will consult with the following stakeholders either through steering groups or consultations:

- Home Office;
- OFCOM;
- OGDs, e.g. Ministry of Justice (justice impacts), Department for Business, Energy and Industrial Strategy (SMBs and business impacts)
- Regulated entities, i.e. online platforms
- Civil society groups
- Wider society

### What will the review consider?

414. The monitoring and evaluation plan will consider how the programme achieves the proposals' objectives: user safety; preserving freedom of speech; law enforcement; efficiency and evidence. This will allow government to ensure the policy is meeting the defined objectives throughout the policy process, and enable emerging evidence to inform ongoing adjustments to the intervention as well as informing ongoing implementation. The review will consider:

- Whether the online safety framework has achieved its stated objectives



- Whether the impacts of the policy were in line with those estimated in previous IAs (both primary and codes of practice)
- Whether the policy has resulted in any unintended consequences
- How well the regime is functioning in practice and whether there are any areas which could be improved through changes to legislation (or recommendations to the regulator)

415. The current regulatory landscape used to tackle illegal or harmful online content is both fragmented and limited. No single regulatory body is accountable for ensuring internet users are safe from online harm. Regulatory activity is dispersed by harm type and by the nature of the service or platform upon which content or activity is hosted. As a result, there is currently no comprehensive monitoring and evaluation strategy in place under the *status quo*. That said, existing data and evidence will feed into the design of the evaluation, and enable the government to determine the types of measures to be used to evaluate each of the objectives. This review will draw on the following sources of evidence (note, this is not an exhaustive list):

- Evidence from OFCOM monitoring and evaluation of codes of practice
- Reviews of DCMS' business and user support measures, including media literacy projects, investment in the safety technology sector, safety by design, and adult and child safety initiatives. A number of initiatives within these projects look specifically at designing evaluation frameworks and identifying key metrics relating to internet safety which will be used in the review of the online safety framework.
- Official data sources on incidents of crime and prosecutions flagged with an online component. These monitoring plans will be developed to allow the tracking of the inputs, outputs and outcomes of an intervention based on regularly available published data from the Home Office and Ministry of Justice.
- Transparency reports from Category 1 organisations where relevant businesses publish transparency reports setting out more information about the prevalence of harm and the effectiveness of their safety systems and processes.
- OFCOM's reporting on compliance
- Information provided by industry, civil society and wider society. The review will include consultations with the main affected groups to identify: the realised compliance costs; unintended effects of the regulation; impacts on freedom of expression, innovation, competition, and trade and investment; and recommendations for how the regime can be improved.
- Expert research. It is likely that as part of the PIR process the Department will commission research and analysis of the online safety regime which will focus on the economic impact of the regime.

# Annex A: Changes to the policy since the OHWP

This section summarises changes between the OHWP and the current policy position. The consultation on the White Paper ran from 8 April 2019 to 1 July 2019. It received over 2,400 responses ranging from companies in the technology industry including large tech giants and small and medium sized enterprises, academics, think tanks, children’s charities, rights groups, publishers, governmental organisations and individuals. In parallel to the consultation process, the government has undertaken extensive engagement over the last 12 months with representatives from industry, civil society and others. These changes have been made in response to concerns raised by stakeholders during the consultation process.

***Please note, this does not cover everything in the government response but instead focuses on areas where there have been updates to the policy as a result of feedback from stakeholders.***

White Paper position & stakeholder concern	Government response
<p><b>Journalistic content:</b> The White Paper committed to ensuring protections for freedom of expression within the regulatory framework. During consultation, there were calls from stakeholders to exclude journalistic content from scope, to protect freedom of expression and avoid negatively affecting the public’s ability to access information or undermining quality news’ media.</p>	<p>In order to protect media freedoms, legislation will include robust protections for journalistic content shared on in-scope services. The government is committed to defending the invaluable role of a free media and is clear that online safety measures must do this. The government will continue to engage with a range of stakeholders to develop our proposals.</p>
<p><b>Services in scope:</b> The White Paper set out that the regulatory framework will apply to companies that provide services or tools that allow, enable or facilitate users to share or discover user-generated content, or interact with each other online. However, many parties expressed a need for clarity around organisations in scope. There were calls to exclude business-to-business services due to the lower risk of harm on those services.</p>	<p>The government will be maintaining a broad regulatory scope encompassing services that host user generated content and facilitate interaction between users, as well as search engines.</p> <p>Specific exemptions have been introduced for low-risk services. For example, reviews and comments by users on a company’s website which relate directly to the company, its products and services, or any of the content it publishes, will be out of scope.</p>
<p><b>Definition of harm:</b> The White Paper set out an initial list of harms in scope but made clear this was, by design, neither exhaustive nor fixed. Companies and stakeholders wanted more detail on the breadth of both services and harms in scope. There were calls to protect freedom of expression and a focus on protecting children.</p>	<p>The legislation will define the harmful content and activity in scope of the regime. A limited number of priority categories of harmful content will be set out in secondary legislation.</p> <p>Some categories of harmful content will be explicitly excluded, to avoid regulatory duplication. This will provide legal certainty for companies and users and prioritise action on the biggest threat of harm.</p>
<p><b>Sale of Unsafe Goods:</b> The White Paper did not set out a definitive position on whether the sale of unsafe goods would be in scope of the new regulatory framework. A number of organisations suggested that economic harms should be in scope, noting that such activity could also lead to significant psychological harm. Others argued that the scope of the</p>	<p>The Office for Product Safety and Standards has a clear remit for consumer product safety, including products sold online. In order to avoid regulatory duplication the sale of unsafe products will be excluded from the online safety regulatory framework.</p>

<b>White Paper position &amp; stakeholder concern</b>	<b>Government response</b>
<p>regulatory framework was too broad, and that any further extension would pose disproportionate regulatory burdens on businesses.</p>	
<p><b>Duty of care and principles of the regulatory framework:</b> The White Paper stated that there would be a new statutory duty of care to make companies take more responsibility for the safety of their users. Industry sought greater reassurance and certainty about how it would be proportionate in practice, particularly for small and medium-sized enterprises and how flexibility would be balanced with certainty about what the duty of care requires of companies. Rights groups and industry also emphasised the need to provide more certainty about how safety would be balanced with freedom of expression, particularly in relation to legal but harmful content.</p>	<p>In order to provide more clarity and targeted effectiveness, the duty of care has been refined. It will cover content and activity that could cause harm to individuals. The duty of care will not cover harms to society more broadly. The legislation will also introduce specific provisions targeted at building understanding and driving action to tackle disinformation and misinformation in the longer term.</p>
<p><b>Differentiated expectations on companies:</b> The White Paper set out that all services in scope will be required to address illegal and legal but harmful content and activity. The consultation responses flagged concerns about the broad scope of harms, calling for greater clarity and highlighting the subjectivity inherent in identifying many of the harms, especially those which are legal. Many respondents objected to the latter being in scope. There were concerns that proposals could impact freedom of expression online.</p>	<p>The initial government response built on the original position, confirming a differentiated approach for illegal content and activity versus content that is legal but harmful. Only companies providing Category 1<sup>184</sup> services will have to take action in respect of adult users accessing legal but harmful content on their services.</p> <p>All in-scope companies will be expected to assess whether children are likely to access their services, and if so, take measures to protect children on their services including reasonable steps to prevent them from accessing age-inappropriate and harmful content. This includes, for example, the use of age assurance and age verification technologies, which are expected to play a key role for companies in order to fulfil their duty of care.</p>
<p><b>Codes of practice:</b> The White Paper stated that the independent regulator would set out how companies could fulfil the duty of care in codes of practice. Some respondents argued that too many codes of practice would cause confusion, duplication, and potentially, an over reliance on removal of content by risk averse companies.</p>	<p>There will not be a code of practice for each category of harmful content. The codes of practice will focus on systems, processes and governance that in-scope companies need to put in place to uphold their regulatory responsibilities.</p>
<p><b>Using technology to identify illegal child sexual exploitation and abuse content and activity:</b> The White Paper set out that some private channels would be in scope of the</p>	<p>The regulator will set out how companies can fulfil their duty of care in codes of practice, including what measures are likely to be appropriate in the context of private communications. Companies in scope will</p>

<sup>184</sup> Category 1: High risk, high reach companies

<b>White Paper position &amp; stakeholder concern</b>	<b>Government response</b>
<p>online safety regime, however companies would not be required to scan or monitor for illegal content on these services, reflecting the importance of privacy. Stakeholders argued that private communications should either fall out of scope or be subject to very limited requirements, to protect user privacy. By contrast, some online safety organisations and children’s charities argued private communications should be in scope because there is a high risk of harmful activity - such as child grooming - on private channels.</p>	<p>need to consider the impact on users' privacy and ensure users understand how company systems and processes affect user privacy.</p> <p>The regulator will have the power to require companies to use automated technology that is highly accurate to identify illegal child sexual exploitation and abuse activity or content on their services. Recognising the importance of protecting users’ privacy, the government will ensure this will be used only where there are no alternative measures that are capable of achieving the same aim and subject to stringent legal safeguards to protect users’ rights.</p>
<p><b>Reporting to law enforcement:</b> The White Paper stated that the regulator would provide guidance on when companies should proactively alert law enforcement and other relevant government agencies about specific illegal content. Stakeholders argued that there should be new, mandatory reporting requirements for child exploitation and sexual abuse content to increase reporting and standardise the approach.</p>	<p>The government is minded to introduce a requirement for companies to report child sexual exploitation and abuse identified on their services, with these reports being made to a designated body.</p>
<p><b>Disinformation and misinformation:</b> In the White Paper, disinformation was included in an indicative list of harmful content or activity that would be within scope of the legislation, because it can be harmful to both individuals and to society. A range of stakeholders raised concerns about including disinformation and misinformation in scope of the regulation because of the impact this might have on freedom of expression. Many stakeholders are concerned about the threat that disinformation and misinformation poses to individual users, as well as its potential broader impact on public safety, national security and community cohesion.</p>	<p>The legislation will introduce specific provisions targeted at building understanding and driving action to tackle disinformation and misinformation. For example, establishing an expert working group on disinformation and misinformation, measures to improve transparency about how companies deal with disinformation and building on OFCOM’s existing duties to promote media literacy.</p> <p>Companies will need to address disinformation and misinformation that poses a reasonably foreseeable risk of significant harm to individuals (e.g. relating to public health).</p>
<p><b>Enforcement:</b> The White Paper set out that the regulator will have a range of enforcement powers to take action against companies that fail to fulfil their duty of care. Stakeholder feedback expressed an overall preference for the regulator to begin its operations by supervising companies and supporting compliance through advice, and that any further enforcement measures should be used proportionately and following a clear process.</p>	<p>The principles and objectives underlying the enforcement proposals have not changed fundamentally, but the government has provided further details on what enforcement activity will look like. This includes refining the additional enforcement powers that the government consulted on. The most notable developments are in our approach to nominated representatives, senior management liability and business disruption measures.</p>

# Annex B: Business and User Support Measures

## Introduction to Business & User Support Measures

An important consideration, beyond the regulatory framework itself, is the development of business and user support measures. Although separate from the core regulatory provisions, these measures will work in conjunction to help both businesses and users have a greater understanding of online harms and adapt their behaviour accordingly. These measures will work alongside the general regulation in making the UK the safest place to be online and the safest place to start an online business.

### Why are these necessary?

These measures are necessary for two main reasons: to empower users to feel confident in protecting themselves online, and to help businesses understand ways that they can create platforms that are safe for users. The need for these measures was identified by research looking at the current online landscape and existing media initiatives.

First, a project looking at users' experiences online found that: the level of media literacy in the UK is limited for adults and children; there were gaps in provision for people most vulnerable to online harms, including those with protected characteristics; there is limited evidence of effective evaluation of current interventions, and, there is a strong rationale for government intervention in online safety and digital media literacy on the grounds of equity, efficiency, and effectiveness\*.<sup>185</sup> There is also an increasing appetite for support measures amongst users; OFCOM recently reported that parents were more than twice as likely in 2019 than 2018 to seek out resources online to protect their children from online harms.<sup>186</sup>

Second, there is evidence that even when companies want to improve the safety of their products, there is a lack of understanding about the best ways of doing this. For example, research found that companies lacked knowledge on the risks posed to children when using their platform, and an understanding of best practice for designing these out. A key example of this is companies struggling to prevent children from lying about their age to create accounts on 13+ platforms or those with age-inappropriate content <sup>187</sup>.

Together, business and user support measures will be introduced to fill these gaps, supporting businesses to create safer online spaces and empowering users to navigate the internet in a safe way.

### What are they?

The business and user support measures are intended to support businesses and users to create and maintain a safe online environment. They include the following policy programmes seen in the table below.

<sup>185</sup> RSM, Mapping Exercise and Literature Review, RSM, **\*not yet published**

<sup>186</sup> Children's Media Use and Attitudes Report 2019, OFCOM, 2020

<sup>187</sup> GCHQ-led project, supported by DCMS, which found that for children 'online harms are endemic to using the internet' (from Business case: Strategic case)

Support measure	Description and responsible body
Child and adult online safety	This includes national research projects into the prevalence and impact of online harms on children and adults, as well as initiatives to engage with young people to help develop tools, interventions and guidance to help keep them safe online. Responsibility for delivering child and adult online safety policy will remain within DCMS
Online Safety Technology	This consists of initiatives to grow and support SMEs in the UK online safety tech industry, and includes defining the sector, export support, an innovation fund to generate solutions to safety technology problems, and funding for running a series of promotional events for the industry. This is a fast-growing sector. Responsibility for delivering online safety technology measures and policy will remain within DCMS.
Safety by Design	<p>This comprises programmes to produce and promote guidance on a safety by design approach and industry best practice. It also includes research projects and the development of a higher education module on safety by design.</p> <p>Responsibility for the promotion of safety by design measures (with the exception of age assurance standards) will be transferred to the regulator as it becomes fully operational in 2023. DCMS will support planning for this transfer of responsibility. One intervention within this workstream, the Age Assurance standard, will remain within DCMS after responsibility for delivery of the other measures is transferred.</p>
Media literacy	<p>This includes developing a media literacy framework to evaluate the outcomes of media literacy initiatives. It also includes piloting work with civil society groups and training initiatives for teachers, support workers, librarians and community groups, as well as a communications campaign.</p> <p>Some media literacy measures will be transferred to the regulator as it becomes fully operational in 2023. OFCOM already has some responsibilities on media literacy which it will retain and deliver in parallel. DCMS will support planning for this transfer of responsibility.</p>

# Annex C: A detailed overview of the chosen policy position

## Part 1-Who will the new online safety framework apply to?

### Services in Scope

#### **Original Policy Position**

The Online Harms White Paper<sup>188</sup> proposed that the regulatory framework should apply to companies that provide services or tools that allow, enable or facilitate users to share or discover user-generated content, or interact with each other online. It noted that regulatory requirements would need to be flexible, risk based and proportionate to the needs of different businesses and that search engines would be in scope of the regulatory framework.

#### **Stakeholder Feedback**

Since the White Paper, the government has undertaken significant consultation with a broad range of stakeholders. They were broadly supportive of the proposed approach. Although many did express a need for greater clarity around which organisations would be in scope. There were also calls for business-to-business services to be exempt from scope due to the lower risk of harm on those services.

#### **Interim Response**

Following the consultation, the Initial Government Response confirmed that only a small proportion of UK businesses (estimated at the time to be less than 5%) are likely to fall within the scope of the regulatory framework. It also confirmed that business-to-business services would be out of regulatory scope.

#### **Final Policy Position**

The Full Government Response confirmed that the government will be maintaining the broad regulatory scope. Companies will therefore fall into scope if they provide services which:

- a): Host user generated content which can be accessed by users in the UK; and/or
- b): Facilitate private or public interaction between service users, one or more of whom is in the UK.

This scope covers a broad range of services, including (among others) social media services, consumer cloud storage sites, video sharing platforms, online forums, dating services, online instant messaging services, peer-to-peer services, and video games that allow interaction with other users and online marketplaces.

Only companies with direct control over the content and activity on a service will be subject to the duty of care. This means that business-to-business services will remain outside of the scope of the regulatory framework. It also means that services which play a functional role in enabling online activity will remain out of scope, including internet service providers, virtual private networks, browsers, web hosting companies, content delivery service providers, device managers, app stores, enterprise private networks and security software.

The new regulatory framework will also apply to search engines. While they do not host user generated content or facilitate user interaction, there are significant actions and interventions that they can undertake to reduce the risk of harm.

#### **Scope Changes**

The scope of the regulatory framework has been refined since the publication of the Online Harms White Paper, based on stakeholder feedback and further analysis of the risk posed by different types of services.

---

<sup>188</sup> [Online Harms White Paper](#) (April 2019)

As set out in the White Paper, the regulatory framework will apply to private communication channels, such as online instant messaging services and closed social media groups. However, regulatory expectations for private channels will respect users' higher expectations of privacy.

The government also recognises that many companies offer services which may fall into scope of the regulatory framework, but which pose a very low risk of harm. The government is committed to avoiding undue regulatory burdens on companies, particularly in light of the Covid-19 pandemic, so we will exempt a number of services that have been defined as low risk from regulatory scope. The exemption will apply to specific services rather than entire companies, hence a company which maintains a service that benefits from the exemption will not be exempt from scope if it provides another service which is not exempt from scope.

#### *Exempt Services*

*Business Services:* Online services which are used internally by organisations will be exempt. This includes (but is not limited to) intranets, customer relationship management systems (CRMs), enterprise cloud storage, enterprise tools and enterprise conferencing software.

*Online services managed by educational institutions, including early years, schools and further and higher education providers:* This includes platforms used by parents, teachers, students and alumni to communicate.

*Email and Telephony:* Email communication, voice-only calls and SMS/MMS remain outside of regulatory scope.

#### **Low Risk Functionality Exemption**

In-order to remove low-risk businesses from regulatory scope, the legislation will exempt user comments on digital content providing that they are in relation to content directly published by a service. This exemption will encompass reviews and comments on products and services delivered directly by a company as well as 'below the line' comments on articles and blogs. This approach is designed to avoid imposing unnecessary familiarisation costs on businesses when they are likely to be required to take action to comply with the duty of care. It will also help to ensure and protect freedom of speech.

The detail of the exemption will be set out in primary legislation, with a power in secondary legislation to commence or stop these provisions should they not work (essentially acting as an on/off switch). This approach has been taken because secondary legislation is considerably easier to amend. Should evidence of harm on low risk functionalities emerge, the Secretary of State would retain the right to rescind the exemption, although this action would be subject to the affirmative procedure in order to provide democratic accountability to an expansion in online harms scope.

#### **Journalism and democratic content**

##### **Original Policy Position**

The White Paper committed to ensuring protections for freedom of expression within the regulatory framework. The government reaffirmed its commitment to the protection of media freedom in the 2019 Conservative and Unionist Party manifesto.

##### **Stakeholder Feedback**

There were strong calls from the media to exclude journalistic content from the scope of regulation in order to protect freedom of expression and avoid negatively affecting the public's ability to access information or undermining quality news media.

Journalistic content is shared across the internet as well as on social media, forums and a range of other websites and journalists also use social media to report directly to their audiences. This content is therefore subject to in-scope services' existing moderation processes and is subject to those services' terms and conditions. Media stakeholders have raised concerns that this could result in journalistic content being removed for vague reasons, with limited opportunities for appeal.



### ***Final Policy Position***

Freedom of expression is at the heart of the regulatory framework and was of great stakeholder concern. There will therefore be strong safeguards in place to ensure media freedom is upheld. Content and articles published by news media on their own sites will not be considered user generated content and thus will be out of regulatory scope.

Legislation will also include robust protections for journalistic content on in-scope services. Firstly, the legislation will provide a clear exemption for news publishers' content. This means platforms will not have any new legal duties for these publishers' content as a result of our legislation. Secondly, the legislation will oblige Category 1 companies to put in place safeguards for all journalistic content shared on their platforms. The safeguards will ensure that platforms consider the importance of journalism when undertaking content moderation, and can be held to account for the removal of journalistic content, including with respect to automated moderation tools.

Legislation will also include protections for democratic content on Category 1 providers (see 'Designation of Category 1 services' below). Category 1 providers will have a duty to set policies for such content. These policies must be enforced consistently and transparently.

### **Advertising**

The online advertising ecosystem is complex and includes services that are both within and beyond the scope of the online safety framework.

As the government considers further action in relation to advertising, it will seek to avoid duplication between overlapping areas ahead of future regulatory requirements. Nevertheless, some types of advertising will still fall in the scope of the online safety regulatory framework. The definition of user-generated content will encompass organic and influencer adverts that appear on in-scope services and encompasses images or texts posted from users' accounts to promote a product, service or brand. As these are often indistinguishable from other forms of user-generated content, it is important and reasonable that companies should ensure their systems and processes reduce the risk of harm from such content.

The Advertising Standards Authority will retain responsibility for overseeing the self regulation of advertising more broadly. It will continue to regulate the content of individual adverts and advertisers' compliance with the advertising codes.

Last year, the Secretary of State for DCMS announced a separate review of the way in which the advertising market is regulated in the UK, which is being considered through the online advertising programme. This is considering whether any regulatory gaps in relation to advertising may exist and ensuring that advertising regulation is fit for purpose in a changing advertising landscape. It will consider a range of measures, including the potential for changes to the new regulatory landscape.

## **Part 2-What harmful content or activity will the new regulatory framework apply to, and what actions will companies need to undertake?**

### **Definition of harm**

#### ***Original White Paper Position***

The White Paper set out an initial list of harms in scope but made clear that this was, by design, neither exhaustive nor fixed. As a result it does not provide a single, comprehensive definition of 'harm' that the new regulatory framework will seek to address. A static list could prevent swift regulatory action to address new forms and types of online harm. It also set out specific exclusions from scope when there are existing government initiatives to tackle these harms.

#### ***Stakeholder feedback***

Companies and stakeholders wanted more detail on the breadth of both services and harms in scope with stakeholder concerns that the list of harms was too broad. There were calls to protect freedom of expression and a focus on protecting children. Others suggested that more work should be done to increase education and public awareness of online harms.

### ***Final Policy Position***

The legislation will set out the general definition of content and harmful content that is in scope. This approach will maintain the flexibility to address different types of harm that could potentially arise in future, prioritising the most serious harms, whilst providing clarity and legal certainty for companies and a clearly defined statutory remit for OFCOM.

The legislation will set out that online content and activity should be considered harmful, and therefore be in scope of the regime, where it is illegal or gives rise to a foreseeable risk of a significant physical or psychological impact on individuals. Companies will not have to address content or activity that does not pose a reasonably foreseeable risk of harm, or which has a minor impact on users. Harms to organisations will not be in scope of the regime.

A limited number of priority categories of harmful content posing the greatest risk to users, will be set out in primary and secondary legislation. This will create the higher level of clarity that has been requested from stakeholders and support the prioritisation of companies' and the regulator's efforts. These will cover (i) terrorism offences, child sexual abuse and exploitation offences and other priority categories of criminal offences (for example hate crime and the sale of illegal drugs and weapons), (ii) priority categories of harmful content and activity affecting children, such as pornography and violent content, and (iii) priority categories of harmful content that is legal when accessed by adults, but which may be harmful to them, such as abuse and content relating to eating disorders or suicide.

This list of priority harms is not exhaustive; indeed companies will be expected to address all illegal content and activity (unless explicitly exempted) if it constitutes a UK criminal offence and must therefore take reasonably practicable steps to minimise the risk of such content occurring on their services. For priority offences, companies will need to consider through a risk assessment whether further systems and processes are required to identify, assess and address such offences. In certain cases companies may be required to proactively identify and block this material if other steps have not been effective.

In line (and building on) the position taken in the White Paper, a number of harms will be excluded from scope when there are existing alternative legislative, regulatory and other government initiatives in place. The following will be excluded from scope:

- Harms resulting from breaches in competition law
- Harms resulting from breaches in intellectual property rights
- Harms resulting from breaches in data protection legislation
- Harms resulting from breaches in consumer protection law
- Harms resulting from cybersecurity breaches and hacking.

The online safety regulatory framework will not aim to tackle harm occurring through the dark web. A law enforcement response to tackle criminal activity on the dark web is more suitable than a regulatory approach.

### **Differentiated expectations on companies**

#### ***Original White Paper Policy Position***

The White Paper set-out that all in-scope companies will be required to address illegal and legal but harmful content and activity. It stated that the regulatory approach would impose more specific and stringent requirements for illegal harms than for those harms that are legal but still have the potential to cause harm. It acknowledged that the impact of harmful content can be particularly damaging for children and place particular emphasis for keeping children safe online.

#### ***Consultation responses and stakeholder engagement***

The consultation responses flagged concerns around the broad scope of harms. There were also

concerns around the high level of subjectivity around certain harms, particularly when they are legal with many respondents objecting to the latter being in scope. There were related concerns that the proposals could impact on freedom of speech. Respondents welcomed the approach to the protection of children.

### ***Final Policy Position***

The final policy position has been amended based on stakeholder feedback around clarity and the subjectivity of harm. The regulatory framework will establish differentiated expectations on in-scope companies with regard to different types of content and activities. This will ensure that companies prioritise tackling relevant illegal content and activity on their services and that children are protected from age-inappropriate and harmful content. The differentiated approach can be summarised as follows:

- All companies will be expected to take action in relation to all relevant illegal content and activity,
- All companies will be required to assess the likelihood of children accessing their services and if so, they will be required to provide additional, proportionate protections for children using these services.
- Only companies which are defined as 'Category 1' services will be required to take action with regards to legal but harmful content accessed by adults. This is because services offering extensive functions for interacting and sharing content pose an increased risk of user harm in relation to legal but harmful content. Companies providing 'Category 1' services will be required to undertake regular risk assessments assessing the risks to adult users (including vulnerable users). These companies will be required to set clear and accessible terms and conditions which explicitly state how they will manage priority categories of legal but harmful content set in legislation and any others identified through their risk assessment. These terms and conditions must be enforced consistently and transparently.

### ***Designation of 'Category 1' Services***

Primary legislation will set-out high-level factors which lead to a significant risk of harm occurring to adults through legal but harmful content, these are the size of a service's audience and the functionality offered. The government will then determine and publish thresholds for each of the criteria with OFCOM required to provide non-binding advice to the governments on the setting of thresholds which will be set in secondary legislation. OFCOM will then be required to assess services against these thresholds and publish a register for those services that meet the threshold.

## **Duty of Care**

### ***Original White Paper Policy Position***

The White Paper stated that there would be a new statutory duty of care to ensure companies are responsible for the safety of users on their services. This duty would be risk-based and proportionate and would be focussed on systems and processes, not individual pieces of content. Important principles would apply to the regulatory framework including users rights to freedom of expression and privacy, innovation and protecting small and medium businesses.

### ***Consultation responses and stakeholder engagement***

Many stakeholders welcomed the approach set-out in the White Paper, noting that this would underpin an effective, future-proofed framework. Nevertheless, industry sought greater reassurance about how it would be proportionate in practice, particularly for small and medium sized enterprises; and how flexibility would balance with certainty about what the duty of care requires of companies. Rights groups and industry also emphasised the need for the government to provide more certainty about how safety would be balanced with freedom of expression, particularly in relation to legal but harmful content.

### ***Final Policy Position***

To respond to this stakeholder feedback, the duty of care has been refined. It will now cover content and activity that could cause harm to individuals rather than harms to society more broadly.

The primary responsibility of each company in scope will be to take action to prevent user-generated activity and content on their services causing significant physical or psychological harms to individuals. To do this they will complete an assessment of the risks associated with their service and take reasonable steps to risk the changes of identified harm occurring.

The steps will depend on the risk and severity of the harm occurring, the number, age, and profile of users and company size. Search engines will need to assess the risk of harm occurring across the entire service with OFCOM providing guidance for search engines on regulatory expectations. Companies will need to consider users rights, including freedom of expression, as part of their risk assessments and when they put systems and processes in place.

Companies will fulfill the duty of care by putting in place systems and processes that improve user safety on their services. These will include, for example, user tools, content moderation and recommendation procedures. Companies must also provide users with effective and accessible processes for users to report harm and seek redress. There will also need to be processes for users to challenge wrongful content takedown. OFCOM's codes of practice will set-out expectations for these mechanisms and the proposed Safety by Design Framework will support companies to understand how they can improve user safety through product design choices.

Expectations on companies will be risk-based and proportionate. For example, the smallest and lowest risk companies might only need to give a contact email address whilst larger companies will be required to provide a fuller suite of measures.

## **Codes of Practice**

### ***Original White Paper Policy Position***

The White Paper stated that the independent regulator would set out how companies could fulfill the duty of care in practice and that the government would produce voluntary interim codes of practice on preventing terrorist use of the internet and child sexual abuse and exploitation due to the serious nature of these harms.

### ***Consultation Responses and Stakeholder Engagement***

Some respondents argued that too many codes of practice would cause confusion, duplication and potentially an over-reliance on removal of content by risk-averse companies.

### ***Final Policy position***

OFCOM will have the duty to issue statutory codes of practice that set out the steps companies can take to fulfill the duty of care. The codes will focus on systems, processes and governance that companies need to put in place. Companies may take alternative steps to those set out in the codes of practice, providing that they can demonstrate to OFCOM that those steps are as effective or exceed those set out in the codes.

The government recognises the concerns from stakeholders over the number of codes and confirms that there will not be a code of practice for each category of harmful content and that the regulator will decide which codes of practice to produce, with the exception of the codes on child sexual abuse and exploitation and preventing terrorist use of the internet due to the serious nature of these harms. The interim codes of practice on child sexual abuse and exploitation and preventing terrorist use of the internet were published alongside the full government response.

OFCOM will have a duty to consult with interested parties on the development of the codes of practice, which is consistent with usual regulatory practice. It will also be required to consult bodies, organisations and interests specified in legislation who have specific knowledge and expertise related to policy objectives.

### ***Changes to the position on ministerial sign-off and Parliamentary accountability.***

The White Paper set out that the government would have the power to direct the regulator in relation to the codes of practice on child sexual abuse and exploitation and preventing terrorist use of the internet and would have the right to sign off those codes. For the codes of practice on preventing terrorist use of the internet and child sexual exploitation and abuse the Home Secretary will be able to direct the regulator for reasons relating to national security and public safety. This reflects the Home Secretary's responsibility for national security and the government's response to online child sexual exploitation and abuse. However, the Home Secretary will not have specific sign off power over the codes.

For the other codes, which will not relate to specific harms, the Secretary of State for Digital, Culture, Media and Sport will have the power to issue a direction to reject a draft code in exceptional circumstances for reasons relating to government policy. Ministers would publish the letter of direction to the regulator, which would also set out modifications the regulator must make when revising the code. The power could be used only at the end of the drafting process when the codes are submitted by OFCOM to the Secretary of State for the Department for Digital, Culture, Media and Sport. OFCOM will be responsible and accountable for all the codes of practice.

The objectives for the codes will be set out in secondary legislation (subject to the affirmative procedure) to provide clarity for the framework and scrutiny of the underpinning objectives. The codes will be laid before Parliament using the negative resolution procedure. Under this approach, no debate or vote is required unless either house demands one within 40 sitting days. This approach balances Parliamentary accountability with the need for the process of updating the codes to be responsive to emerging harms and changing technology.

## **Using technology to identify illegal child sexual exploitation and abuse content and activity**

### ***Original White Paper Policy Position***

The White Paper set-out that some private channels would be in regulatory scope, however companies would not be required to scan or monitor for illegal content on these services.

### ***Consultation responses and stakeholder feedback***

Some consultation responses, including industry and civil liberties groups argued that private communications should either fall out-of-scope or be subject to very limited requirements in-order to protect user privacy. By contrast, some online safety organisations and children's charities argue that private communications should be in scope because there is a risk of harmful activity - such as child grooming- on private channels.

### ***Final Policy Position: Requirement to monitor private channels in certain circumstances***

The scope, severity and complexity of child sexual exploitation and abuse is highly concerning and private channels are frequently exploited by offenders. In-light of this, the regulator will have the power to require companies to use automated technology (when it is highly accurate) to identify illegal child sexual abuse and exploitation activity on their platforms including on private channels.

Robust safeguards will be in place to govern the usage of this technology. The regulator will only be able to require the use of tools that are highly accurate in identifying the flagging of legal content for human review. Before the power can be used, the regulator will need to seek approval from Ministers on the basis that sufficiently accurate tools exist. To support transparency, the regulator will report annually to the Home Secretary and will need to lay a report before Parliament on the use of the power, including on the accuracy of available tools.

As a further safeguard, prior to the regulator requiring a company to identify child sexual exploitation and abuse material, they will have to undertake the following steps:

- have gathered evidence which it assesses as demonstrating persistent and prevalent child sexual abuse and exploitation on the service, which the company has failed to address
- be satisfied that no alternative, less intrusive approaches are available to address the problem
- Issue a public notice of the regulators intention to require a company to use automated technology to identify child sexual abuse and exploitation, to ensure users are fully informed.

In exercising the power, the regulator will balance users' rights to privacy and freedom of expression with the rights of children to be protected from sexual exploitation and abuse.

## **Using technology to identify terrorist content and activity on public facing services**

### ***Original White Paper Policy Position***

The White Paper set-out that the regulator would not compel companies to undertake general monitoring of their public services. Instead the regulatory framework would increase the responsibility of online services in a way which is compatible with the European Unions' e-Commerce directive, which limits their liability for content until they are aware of its existence, and they have failed to remove it from their services in good time. However, the White Paper noted that there was a strong case for mandating specific monitoring for tightly defined categories of illegal content when there is a threat to national security and the safety of children and that the use of automated technology was a reasonable step that companies could take in advance of legislation to remove illegal content and activity.

### ***Consultation responses and stakeholder engagement***

Industry welcomed the commitment to maintaining existing intermediary liability provisions set-out in the e-commerce directive, including the prohibition of illegal monitoring.

### ***Final Policy Position: Requirement to monitor public channels for terrorist content where possible***

Companies liability for specific pieces of content will remain unchanged. Once a company is aware of illegal content, it will be required to remove it quickly otherwise it could be deemed liable for that content.

Regarding the usage of technology to identify illegal content; the regulator will be given an express power in legislation to require a company to identify and remove illegal terrorist content for their public channels when this is the only effective, proportionate and necessary action available. There will be identical safeguards in place regarding the use of this power in relation to the use of technology to identify child sexual exploitation and abuse material, these are: (i) that the technology is highly accurate, ii) that there is evidence of persistent and prevalent terrorist activity and iii): there is no reasonable alternative. This will ensure that the approach taken is proportionate, protects national security and protects users online rights.

## **Data retention and reporting to law enforcement**

### ***Original White Paper Policy Position***

The White Paper stated that the regulator would provide specific guidance in its Codes of Practice on the content that companies should preserve following removal and for how long. It also set-out that the regulator should provide guidance on whether companies should proactively alert law enforcement and other relevant government agencies about specific illegal content.

### ***Consultation responses and stakeholder engagement***

Stakeholders, including the National Crime Agency and National Centre for Missing and Exploited Children, argued that there should be a new, mandatory reporting system for child exploitation and sexual abuse to increase reporting and standardise the reporting approach.

### ***Final Policy Position: Reporting of Child Sexual Abuse and Exploitation on Services***

In-line with the White Paper, the regulator will also set-out in the preventing terrorist use of the internet and child sexual abuse and exploitation codes of practice, the reasonable steps that companies that companies should take to retaining data and reporting these types of content. This would include guidance on how long companies should retain data for and the circumstances in which this should be reported to law enforcement.

Following the White Paper consultation and further engagement with law enforcement and other agencies, the government is minded to introduce a requirement on companies to report child sexual exploitation and abuse identified on their services. Further works This would be a stand-alone legislative requirement rather than part of the duty of care. Further work is being undertaken to explore a suitable body to receive these reports and ensure that the system does not duplicate existing reporting requirements.

## **Disinformation & Misinformation**

### ***Original White Paper Position***

The White Paper did not set out a definitive position on how disinformation and misinformation would be addressed under the regulatory framework. Disinformation was included in an indicative list of harmful content and activity that would be in scope of the legislation, because it can be harmful to both individuals and society.

### ***Consultation response & Stakeholder engagement***

A range of stakeholders, including civil society organisations, raised concerns about including disinformation within regulatory scope because of the impact on freedom of expression. Many stakeholders are concerned about the threat that disinformation and misinformation poses to users as well as the broader impact on public safety, national security and community cohesion.

### ***Final Policy Position***

The government recognises the threat of mis and disinformation, particularly in-light of the Covid-19 pandemic and has undertaken a number of different interventions to respond to this threat. This includes the standing-up of a Department for Digital, Culture, Media and Sport led Cross-Whitehall Counter Disinformation Unit, to provide a comprehensive picture of the extent, scope and reach of disinformation and misinformation.

In relation to the regulatory framework, the duty of care applies to content or activity that could cause significant physical or psychological harms to individuals but not to society in general. Under our proposals, disinformation and misinformation will be in scope of the duty of care so far as the content or activity could cause harm to individuals. An example of this could be content which suggests that users should go against established medical advice, such as avoiding vaccinations. Harms to society will not be part of online harms legislation in general.

As the majority of misinformation and disinformation is not illegal, only companies providing Category 1 services will be required to deal with it. As with other legal but harmful content, these companies will need to make clear what content is acceptable on their services in their terms and conditions and will be required to enforce this. Companies whose services are likely to be accessed by children will also need to take steps to protect these users from harmful misinformation and disinformation.

## **Part 3-The Regulator**

### **Body (new vs. existing regulator) and identity of regulator**

#### ***Original White Paper Position***

The White Paper stated that the regime will be overseen and regulated by an independent regulator. It also explained that the government would consider whether a broader restructuring of the regulatory landscape would minimise the risk of duplication and minimise the burden on businesses. The White Paper also stated that the regulator will be an independent body that can command public confidence in its independence, impartiality, capability and effectiveness.

#### ***Consultation responses and stakeholder engagement***

Engagement with stakeholders emphasised the need for there to be consistency between existing and new regulatory regimes, and for the regulator to be equipped to function effectively. Views on the identity of the regulator were balanced, highlighting the benefits and risks of a new body over an existing one.

#### ***Interim Response Position***

The government has examined a range of options including creating a new body or appointing an existing regulator. These options were assessed against a range of key criteria, including effectiveness, efficiency and strategic coherence and were informed by feedback from the consultation response.

In the Interim Response published in February 2020, the government announced that it was minded to give OFCOM the role of the independent online harms regulator. This was based on its organisational experience, robustness and experience of delivering, whilst holding challenging, high profile remits across a range of sectors. OFCOM also offers a strong strategic fit given its role regulating activities that

are increasingly related to online harms and their new responsibilities in regulating UK based video sharing platforms under the Audiovisual Media Services Regulations 2020.

### ***Final Policy Position***

Ministers have decided to confirm the appointment of OFCOM to the role of independent online harms regulator, subject to the passage of legislation. OFCOM will use the time prior to the passage of legislation to further its engagement with in-scope companies and prepare organisationally for taking on its role as the independent online harms regulator. It will be able to set out the expectations on companies and ensure a fuller understanding of compliance ahead of the duty of care coming into force.

As part of this, OFCOM will need to ensure that it has the right skills and expertise to discharge its responsibilities and will be building its capability, particularly in areas of emerging technology. OFCOM has already created an emerging technology directorate and a data science team.

## **Governance, capabilities and infrastructure**

### ***White Paper Policy Position***

The White Paper stated that the regulator will be an independent body and that the government will take steps to ensure that it can command public confidence in its independence, impartiality, capability and effectiveness.

### ***Consultation responses and stakeholder engagement***

Most responses viewed an independent and empowered regulator as critical to the delivery of the regime. There were no particular views on the regulator's governance arrangements.

### ***Final Policy Position***

The independence of regulators from undue influence - from government, other political sources and organisations with an interest in the regulation - is a crucial element of effective regulation. The relationship between government and the regulator will be clearly defined in legislation along with the scope of the regulatory regime and the remit of the regulator. OFCOM's founding legislation already provides it with a high degree of regulatory independence as it is operationally independent from government, giving it the statutory provisions to manage its own affairs.

In some areas, such as the production of the codes of practice and the threshold for companies in scope to pay the annual fee, the government will maintain levers to ensure that the policy intent of the regime is upheld. The government will introduce a power to allow the Secretary of State for Digital, Culture, Media and Sport to issue guidance to the regulator, with clearly defined scope and use. This will allow the government to set out further detail on regulatory processes, but will not stray into operational matters or limit OFCOM's independence. The final version of this guidance will be subject to parliamentary approval.

There will also be an option for the DCMS Secretary of State to issue a Statement of Strategic Priorities in relation to the regulatory framework. This will allow the government to be clear on the overall strategic direction for tackling online harms and to respond at a high level to future changes. The statement will require external consultation, including with OFCOM, and approval by Parliament.

## **Accountability to Parliament**

### ***White Paper Policy Position***

The White Paper was clear that it was important to ensure that Parliament is able to scrutinise the regulator's work.

### ***Consultation responses and stakeholder engagement***

Responses to the consultation showed strong support for Parliamentary oversight of the regulator. Most stakeholders agreed that Parliament should not interfere with the regulator's independence in drafting codes of practice. Several responses suggested a dedicated body for reviewing codes.



### ***Final Policy Positions***

The regulator will be accountable to Parliament. OFCOM will lay its annual report and accounts before Parliament and be subject to select committee scrutiny.

Parliament will also have a role in approving a number of different aspects of the regulatory framework through its scrutiny of secondary legislation. This will include statutory instruments establishing the objectives set by the government for the codes of practice, the codes of practice themselves and the priority categories for harm.

The Secretary of State for Digital, Culture, Media and Sport will undertake a review of the effectiveness of the regime 2-5 years after entry into force, producing a report that will be laid before Parliament and will be subject to a debate. OFCOM will also be required to conduct and publish impact assessments for proposals which affect businesses. This will include the codes of practice but may also include other policy areas such as enforcement, information gathering, transparency, super complaints, media literacy and funding. The regulator will also have a specific duty to assess the impact on small and micro businesses. OFCOM will also be required to consult on impact assessments and report on the impact assessments it has undertaken in its annual reporting to Parliament.

### **Regulator funding model**

#### ***White Paper Policy Position***

The White Paper and the Initial Government Response both outlined that the regulator would be funded by industry in the medium term. The government indicated that it would consider a range of options to fund regulator activity including fees, charges and levies on services in scope.

#### ***Consultation responses and stakeholder engagement***

There was broad agreement amongst stakeholders and consultation respondents that funding for the regulator should largely come from industry. However, it was also felt that the model should be proportionate and practical, in particular through minimising the cost on smaller businesses and ensuring efficient collection of contributions from overseas companies.

#### ***Final Policy Position***

The framework for regulator funding has been significantly developed since the White Paper. OFCOM will be able to raise the required income to cover the costs of running the online safety regime through industry fees. OFCOM will also have the power to require a company to undertake, and pay for, a skilled person report.

Companies at or above a threshold based on global annual revenue will be required to notify the regulator and pay an annual fee. Companies below the threshold will not be required to notify the regulator or pay a fee. The threshold will be set by OFCOM, based on consultation with industry, and will be signed off by Ministers. In-scope companies that fall below the threshold will still have to comply with their other regulatory responsibilities.

The total amount of fees charged to industry will be in proportion to the costs incurred by the regulator in operating the online safety regime. The fees to be paid by individual companies will be tiered. It is expected that the regulator will calculate the fees based on two metrics: a primary metric of global annual revenue and a secondary optional metric based on company activity.

The regulator will also have the power to request a skilled person report on specific issues of concern. When the regulator uses this power, the company under investigation will always be required to cover the direct costs of the skilled person report. The regulator will consider the use of alternative powers to obtain the information it needs, if it determines that paying for the skilled person report could have an adverse financial effect on a company.

### **Interface with other bodies**

#### ***White Paper Policy Position***

The White Paper stated that the government and regulator should work closely with a range of other organisations, both domestic and international, to ensure the successful implementation of the regime. For example industry bodies, other regulators, law enforcement and overseas bodies.

### ***Consultation responses and stakeholder engagement***

Stakeholders showed strong support for consultation and cooperation across regulators. The emphasis was on UK based regulation as the question did not reference international engagement.

### ***Final Policy Position***

OFCOM will work with a range of organisations and stakeholders to deliver the online safety regime and the government will work closely with the regulator to ensure this happens. The relationship will be delivered through a range of means, including co-designation powers, memorandums of understanding, forums and networks.

OFCOM already has a strong network of relationships with other bodies and will continue to cultivate these at home and internationally. Furthermore, OFCOM will play a critical role in enforcement across borders, and will use its good relationships with its international counterparts to facilitate information sharing from other jurisdictions and to achieve a degree of international regulatory alignment. The decision to appoint OFCOM as the online harms regulator is part of the wider programme of work to ensure that the regulatory landscape for digital technologies is coherent, effective and efficient.

## **Part 4- Functions of the regulator**

### **Duties on the functions of the regulator**

OFCOM as the regulator will have certain duties and functions under the framework. It's primary duty will be to improve the safety of users of online services. Regulatory action should be taken in line with the principles of the regulatory framework (set-out in **part 2**).

OFCOM will need to apply the principles of the regulatory framework when it issues code of practice which will set out the steps that companies can take to fulfill the duty of care.

OFCOM will also have to pay due regard to innovation in the exercise of all of its functions and it will have further responsibilities to help all companies to understand and fulfill their responsibilities. This will involve providing appropriate support to companies depending on their size and maturity and providing greater support for small and medium sized businesses. It will also be required to assess the impact of its regulatory activity on its businesses, and in particular small and micro businesses. OFCOM is also subject to the Public Safety Equality Duty and can be expected to pay due regard to the requirements of the Equality Act 2010 assessing the impact of its regulatory activity on user groups with protected characteristics.

### **Promoting Innovation**

#### ***White Paper Policy Position***

The White Paper proposed that the regulator should have a legal duty to pay due regard to innovation, to ensure competition within the regulatory market and to help companies find more efficient ways of working with the regulator.

#### ***Consultation responses and stakeholder engagement***

Engagement suggested that there is significant appetite for government influence and oversight to support innovation.

#### ***Final Policy Position- No required legal duty to promote innovation***

OFCOM, as the independent online harms regulator, must already pay 'due regard' to 'encouraging innovation and promoting competition in relevant markets' when performing its duties as set-out in section 3(4)(d) of the Communications Act 2003. Additional requirements to pay regard to innovation

with regards to online harms would therefore be delivered through practical measures and thus no additional legal power will be required.

## **Transparency**

### ***White Paper Policy Position***

The White Paper sets out that developing a culture of transparency, trust and accountability will be a crucial element of the new regulatory framework. It stated that in-scope companies will be required to publish annual transparency reports. These will cover the prevalence of harmful content on services and the actions being taken to resolve them.

### ***Consultation responses and stakeholder engagement***

Responses to the consultation and White Paper highlighted the importance of transparency in holding companies to account for enforcement of their own standards and upholding freedom of expression. Industry suggested that transparency requirements should be proportionate - noting that a 'one size fits all' approach was unlikely to be effective.

### ***Final Policy Position***

The future transparency reporting requirements are in-line with the position set-out in the White Paper. Companies providing Category 1 services will be required to publish reports containing information about the steps that they are taking to tackle online harms on their services. The Secretary of State for Digital, Culture, Media and Sport will also have the power to extend the scope of companies who will be required to publish transparency reports beyond Category 1 services, by setting additional thresholds based on factors such as the audience and functionalities of the service.

To ensure that the transparency framework is proportionate and reflects the diversity of different companies, the reporting requirements will differ between different types of companies. OFCOM will consider companies' resources and capacity, service type and audience in determining the type of information they need to include in their reports. To support this, the regulator will have flexibility in determining the specific information that companies need to provide (which is still to be determined).

When the regulator has determined that a company should report and set out what they will need to report on, the company will be required to do so or face enforcement action. Companies will be required to publish their reports and make their reports accessible.

The regulator will also be responsible for producing an annual report of its own which will summarise key findings and insights from the reports that companies have produced and highlight best practice. This will play a vital role in helping users and parents highlight differences between different platforms and make informed decisions about which to use.

## **Part 5 - Information gathering and investigation powers**

### ***White Paper Position***

The White Paper sets out that the transparency, trust and accountability framework would be backed by robust information gathering powers, to enable OFCOM to assess companies compliance with the duty of care and develop its own understanding of the risk landscape.

### ***Companies responses and stakeholder engagement***

Respondents to the consultation did not answer specifically on information gathering and investigation powers but highlighted the importance of transparency in holding companies to account for the enforcement of their own standards.

### ***Final Policy Position***

The regulator will have broad information gathering powers to allow it to carry out its functions similar to its existing power for regulating telecommunications (S135 of the Communications Act, 2003). This will give OFCOM the flexibility to determine the specific information that it requires. It will be required to take a proportionate approach when exercising these powers. This power will apply to all companies within

the scope of the duty of care and where necessary to other organisations and persons who have relevant information.

OFCOM will be able to use information from a wide range of sources to support its investigations and enforcement activity. Alongside the information that companies have provided (in their transparency reports and in response to information requests), the regulator will also use complaints data and publicly available information to help determine whether an investigation might be warranted.

The regulator will have additional powers of investigation to support its oversight and enforcement activity. When there are reasonable grounds to suspect that a company may be non-compliant, OFCOM will have the power to enter companies premises and access documentation, data and equipment in order to understand whether companies are taking sufficient measures to fulfill the duty of care. The regulator will also have the power to interview employees.

Finally, OFCOM will also have the power to require a company to seek a skilled person report on specific issues of concern. This power will be particularly useful in areas where specific technical expertise is needed, for instance to validate the effectiveness of automated moderation systems. As with all its powers, OFCOM will be required to take a proportionate approach to this power.

### ***Researcher access to company data***

To support research into online harms and to help the regulator prioritise its actions, OFCOM will be required to conduct research and produce reports on the opportunities, challenges and practicalities of companies providing independent researchers with access to online harms information. As part of this, OFCOM will produce best practice guidance for companies and researchers on how to approach providing independent researchers with access to company information for research into online harms. In preparing this guidance, OFCOM will be required to consult a range of stakeholders including companies, academics, the Information Commissioner's Office, the Centre for Data Ethics and Innovation and UK Research & Innovation.

## **User redress**

### ***White Paper Policy Position***

The White paper committed to ensuring that measures are in place for users to seek redress, and consulted on the proposed super-complaints framework. It also noted that users would be able to alert the regulator to their concerns, and use regulatory decisions in legal proceedings.

### ***Consultation responses and stakeholder feedback***

Organisations overwhelmingly agreed that companies should have effective, accessible and transparent mechanisms for reporting harmful content and felt that current processes fell short with a patchy approach across industry. They agreed that this process should start with reports directly to the service, and noted the importance of making these processes accessible and prominent to all users, especially children.

### ***Final Policy Position***

The response reaffirmed the position set-out in the White Paper, companies will be required to have effective and accessible user reporting mechanisms for the type of content that they have to address as part of their duty of care and they will also be required to have mechanisms for users to report broader concerns about a company's compliance with its duties. The regulator will be able to access information about a company's reporting and redress mechanism as part of its statutory functions. Users will also be able to report their concerns to the regulator, however the regulator will not be able to investigate or arbitrate on individual cases as allowing the regulator to do this would be in contravention of the systems and processes approach and is likely to place undue strain on the regulator. Receiving user complaints will instead be part of OFCOM's horizon scanning, research supervision and enforcement activity.

The existing legal rights for individuals to bring action against companies will not be affected by the new regulatory framework. Users will also be able to use regulatory decisions that are already available as evidence in any legal action that they pursue.

### ***Super Complaints***

As proposed in the White Paper, a super complaints function will ensure that there is an avenue for organisations representing users or those who are affected by harmful content and activity online to alert OFCOM to their concerns about systematic issues.

Under this function, primary legislation will require OFCOM to accept super-complaints demonstrating evidence of a systemic issue that is causing harm, or risks causing harm, to large numbers of users or specific groups of users. This will include those who may suffer disproportionately from online harms. Super-complaints will need to focus on the systems and processes that companies have in place, rather than any specific content issues. The government expects super-complaints to concern issues across multiple in-scope services, as companies can raise concerns about a single company's conduct through OFCOM's enforcement complaints service. However, recognising the dominance of some services, super-complaints regarding one service will be admissible in certain circumstances.

In the event of a super-complaint, the regulator will have a legal duty to assess eligible super complaints and publish a reasoned response within a set time period. It will not have to carry-out a full investigation in this time, although it may choose to do so. In specific circumstances, including when relevant information is not forthcoming, the regulator may choose to temporarily pause this process to ensure it's response is well-informed and accurate. The reasoned response will set out what action the regulator plans to take, or to explain why it doesn't intend to take action. The details of this process will be consulted on and set out in secondary legislation. The criteria organisations must meet to be eligible to submit a super-complaint will also be set out in secondary legislation.

### ***User Advocacy***

OFCOM will also have a legal duty to establish ongoing mechanisms for user advocacy. This will ensure that the regulator understands the experiences of service users (including children) and others who are affected by harmful content and activity and that it can take action to address their concerns. It will also help OFCOM to become aware of issues at an early stage before they can cause significant harm. OFCOM will be required to report on its user advocacy work in its annual report to Parliament.

## **Enforcement**

### ***White Paper Policy Position***

The White Paper set-out that the regulator would have a range of enforcement powers to take action against companies that fail to fulfill the duty of care. It recognised that the powers must incentivise compliance and be used in a proportionate manner.

The White Paper also proposed that companies should have a nominated representative in the UK or European Economic Area, to assist the regulator in taking enforcement actions outside of these areas. The White Paper also consulted on whether senior managers should be personally liable for failures to meet the duty of care.

### ***Consultation responses and stakeholder engagement***

Stakeholder feedback expressed an overall preference for the regulator to begin its operations by supervising companies and supporting compliance through advice, and that any future enforcement measures should be used proportionately and follow a clear process.

### ***Final Policy Position***

The principles and objectives underlying the enforcement proposals have not fundamentally changed, but the government has provided further detail on what enforcement activity will look like. This includes a refinement of the additional enforcement powers that the government has consulted on.

The government recognises the need to balance enforcement with protecting the attractiveness of the UK's technology sector. The approach to enforcement will aim to both encourage compliance and drive cultural change. OFCOM will have a suite of enforcement powers to use against companies who fail to fulfill the duty of care. OFCOM will use its enforcement powers in line with its duties and will ensure that they are used proportionately, taking into account the level of harm as well considering the impact on children.

The regulators enforcement powers are:

- issuing notices of non-compliance;
- issue civil fines of up-to £18 million or 10% annual global turnover, whichever is higher;
- take measures to disrupt a company's business activities in the UK to make it less attractive for a non-compliant company to provide services in the UK, this includes restricting access in the most serious circumstances.

If a company fails to fulfill the duty of care, the regulator may also pursue a parent company that has sufficient control over the non-compliant company. The regulator may also be able to pursue enforcement action against subsidiaries of parent companies, where they are involved in the breach of a group company.

OFCOM will be able to take enforcement action against companies that fail to fulfil the duty of care. In addition, their enforcement powers will also extend to other requirements in the regulations that do not fall under the duty of care. These are for example:

- Failing to register with the regulator (when required)
- Failing to pay the industry fee (when required)
- Failing to provide transparency reporting (when required)
- Failing to respond to information requests
- Failing to cooperate with the regulator, or the skilled person, in relation to the regulator's use of its power to require a skilled person report
- Failing to comply with a use of technology notice

OFCOM will also be able to levy fines against any party that fails to comply with information requests, whether they are in scope of the regulations or not.

Recognising the global nature of online harms, these powers have been designed to be international in-scope meaning that it will be possible for the regulator to take enforcement action against any company, irrespective of where it is based in the world or if it has a physical presence in the UK, if it provides services to UK users. As other countries introduce similar legislation, international cooperation will become an increasingly important tool for the regulator and the government expects the regulator to work internationally to help foster collaboration.

### ***Changes since the White Paper- Nominated representative and senior management liability***

Concerns were raised by consultation respondents in relation to the impact on business and operations of the nominated representatives proposal which would be particularly acute in relation to small businesses. The government has therefore decided not to proceed with this option.

In relation to senior management liability, stakeholders and consultation respondents highlighted the potential negative impacts on the attractiveness of the UK technology sector, and the lack of clarity about what would be expected of a senior manager. Therefore the government will not introduce a general senior manager liability. It will be crucial that the regulator has access to reliable and timely information, and senior managers have an important role to play in ensuring that these requests are met. The government will therefore reserve the right to introduce criminal sanctions for managers who fail to respond fully, accurately and in a timely manner to information requests from the regulator. This power would not be introduced for at least two years after the regulatory framework comes into effect, based on a review of the impact of the framework.

## **Appeals**

### ***White Paper Policy Position***

The White Paper set-out that companies and other individuals will have the ability to seek judicial review of the regulators actions and decisions through the High Court, to provide assurance that the regulator is acting fairly and within its powers. The government also consulted on whether there should be an

additional statutory mechanism of appeal, who should be able to access this, and what the circumstances and standards of appeal should be.

### ***Consultation responses and stakeholder engagement***

Responses were broadly in support of a statutory mechanism in addition to judicial review, with a primary focus on it being affordable and accessible.

### ***Final Policy Position***

The government will now ensure that, in addition to judicial review through the high court, there is an additional statutory mechanism of review by designating an existing statutory body to review appeals. This responsibility will be given to the Upper Tier Tribunal: Administrative Appeals Chamber, part of HM Courts and Tribunal Service. By using an additional statutory appeals body, the regime will seek to save costs and reduce the financial burden on smaller businesses and third parties who wish to appeal decisions. Appeals will be heard by applying the same principles as a judicial review, rather than on the merits of the case.

Any party with a sufficient interest in the matter to which the appeal relates will be able to appeal OFCOM's enforcement decisions and sanctions, in line with judicial review standards. OFCOM decisions to either allocate a business as a Category 1 service, require a business to provide transparency reports, or to give a use of technology notice, may also be appealed by affected service providers.

## **Part 6 - What part will education, technology and awareness play in the solution?**

### **Safety Tech Market**

#### ***White Paper Position***

The White paper set-out the government's ambition to position the UK as a world leader in safety technology. It proposed specific actions to assess the online safety sector's capability and potential, and to explore how organisations can securely access training data to develop artificial intelligence solutions whilst ensuring its use is safe and ethical.

#### ***Consultation responses and stakeholder engagement***

The government also consulted with a range of stakeholders from across industry and civil society, to understand the potential for growth of the safety technology sector. Key themes and opportunities that emerged for the government were:

- To support a data infrastructure that enables greater competition and innovation in safety technology, for example through improving access to datasets that can be used for training artificial intelligence solutions.
- Champion the emerging safety tech sector, including through international trade and improving company's access to funding.
- Strengthen networks for collaboration between the safety tech sector and the wider technology sector and use insights from sector providers to inform policy development.

#### ***Final Policy Position***

Since the White Paper, the government has conducted a detailed study into the safety technology market. These findings, published in the 'Safer Technology, Safer Users: The UK as a world leader in Safety Tech' report in May 2020.<sup>189</sup> This demonstrated that UK safety tech providers are at the cutting edge of technological development, offering products that are helping to protect millions of users worldwide.

The safety tech market is also becoming increasingly attractive for investors with an annual growth rate of 35% in recent years and with revenues expected to exceed £1bn by the mid 2020s. UK companies

---

<sup>189</sup> DCMS May 2020 '[Safer Technology, Safer Users: The UK as a World Leader in Safety Technology](#)'

have around 25% of market share and the sector employs 1,700 full time employees in the UK with regional hubs in London, Cambridge, Edinburgh and Leeds.

The government has also supported the launch of the UK Safety Technology Association (OSTIA), a collective voice for the safety tech sector which will help to increase the visibility of new innovations and technology. In August 2020, DCMS and the Department for International Trade published a directory of UK Safety Tech Providers, designed to help open-up export markets.<sup>190</sup>

### ***Upcoming Measures***

The government will continue to explore a range of measures to support the rapid development of the safety tech market, these are:

#### **New Measures to support the growth of the safety tech sector**

- Launch a Safety Tech Innovation Network, the world's first forum allowing safety tech providers to collaborate and promote their work.
- Deliver a new £2.6 million project to prototype how better uses of data can lead to improved artificial intelligence systems and deliver better outcomes for citizens.
- Organise a series of events, including the Safety Tech Unconference and Expo, to raise awareness and showcase safety tech to potential buyers.
- Collaborate across sectors, including the UK Online Safety Tech Industry Association (OSTIA) to identify opportunities for innovation, adoption and the promotion of safety tech.
- Explore ways in which best practice for online safety can be included in standards and guidance for buying, building and reusing government technology.
- Develop a Safety Tech Sector Strategy, to guide future priorities for sector support.

## **Safety by design**

### ***White Paper Position***

The government is committed to developing a Safety by Design Framework to make it easier for start-ups and small businesses to embed safety during the design and development phase for new products and services.

### ***Consultation responses and stakeholder engagement***

Stakeholders expressed broad agreement that safety standards are improved when organisations build-in safety and the design stage. It was felt that greater guidance was needed on this, particularly for smaller companies. It was also noted that there was a significant gap in resources focussed on product managers, designers and developers.

### ***Final Policy Position***

The government remains committed to supporting the safer design of online products and services. We intend to publish the first phase of guidance by Spring 2021 to provide support for start-ups and SMEs. This guidance will enable businesses to adopt a safety by design approach, allowing them to consider the impact of their choices at each stage of the design and development process. The Safety by Design Framework will also contain:

- High-level design principles to guide product design and development work;
- Practical guidance for implementing safer design choices and safety features; and

<sup>190</sup> [Directory of UK Safety Tech Providers](#)



- Examples of best practice and case studies on platform design.

The guidance will recognise the different experiences, needs and technical capabilities of businesses in scope of the online harms regulatory framework. As part of the development of the framework the government will engage with companies of different sizes, capabilities and sectors to develop and user-test the guidance. We will continue to develop the guidance, in line with our programme of research on platform design. Future phases of the guidance will produce advice tailored to the roles within an organisation that have a responsibility for platform design, this includes product managers, designers and developers. Government will also work closely with industry, technical experts, academia and civil society to ensure that the right approach is adopted. Responsibility for promoting Safety by Design will ultimately pass to the regulator.

## **Media Literacy**

### ***White Paper Policy Position***

The government committed to developing an online Media Literacy Strategy to set out a strategic and coordinated approach to media literacy education for all UK citizens. It further set out that both industry and government have a shared responsibility to empower users with the media literacy skills and knowledge required to make informed decisions about their online safety. Government intended for the regulator to have oversight on industry activity and spend on media literacy education and awareness, and would hold a responsibility to promote online media literacy.

### ***Consultation responses and stakeholder engagement***

While some respondents felt that that the regulator should not have a role in education and awareness, others made a range of suggestions on how the regulator should take specific action. These included overseeing industry activity and spend, creating an evaluation framework for assessing activity and promoting awareness of online safety.

### ***Final Policy Position***

The government recognises the vital role education plays in mitigating online harms by empowering citizens to make safer choices online. Despite progress made by education providers and tech companies in improving media literacy rates in the UK, the government recognises that more needs to be done, and that there are some key challenges that need to be addressed.

As part of this, the government is committed to publishing an Online Media Literacy Strategy following broad consultation with stakeholders. The strategy will review the existing media literacy landscape, identify areas for change, and set out the government's plans to ensure a strategic and coordinated approach to media literacy education for all UK citizens.

### ***Role of the Regulator***

The online safety regime will build on OFCOM's existing responsibilities (under section 11 of the Communications Act, 2003) to promote media literacy, which are currently delivered through the OFCOM led 'Making Sense of Media' research programme. The online harms legislation will strengthen this duty by placing more responsibilities on the regulator to play an enhanced role in improving media literacy for UK users which could include delivering a number of initiatives such as communications campaigns, piloting targeted interventions, and providing training. These responsibilities will include:

- Taking steps to improve media literacy levels of members of the public;
- Encourage the development of technologies that will support media literacy;
- Promote a greater understanding of media literacy, including the public's media literacy knowledge and skills through research; and
- Support and encourage the evaluation of media literacy initiatives, including service design choices and educational programmes through the development and maintenance of a media literacy evaluation framework.

## Annex D: Business stakeholder survey questions

**Instructions:** *The Online Harms White Paper set out the government's ambition to make the UK the safest place in the world to go online, and the best place to grow and start a digital business. It described a new regulatory framework establishing a duty of care on companies to improve the safety of their users online, overseen and enforced by an independent regulator.*

*The government set out the results of the formal consultation and clarified its direction of travel in the initial government response published in February 2020. The initial government response reconfirmed the commitment to the duty of care approach set out in the White Paper and announced a number of further measures to increase proportionality and protect freedom of expression. It also indicated that the government was minded to appoint OFCOM as the regulator.*

*The government has continued to develop its policy proposals since February and has made further, important changes. The full government response - published on the 15th December - confirms that OFCOM will be named as the regulator in legislation, and sets out the intended policy position. The full government response can be found [here](#).*

*Post-publication of the full government response, we are reaching out to a number of key stakeholders to understand in greater detail how organisations and platforms are preparing for regulation and any costs associated with the preparations.*

*All information provided will remain anonymous and no active attempt will be made to identify individual respondents. While it may theoretically be possible - for the member of staff analysing responses - to identify the respondent (without actively attempting to), we will make every effort to minimise this risk to near zero, including by analysing cost questions separately from other questions.*

*The information gathered will not be published as a standalone piece of research and will instead be used to improve the government's evidence base on actions taken by organisations to ensure compliance with the online harms regime, including in its published assessment of the impacts of regulation.*

*We encourage respondents to engage meaningfully with this survey and welcome any information that can be provided.*

*If you are unable to complete the survey in one go, your responses will be saved as long as you return to the survey using the same computer and do not clear browser cookies.*

*No personal data is being collected in this survey.*

**Question 1:** *If applicable, how many active users do you currently have? If possible, please also include information on active UK users. Please skip this question if the term 'user' does not apply to your platform or service.*

**Question 2:** *If applicable, how has the number of active users on your platform changed over the last 12 months? Please skip this question if the term 'user' does not apply to your platform or service.*

**Question 3:** *If applicable, how has the number of active users on your platform changed over the last 12 months? Please skip this question if the term 'user' does not apply to your platform or service.*

**Question 4:** How many unique visitors does your platform get? *Where possible, please provide unique visitor numbers for the last 12 months. However, if you are providing unique visitor numbers for a different time-period, please specify in the box provided.*

**Question 5:** How many people does your organisation employ?

**Question 6:** Is your platform likely to be accessed by children?

**Instructions:** We are now going to ask some questions about the type of user-generated content and peer-to-peer interaction that is enabled on your platform. This information **will not** be used to attempt to identify individual respondents and all answers in this survey will remain anonymous. This information will only be used to categorise and group responses to better understand the actions taken and costs incurred by different types of platforms/organisations.

**Question 7:** With regard to uploading or broadcasting a users' own content, i.e. posting, does your platform enable any of the following? Please select all that apply.

**Question 8:** With regard to sharing content that already exists on the platform, does your platform enable any of the following? Please select all that apply.

**Question 9:** With regards to reacting to content, does your platform enable any of the following? Please select all that apply.

**Question 10:** With regards to sending messages to others, does your platform enable any of the following? Please select all that apply.

**Question 11:** With regards to video or voice calling, does your platform enable any of the following? Please select all that apply.

**Question 12:** With regards to commenting on content, does your platform enable any of the following? Please select all that apply.

**Question 13:** With regards to creating links between bits of content and people or places, i.e. tagging, does your platform enable any of the following? Please select all that apply.

**Question 14:** With regards to users discovering content, does your platform enable any of the following? Please select all that apply.

**Question 15:** Does your organisation currently have a dedicated online user-safety team or function within the business?

**Question 16:** Does your organisation have a separate regulatory or legal compliance department?

**Instructions:** The rest of this survey aims to understand the specific actions - if any - organisations are taking to prepare for the online harms regime. Throughout the remaining questions, 'actions' refers to systems, controls and processes relating to protecting users from online harms, including things like risk assessments, terms and conditions, content moderation, and other actions organisations may take to ensure the safety of their users.

We are aiming to better understand actions organisations currently take to protect users from online harms, any actions taken since the publication of the Online Harms White Paper (April 2019), and any actions organisations expect to take in the future to ensure compliance.

We are aware that in recent times, society has placed a greater value on online safety and that organisations and platforms have responded by investing in this area. While natural progress in this area is interesting in and of itself, this survey focuses on actions (and associated costs) taken by organisations as a direct result of the Online Harms regime, rather than actions which would have been taken regardless of the regime coming into force.

We ask respondents to be as specific as possible when providing costs and ensure monetary values are provided. However, we appreciate that information will be approximations and where this is the case, we welcome cost ranges.

**Instructions:** We are now going to ask you about any actions relating to protecting users from online harms that your organisation currently takes, including any associated costs.

**Question 17:** Does your organisation currently undertake any actions related to protecting users from online harms?

**Question 18:** What actions relating to protecting users from online harms does your organisation currently take<sup>191</sup>? *If helpful, please feel free to provide any explanatory information on specific actions in the boxes provided. If your organisation takes additional actions not explicitly noted in this list, please use the 'other' boxes provided at the bottom.*

**Question 19:** What currency are you providing cost information in?

**Question 20:** For the actions relating to protecting users from online harms you currently undertake, what is the cost?<sup>192</sup> *Please be as specific as possible and provide monetary values. Costings can be approximations/ranges. We would really like to understand the cost of specific actions relating to protecting users from harms. However, if you do not have information regarding the cost of specific actions, you can leave this question blank and provide an estimate of the total annual cost in the next question.*

**Question 21:** What is the total annual cost of actions relating to protecting users from online harms you currently undertake? *You can skip this question if you were able to provide cost information next to the specific actions in the previous question and the sum of those costs is a fair reflection of your total*

**Instructions:** We are now going to ask you about any actions relating to protecting users from online harms that your organisation may have taken since the publication of the Online Harms White Paper, including any associated costs.

**Question 22:** Has your organisation already taken any actions relating to protecting users from online harms since the publication of the Online Harms White Paper (April 2019) specifically in order to ensure

---

<sup>191</sup> For all questions related to actions, prompted answers included risk assessments, terms and conditions, acceptable use policies, human content moderation processes, automated content moderation systems, reporting mechanisms, flagging mechanisms, content rating capabilities, age verification, age assurance, parental controls, complaints handling processes and procedures, tools to improve media literacy, and transparency reporting.

<sup>192</sup> The survey provided space for respondents to insert annual costs next to each specific action.

compliance with the future online harms regime? *This should not include actions your organisation has taken to protect users regardless of the online harms regime coming into force.*

**Question 23:** What actions relating to protecting users from harms have you taken since the publication of the Online Harms White Paper (April 2019) specifically to ensure compliance with the future online harms regime? *If helpful, please feel free to provide any explanatory information on specific actions in the boxes provided. If your organisation has taken additional actions not explicitly noted in this list, please use the 'other' boxes provided at the bottom.*

**Question 24:** For the actions relating to protecting users from online harms you have undertaken since the publication of the Online Harms White Paper (April 2019), specifically in order to ensure compliance with the future online harms regime, what was the cost to your organisation? *Please be as specific as possible and provide monetary values. Costings can be approximations/ranges. We would really like to understand the cost of specific actions relating to protecting users from harms. However, if you do not have information regarding the cost of specific actions, you can leave this question blank and provide an estimate of the total cost in the next question.*

**Question 25:** What is the total cost of actions relating to protecting users from online harms you have taken since the publication of the Online Harms White Paper (April 2019), specifically in order to ensure compliance with the future online harms regime? *You can skip this question if you were able to provide cost information next to the specific actions in the previous question and the sum of those costs is a fair reflection of the total cost.*

**Instructions:** We are now going to ask you about any actions relating to protecting users from online harms that your organisation expects to take in the future, in order to ensure compliance with the online harms regime.

**Question 26:** Based on your understanding of the future online harms regime, as set out in the Full Government Response to the Online Harms White Paper, do you expect your organisation to take further actions relating to protecting users from online harms, specifically in order to ensure compliance with the online harms regime? *These actions should be additional to what your organisation currently does and where possible, should not include planned actions your organisation will take to protect users regardless of the online harms regime coming into force.*

**Question 27:** Based on your understanding of the future online harms regime, as set out in the Full Government Response to the Online Harms White Paper, what additional actions will you have to take, in the future, specifically in order to ensure compliance with the online harms regime? *These actions should be additional to what your organisation currently does and where possible, should not include planned actions your organisation will take to protect users regardless of the online harms regime coming into force.*

**Question 28:** For the actions relating to protecting users from online harms you expect to take in the future specifically in order to ensure compliance with the online harms regime, what do you expect the cost to your organisation to be? *These actions should be additional to what your organisation currently does and where possible, should not include planned actions your organisation will take to protect users regardless of the online harms regime coming into force. Where possible, please be as specific as possible; however, costings can be approximations/ranges. We would really like to understand the cost of specific actions relating to protecting users from harms. However, if you do not have information regarding*

*the cost of specific actions, you can leave this question blank and provide an estimate of the total expected cost in the next question.*

**Question 29:** What is the total cost of actions relating to protecting users from online harms you expect to take in the future, specifically in order to ensure compliance with the future online harms regime? *You can skip this question if you were able to provide cost information next to the specific actions in the previous question and the sum of those costs is a fair reflection of the total cost.*

**Question 30:** To comply with the future online harms regime, are there any additional costs to your organisation which you have not included in your previous answers? *If 'yes', please provide as much information as possible in outlining these additional costs, including costings and/or FTE days where appropriate.*

# Annex E: Rapid evidence assessment of NetzDG methodology

The Network Enforcement Act (NetzDG) is a German law aimed at combating hate speech online which came into effect on 1 January 2018. The law applies to platforms with more than 2 million users and targets illegal harms only. The principal focus of NetzDG is upon transparency with platforms and effective complaint management.

After two years after its implementation, a rapid evidence assessment (REA) has been carried out to provide an overview of the impact of the NetzDG in Germany, specifically in relation to compliance costs faced by businesses, the legislation's effect on market innovation and whether it has reduced the prevalence of harms. The main objective of this REA was to enhance the knowledge base in regards to existing online regulation, specifically the German NetzDG, to inform the implementation of the Online Safety Bill.

A rapid evidence assessment (REA) is a review of the evidence landscape which focuses on one topic and the synthesis of evidence gathered to answer specific questions. It is a streamlined version of the systematic review process<sup>193</sup>.

The following research questions were developed to identify the impact of the policy in three key areas:

- What is the cost to businesses of NetzDG?
- What (if any) impact has NetzDG had on market innovation?
- What impact has NetzDG had on the prevalence of online harms?

The inclusion criteria was developed using the PICO strategy - population, intervention, comparison, outcomes.

- P - Businesses operating in Germany
- I - Introduction of NetzDG
- C - pre-NetzDG
- O - Impact on cost, innovation or prevalence of harms

*In addition, evidence included in the REA had to be written or translated into English and published since 2017. As a result grey literature was excluded from this REA<sup>194</sup>.*

When assessing the NetzDG evidence landscape we took certain steps to ensure our search was as systematic as possible given time and resourcing constraints. This involved the following steps:

---

<sup>193</sup> A systematic review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question

<sup>194</sup> Grey literature refers to materials and research produced by organisations outside of the traditional commercial or academic publishing channels.



When reviewing the literature the quality of each source was considered using a RAG rating<sup>195</sup>. Any sources from reports rated as red would have been excluded from the overall findings; however, all sources gathered were rated green or amber.

### Limitations

The REA did have certain limitations predominantly stemming from a lack of access to academic databases and data-scanning software. Subsequently, the evidence used in this REA was taken from Google Scholar on the week commencing 21st December 2020 and the data scanning was done manually. Lastly, the inclusion criteria was restricted to sources in English, this excluded articles written in other languages. This REA was not piloted. This would have involved using the search strings to collect the first five sources for each research question, these sources would have then been reviewed. Depending upon the relevance of the initial findings the search strings, research questions and/or criteria would have been adapted to yield more relevant sources.

As a result, this REA risked excluding relevant information that was written in German or that which did not appear on Google Scholar in December 2020. Reports from the German Government, federal or otherwise, were not included in our research, however, to our knowledge such reports do not exist or are not published.

---

<sup>195</sup> Red (significant concerns around robustness of evidence), Amber (non-significant concerns regarding robustness of evidence), Green (no concerns regarding robustness of evidence).



## Annex F: Consultation questions

**Consultation question 1:** Do you agree with the estimates for costs incurred in the transition period, including estimates for familiarisation, changes to user reporting mechanisms, and revising terms of service? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 2:** How (if at all) will the inclusion of user-generated fraud affect transition costs? The government welcomes any evidence you can provide to understand the impacts.

**Consultation question 3:** Are you able to identify any other costs businesses would incur during the transition period? The government welcomes any evidence you can provide.

**Consultation question 4:** Do you agree with this assessment of the incremental cost of producing a risk assessment? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 5:** Would the cost of producing a risk assessment differ for Category 1 services (those expected to address legal but harmful content accessed by adults)? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 6:** Do you agree with this assessment of the proportion of platforms that will require additional content moderation to ensure compliance? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 7:** Do you agree with the estimates for the incremental cost of additional content moderation? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 8:** How would the cost of additional content moderation differ for platforms required to address legal but harmful content? The government welcomes any evidence you can provide.

**Consultation question 9:** Do you agree with the estimates for the cost of transparency reporting? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 10:** Do you agree with the estimates of potential compliance costs? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 11:** How (if at all) will the inclusion of user-generated fraud affect compliance costs? The government welcomes any evidence you can provide to understand the impacts.

**Consultation question 12:** Are there any additional costs associated with compliance not considered in this IA? The government welcomes any evidence you can provide.

**Consultation question 13:** Do you agree with the assessment that the costs to individuals (i.e. not businesses or civil society organisations) will be negligible? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 14:** Are there any additional indirect costs or benefits not discussed in this IA? The government welcomes any evidence you can provide.

**Consultation question 15:** Do you agree with the assumptions used in this assessment of the costs and benefits of the policy? The government welcomes any evidence you can provide to refine the assumptions.

**Consultation question 16:** Do you agree with the assessment of the impacts on small and micro businesses? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 17:** Is there anything additional the government can do to support small and micro businesses in the implementation of this regime?

**Consultation question 18:** Do you agree with the assessment of the impacts on trade and innovation? The government welcomes any evidence you can provide to refine the estimates.

**Consultation question 19:** Do you agree with the points made in the equalities IA? The government welcomes any evidence you can provide.

**Consultation question 20:** Do you agree with the assessment of the impacts on competition in the market? The government welcomes any evidence you can provide.