Title: The Online Safety Bill **IA No:** **RPC Reference No:** RPC-DCMS-4347(4) **Lead department or agency:** Department for Digital, Culture, Media and Sport **Other departments or agencies:** Home Office	Impact Assessment (IA) **Date:** 31/01/2022 **Stage:** Final stage **Source of intervention:** Domestic **Type of measure:** Primary legislation **Contact for enquiries:** soh-analysis-team@dcms.gov.uk

Summary: Intervention and Options

RPC Opinion: Fit for purpose

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**
NQ	NQ	NQ	Qualifying provision

What is the problem under consideration? Why is government action or intervention necessary?

The internet is a powerful force for good, but illegal and harmful content is widespread online. A lack of transparency, perverse incentives, and an inconsistent voluntary approach towards fighting harm online has limited the effectiveness of market solutions. Therefore, the Government must act to protect users online.

What are the policy objectives of the action or intervention and the intended effects?

The policy objectives are as follows:

- **to increase user safety online**
- **to preserve and enhance freedom of speech online**
- **to improve law enforcement's ability to tackle illegal content online**
- **to improve users' ability to keep themselves safe online**
- **to improve society's understanding of the harm landscape**

What policy options have been considered, including any alternatives to regulation? Please justify preferred option (further details in Evidence Base)

- **Option 0 - do nothing**: a continuation of platforms being liable for illegal content that they "host" only, with no existing legal framework to tackle content and activity which is harmful but not illegal.
- **Option 1 - online safety framework**: 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.

Option 1 is the Government's preferred option as it is likely to achieve reductions in online harm while maintaining a proportionate and risk-based approach.

Is this measure likely to impact international trade and investment?		Yes		
Are any of these organisations in scope?	**Micro** Yes	**Small** Yes	**Medium** Yes	**Large** Yes
What is the CO_2 equivalent change in greenhouse gas emissions? (Million tonnes CO_2 equivalent)		**Traded:** n/a	**Non-traded:** n/a	

Will the policy be reviewed? It will be reviewed. **If applicable, set review date:** 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 : Alison Kilburn Date: 27/01/2022

Summary: Analysis & Evidence **Policy Option 1**: online safety framework

Description: 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

FULL ECONOMIC ASSESSMENT

Price Base Year 2019	PV Base Year 2020	Time period 10	Net Benefit (Present Value (PV)) (£m)		
			Low: -£1,787m	**High:** -£3,291m	**Best Estimate:** -£2,507m

COSTS (£m)	Total Transition (Constant Price) Years		Average Annual (excl. Transition) (Constant Price)	Total Cost (Present Value)
Low	£50.7m		£206.5m	**£1,787m**
High	£95.4m		£379.1m	**£3,291m**
Best Estimate	£65.3m		£289.7m	**£2,507m**

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 (£9.6m-£17.5m), ensuring they have a user reporting mechanism in place (£17.7m-£33.8m), updating terms of service (£17.8m-£33.6m).

Businesses are expected to incur the following ongoing compliance costs: producing risk assessments (£17.5m-£48.7m), potential additional content moderation (£1,319.1m-£2,486.2m), employing age assurance systems (£17.9m-£89.6m), transparency reporting (£0.8m-£10.3m), conducting due diligence on advertisers (£63.1m-£221.4m), offering optional user verification (£8.7m-£13.5m), producing FoE and privacy IAs (£1.1m - £11.5m) and paying an industry fee (£313.9m)

Government is expected to incur the following costs (all in 10 year PV): justice impacts (£0.3m).

The creation of an additional offence related to cyberflashing is expected to be implemented through the Online Safety Bill, with the potential for additional costs to law enforcement and the criminal justice system. This is not accounted for within this impact assessment but an additional impact assessment will be produced by the Ministry of Justice, the analysis of which will be incorporated into an updated Online Safety Bill impact assessment.

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

The following costs to businesses have not been monetised: fines for non-compliance (out of scope), cost to internet service providers (ISPs) and payment service providers (PSPs) of business disruption measures, cost to industry and government stemming from the requirement to report online child sexual abuse (CSA). 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 potential pass through from the fraudulent advertising duty, innovation impacts, competition impacts, freedom of expression implications, privacy implications, and trade impacts - these have all been thoroughly assessed qualitatively.

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

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

Based on a subset of quantified online harm,[1] this IA estimates that this option would need to reduce online harm on an average annual basis by between 1.5-2.7% (central: 2.1%) in order to break even. This equates to between £199 million - £365 million (central: £278 million) average annual benefit over the appraisal period.[2]

Given the difficulties in monetising the impact of online harm, 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.

A number of illustrative scenarios are also estimated to indicate how the benefit-cost ratio (BCR) would change if different illustrative assumptions were made about the effectiveness of Option 1 in reducing harm. Under a scenario in which harm is reduced by 3%, the BCR for the Online Safety framework is estimated to be 1.46, and under a scenario in which harm is reduced by 5%, the BCR for the Online Safety framework is 2.43.

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

This proposal is expected to accrue the following non-monetised benefits:

- Improved efficacy of law enforcement and crime prevention for illegal content and behaviour online, expected to accrue as either cost savings or improved outcomes, for example, through minimising the creation and spread of illegal harm.
- Increases in levels of media literacy or the ability of users to keep themselves safe online.
- Expansion of the SafetyTech sector through increased growth (as measured by revenue, profit and GVA), businesses, and jobs.
- An increase in the evidence base underpinning online harm.
- A reduction in the non-monetised impacts of online harms (namely, those not captured in the break-even analysis).

Key assumptions/sensitivities/risks	**Discount rate (%)**	3.5%

The key assumptions for this option are:

- the number of platforms in scope of the framework
- the risk categorisation of in-scope platforms (used as a proxy for proportionate requirements stemming from future codes of practice)
- the incremental cost of potential changes to content moderation practises
- growth rate of online harm over the appraisal period

All key assumptions are tested.

BUSINESS ASSESSMENT (Option 1)

Direct impact on business (Equivalent Annual) £m:			**Score for Business Impact Target (qualifying provisions only) £m:** illustrative only
Costs: 250.6	**Benefits:** 0	**Net:** 250.6	

[1] This includes contact CSA, modern slavery, hate crime, illegal sale of drugs online, cyberstalking, fraud facilitated by user generated content, cyberbullying, and intimidation of public figures.

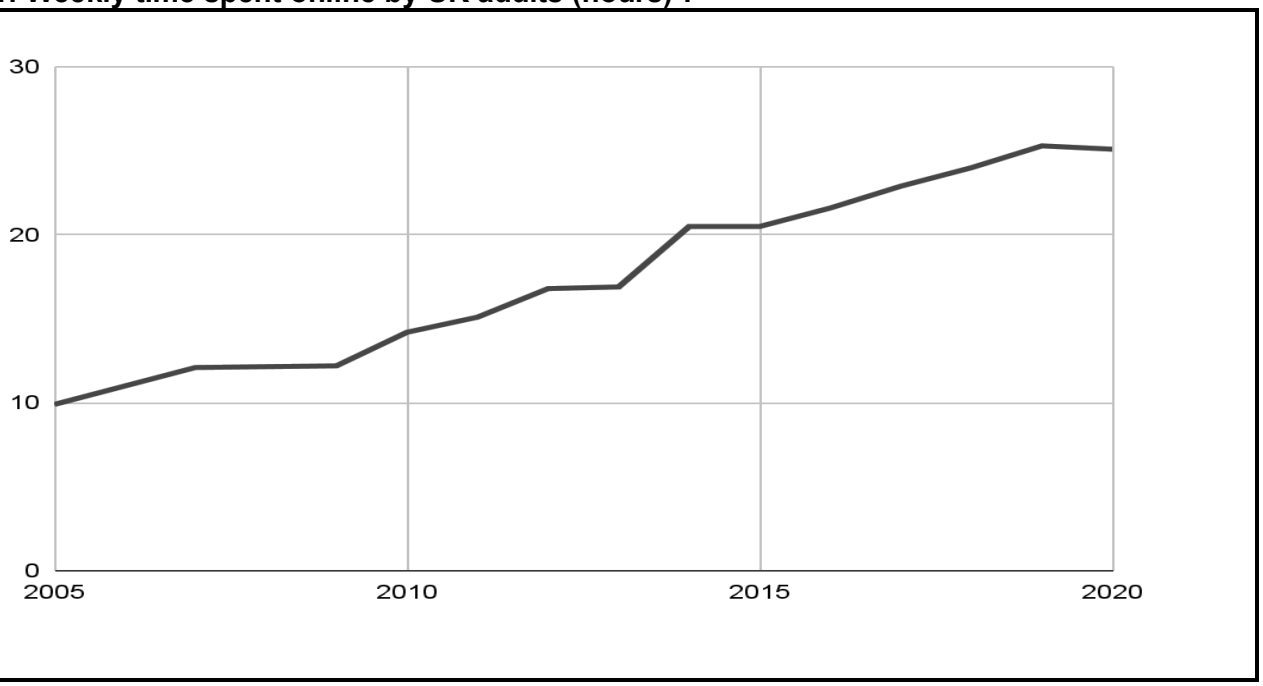
[2] Averages are calculated from the second year onward, the year in which compliance costs are assumed to be incurred and benefits are expected to accrue.

Policy rationale

Problem under consideration

Individuals in the UK are spending an increasing proportion of their time online. This has been part of a long-term trend over the last fifteen years. Between 2005 and 2020, the weekly time spent online by UK adults has significantly increased from 9.9 hours to 25.1 hours (see Figure 1). Being online is an integral part of everyday life for the majority of people in the UK. For both children aged 5-15 years and adults aged 18-54 years, going online is almost universal.[3] During the pandemic, internet use among adult internet users in the UK was most pronounced in April 2020, averaging 4 hours and 2 minutes online each day. Similarly, three-quarters of British parents reported that their child's screen time averaged nine hours per day at the height of the first lockdown – nearly double the screen time prior to the outbreak.[4]

Figure 1: Weekly time spent online by UK adults (hours)[5].



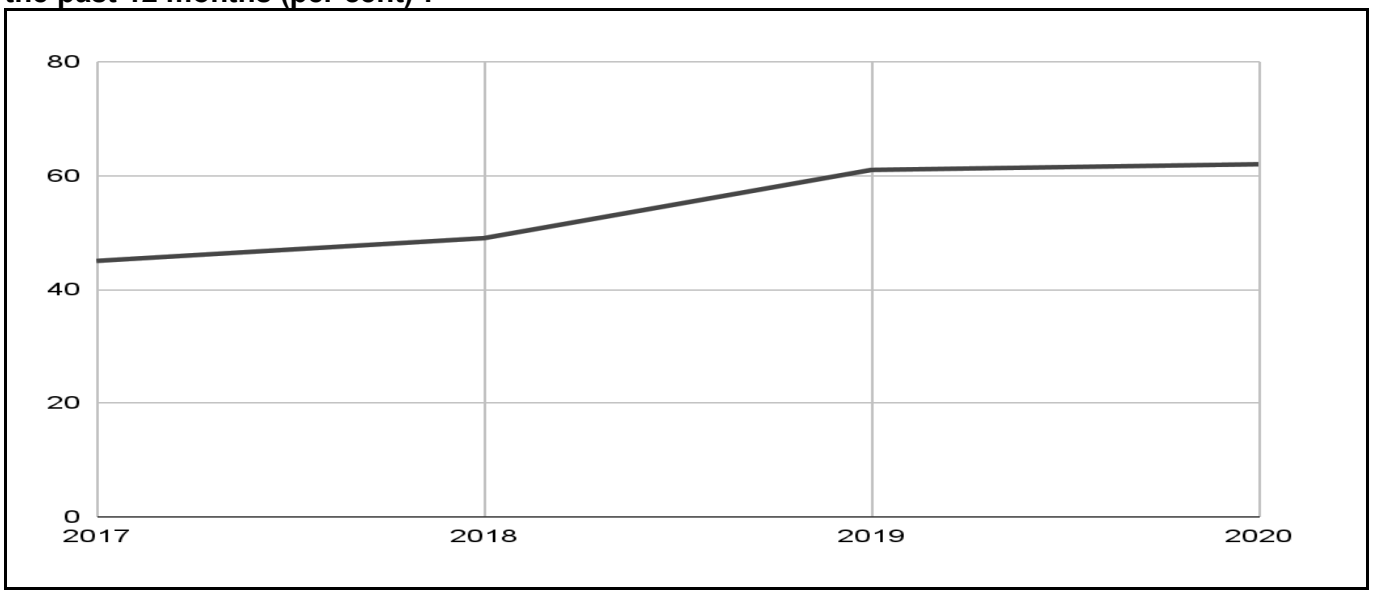
This line graph illustrates the increasing amount of time spent online each week by UK adults, from 9.9 hours in 2005 to 25.1 hours in 2020.

1. **The internet is a vital part of so many everyday activities.** Over the years, reliance upon the internet for communication, access to information, entertainment, and e-commerce has dramatically increased in the UK. The internet is a place for socialising with 92% of UK internet users going online to communicate with others and 82% of them having a social media profile. The internet also acts as a source of entertainment with 74% of internet users watching TV content online.[6] Findings presented in the Reuters Digital News Report 2020 suggest that in the week prior, 79% of respondents (all of whom are news users) sourced news online compared to 71% accessing news through TV.[7] However, while the internet is a powerful force for good, illegal and harmful content and activity is widespread online.

2. **UK users are becoming increasingly concerned about the content they interact with and their experiences online.** 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.[8]

[3] Online Nation - 2021 report (Ofcom). The percentage of individuals who go online in each age group: 18-24 (97%), 25-34 (98%), 35-44 (99%) and 45-54 (92%).
[4] Study suggests lockdown could have permanently altered families' tech habits (October 2020)
[5] Adults' Media use and Attitudes report' (2005-2019) - Ofcom
[6] Adults' Media Use and Attitudes report - Ofcom (2020/21)
[7] Reuters Institute Digital News Report 2020 - Reuters
[8] Internet users' experience of online harms - Ofcom and ICO (2020)

Figure 2: Adult internet users that have had at least one potentially harmful experience online in the past 12 months (per cent)[9].



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.

3. **Parents are also increasingly concerned about nearly all aspects of their child's online use.[10]** Only 51% of parents of 12-15 year olds think that their child has a good balance between screen time and other things.[11] 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 agreed with this in 2019, compared to two-thirds in 2015.[12] [13] Whilst most parents are aware of the available parental safety controls, there is limited use of such features. 66% of parents are aware of content filters via parental control softwares yet only 29% of parents actually use them. In addition, for most social media platforms, the minimum age requirement is 13; however, Ofcom's research showed that 42% of children under the minimum age requirement (that is, aged between 5 and 12 years old) used social media.[14]

4. **Research commissioned by 5Rights[15] claims that the current design of social media platforms does not always prioritise the safety of its users, particularly that of children.** The findings indicate that the designers' objectives focus on increasing time spent, users, and activity on the platform. This is in part done through the use of algorithms which amplify the type of content that a profile appears to show interest in, potentially resulting in the promotion of harmful content. The report used child-aged avatars to assess which content the algorithms amplified which worryingly included sexualised images, content promoting eating disorders or weight loss and self-harm, despite the platforms recognising that these accounts were registered as children.[16]

5. **The scale of illegal child sexual abuse (CSA) content online is significant.** In 2020, there were 21.7 million reports of CSA content[17] referred to the National Center for Missing and Exploited Children (NCMEC),[18] an increase of 28% from 2019[19]. Reports of CSA content online also appear to have increased through the pandemic, with the Internet Watch Foundation (IWF) recording a record number.[20]

---

[9] Internet users' concerns about and experience of potential online harms - Ofcom (2017-2020)
[10] Children and parents: media use and attitudes report 2020 - Ofcom
[11] Children and parents: media use and attitudes report 2020 - Ofcom
[12] Children and parents: media use and attitudes report 2019 - Ofcom
[13] Children and parents: media use and attitudes report 2015 - Ofcom
[14] Children and parents: media use and attitudes report 2020/21 (Ofcom, 2021)
[15] 5Rights is a charity focusing on the protection of children's privacy and data online.
[16] Pathways: How digital design puts children at risk - 5Rights (2021)
[17] Reports can contain multiple pieces of content including images, videos or other files. In 2019, 16.9 million reports totaled 69.1 million pieces of suspected CSA material and other incident related content. In 2020, 21.7 million reports included 64.5 million pieces of suspected CSA material and other incident related content.
[18] By the numbers - NCMEC (2020)
[19] By the numbers - NCMEC (2019-2020)
[20] IWF has record month as public reports of child sexual abuse surge - IWF (2020)

6. **Online platforms are also used as a tool to promote extremist content.** 3% of UK adults and 5% of children aged 12-15 have encountered material online promoting terrorism/radicalisation.[21] Evidence from a Ministry of Justice sample of extremist prisoners found that for cases prior to 2005, 83% were radicalised face to face with only 17% radicalised using a mixture of online and face to face. For cases from 2015 to 2017 this had increased dramatically to 56% radicalised using a mixture of online and offline, 27% purely online and 17% just offline, showing a reversal of the pathways to extremism.[22]

7. **Online fraud facilitated by UGC (user-generated content) continues to pose a major threat to UK users with large sums being lost to criminals each year.** Fraud is the UK's most common crime type: in the year ending March 2020 there were 3.7 million instances of fraud in England and Wales[23], and over half of these had some online element.[24] [25] Online fraud not only has a significant financial impact on the victim but can also take an emotional toll, this is particularly relevant for romance scams online. In the year ending February 2020, the National Crime Agency (NCA) reported victim losses of over £60 million from romance fraud alone.

8. **As spend on digital advertising increases and consumers shift their purchasing online, victims of scam adverts are incurring significant financial losses.** Tackling fraudulent advertising is vital as more people see scam adverts than fraudulent UGC, with 63% having seen a scam advert and almost half seeing them at least monthly. One in four (23%) people who have experienced a mental health problem have been victim to an online scam, three times the rate among people who have never experienced a mental health problem (8%).[26]

9. **Content and activity that is harmful but not illegal is also widespread online.** One study of children between 8-18 years old presenting to hospital following self-harm found that 26% of them had viewed self-harm and suicide content online.[27] Online advocacy of self-harm poses a clear threat to people's wellbeing - according to an online study of European children, 10% of children aged 11-16 years had visited pro-eating disorder sites and 5% had visited suicide sites.[28]

10. **A significant proportion of children access pornography online both inadvertently and intentionally.** Although legal age restrictions on pornography exist, under the status quo, children can easily access pornography across a range of online platforms. 51% of children aged 11-13 years old have seen pornography and this number is likely conservative.[29] Many children - some as young as 7 years old - stumble upon pornography online, with 61% of 11-13 year olds describing their viewing as mostly unintentional.[30]. There are clearly not enough safeguards to protect children from accessing pornographic material online.

11. **Online abuse is widespread online.** The nature of the internet provides a channel through which abuse and hate speech can spread anonymously and instantaneously. Research conducted by Ofcom highlights that 9% of UK adult internet users had experienced hate speech or speech encouraging violence in the four weeks prior to completing the survey.[31] The prevalence of hate speech online is particularly concerning among individuals with protected characteristics. In the first six months of 2018, 22% of reported antisemitic incidents in the UK involved social media.[32] Similary in 2017, one third of reported anti-Muslim or Islamophobic incidents occured online.[33] Experiences of online abuse also vary depending on gender identity and sexual orientation. Findings of the Online Hate Crime Report 2020

---

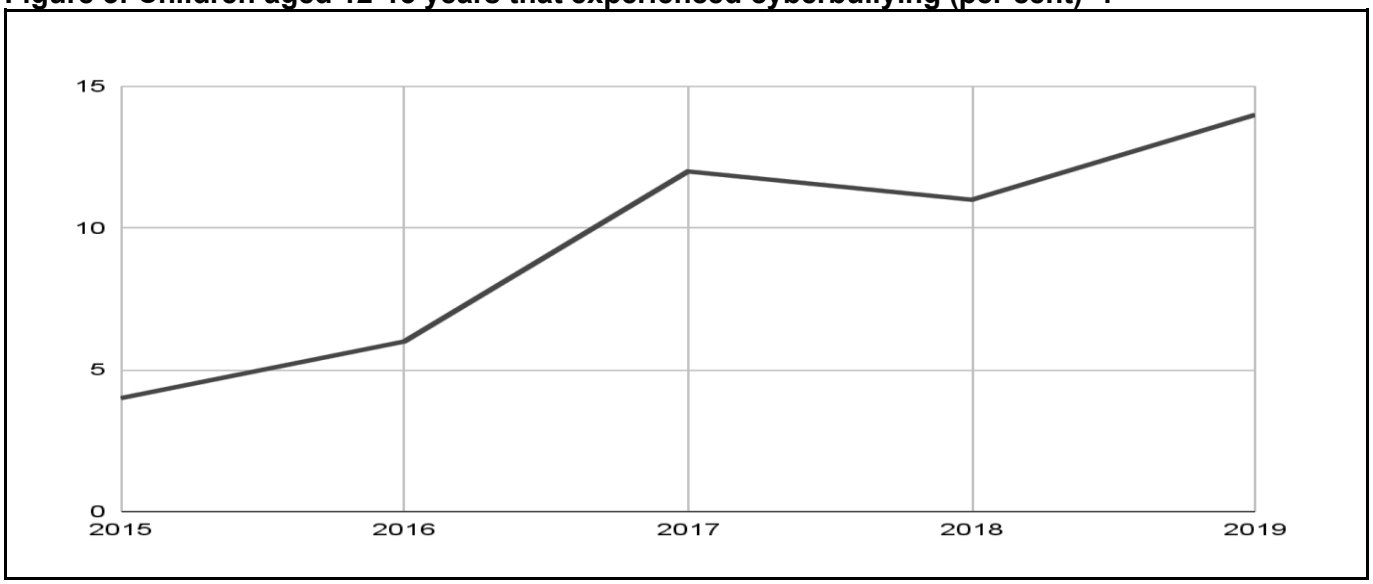
[21] Internet users' concerns about and experience of potential online harms - Ofcom (2020)
[22] Exploring the role of the Internet in radicalisation and offending of convicted extremists (MoJ, 2021)
[23] CSEW - ONS (year ending March 2020)
[24] Nature of crime: fraud and computer misuse - ONS (year ending March 2020)
[25] Cyber fraud represents cases where the internet or any type of online activity was related to any aspect of the offence.
[26] Caught in the Web - Online Scams and Mental Health (Money and Mental Health Policy Institute, 2020)
[27] Suicide and Self-Harm Related internet Use - Padmanathan et al. (2018)
[28] Risks and safety on the internet: the perspective of European children: full findings and policy implications from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries - LSE (2011)
[29] Young People and Pornography, BBFC, Revealing Reality, 2020
[30] Ibid.
[31] Online Nation 2021 Report - Ofcom
[32] Adult online hate, harassment and abuse: a rapid evidence assessment - LSE (2019)
[33] Adult online hate, harassment and abuse: a rapid evidence assessment - LSE (2019)

indicate that 80% of LGBT+ respondents[34] had experienced online abuse, resulting in 40% of these individuals reducing the use of their online accounts.[35]

12. **Recent events have shone a spotlight on the targeting of certain public figures.** Throughout the 2020 European Football Championship, England's men's football team were subjected to online abuse. Across England's three group games, over 2,000 abusive tweets were directed towards or naming the players and Gareth Southgate. This included 44 explicitly racist tweets and 58 that attacked players for their anti-racist actions.[36] In addition, there were increasing levels of abuse towards UK Members of Parliament during the pandemic with those from minority ethnic backgrounds and women frequently targeted during this period.[37]

13. **Children are experiencing cyberbullying at concerning rates.** Based on figures from the Office for National Statistics (ONS), one in five children aged 10-15 years in England and Wales experienced at least one type of online bullying behaviour in the year ending March 2020. The prevalence of online bullying is significantly higher for children with a long-term illness or disability (26%) than those without (18%). 22% of children said that these incidents affected them a lot emotionally and one in four young people now have anticipatory anxiety about being abused online.[38]

Figure 3: Children aged 12-15 years that experienced cyberbullying (per cent)[39].



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

Rationale for intervention

Negative externalities

14. **Online harm encompasses a number of negative externalities, harmful content has consequences beyond that of the direct impact upon the victim.** Internet users are less likely to validate online information sources,[40] and with 65% of adults in the UK using the internet as their main source of news,[41] this increases the likelihood of individuals falling victim to misleading and manipulative content. From analysing self-reported mask wearing and reasons given for not wearing a face mask, one randomised control trial found that after exposure to misinformation about mask wearing during the pandemic, 4.6% of people changed their behaviour. One study estimated that this change in behaviour

[34] Lesbian, gay, bisexual and transgender and related communities.
[35] Online Hate Crime Report 2020 - Galop
[36] Revealed: shocking scale of Twitter abuse targeting England at Euro 2020 (Guardian, 2021)
[37] MP Twitter Engagement and Abuse Post-first COVID-19 Lockdown in the UK: White Paper - Farrell et al (2021)
[38] Online bullying in England and Wales - ONS (year ending March 2020)
[39] Children and Parents: media use and attitudes report - Ofcom (2015-2019)
[40] Adults' Media Use and Attitudes report - Ofcom (2020/21)
[41] News Consumption in the UK: 2020 - Ofcom (2020)

on an individual level could have resulted in 2,187 additional hospitalisations and 509 additional deaths.[42]

15. **Similar negative externalities occur with exposure to content and activity that is harmful but not illegal.** 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.[43] In some cases, children can develop long-term behavioural difficulties including alcohol consumption and substance abuse.[44]

16. **Pornography can result in significant short- and long-term impacts on children.** Children - many who have stumbled upon pornography - feel a range of negative emotions after seeing this content, such as feeling shocked, confused, disgusted, sick, scared, and upset.[45] Exposure to this content at such a young age can negatively impact how they view their own body, for example by comparing themselves to the people featured in pornography, or seeing the people in pornography as examples of what a normal naked body looks like.[46] Several longitudinal studies have found an association between adolescents' pornography consumption and subsequent body dissatisfaction (as well as increased sexual and relational dissatisfaction).[47]

17. **Worryingly, children's exposure to pornographic content has also been shown to affect attitudes and sexual behaviour.** Evidence suggests that pornography can influence young people's sexual behaviours and expectations towards more "rough" and "forceful" sexual encounters.[48] In one longitudinal study, 10-15 year olds who consumed violent pornography were six times more likely to be sexually aggressive than those who did not consume it, or than those who consumed less aggressive pornography.[49] Also, in another study, 29% of children who intentionally access pornography did not think consent was needed if "you knew the person really fancies you", in comparison to only 5% of those who had mostly seen pornography by accident.[50]

18. **Online abuse can also influence people's willingness to speak out, fundamentally impacting on a functioning democracy**.[51] An international survey of female journalists found that 64% had experienced online abuse – death or rape threats, sexist comments, cyberstalking, account impersonation, and obscene messages.[52] 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.

Information Asymmetry

19. **In addition to negative externalities, information asymmetries exist between users and online platforms.** While efforts have been made to increase the transparency between the two, there remains a lack of clarity among users with regards to the risk of exposure to harmful online content. This extends to information provided on platforms' websites. One survey indicates that many people do not engage with this information and those that do engage struggle to understand the information provided.[53] As a result, 97% of 18-34 year olds agree to a platform's terms of service without even reading them.[54]

---

[42] The Cost of Lies - London Economics (2020)
[43] The Annual Bullying Survey 2017 - Ditch the Label
[44] The Relative Importance of Online Victimization in Understanding Depression, Delinquency, and Substance Use - Mitchell et al. (2007)
[45] Researching the Affects That Online Pornography Has on U.K. Adolescents Aged 11 to 16 (Martellozzo et al. 2020)
[46] Young People and Pornography, BBFC, Revealing Reality, 2020
[47] What is the IMPACT of pornography on young people? A RESEARCH BRIEFING for educators, PSHE, 2020
[48] Young People and Pornography, BBFC, Revealing Reality, 2020
[49] X-rated material and perpetration of sexually aggressive behavior among children and adolescents: is there a link? (Ybarra et al., 2011)
[50] Young People and Pornography, BBFC, Revealing Reality, 2020
[51] Online abuse against women MPs chilling - Amnesty International, 2020
[52] IFJ global survey shows massive impact of online abuse on women journalists - IFJ (2018)
[53] Understanding how platforms with video sharing capabilities protect users from harmful content online - Ernst & Young (2021)
[54] You're not alone, no one reads terms of service agreements - Business Insider, Survey conducted by Deloitte (2017)

20. **Children's attitudes towards the privacy of their online profiles highlights their lack of understanding of potential exposure to online harm**. 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%.[55] It is therefore difficult for users to make an informed decision as to how they use online platforms and what content they access.

Government Intervention

21. **Online platforms have failed to effectively address online harm and ensure the safety of their users.** In the absence of regulations, harmful online content is addressed on a voluntary basis. Measures taken to improve the safety of online platforms can be delayed and reactive, resulting from governmental or societal pressure. Social media platforms' actions following widespread concern over the prevalence of COVID-19 misinformation illustrates the reactive nature of market solutions.

22. **The need for regulation is recognised by the sector 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 harm.[56] 61% of UK adult internet users believe that individuals must be protected from seeing inappropriate or offensive content online.[57] There is international recognition for the need to regulate online platforms (see international context section below). One study has found that 60% of US consumers support more government regulation of platforms.[58] The findings also suggest that there are significant concerns about the practises of online platforms and the power that the larger platforms hold.[59]

23. **Absent regulation, there is the potential for 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 re-tweeted than real news.[60] The high levels of interaction with fake news, including the anti-vaxx rhetoric, generates a higher profit for platforms. Between July and August 2020, interactions on posts criticising COVID-19 vaccines on six UK Facebook pages increased by 350%.[61] Removing this content could therefore result in a short-term loss of profit and reduced user engagement.

24. **Some platforms will face incentives to address harmful content in order to maintain advertising revenue; however, this incentive does not appear to be driving sufficient change.** In 2019, it was projected that by 2020, UK advertisers would be spending almost two-thirds of their budget online[62] and it is unlikely that this would be significantly affected by a platform's moderation activities. This is because advertisers have an inelastic demand for social media advertisements on the largest platforms. This is a result of smaller platforms being unable to offer advertisers such a large and engaged user base that is provided by the more popular social media platforms. Advertisers rely on the popularity of online platforms with young consumers; in 2020 a third of 12-15 year olds said they used social media and messaging services to follow companies and brands that they like.[63] The main social media platforms also provide a unique method of marketing, namely user-generated content (UGC)[64] through the use of influencers. UGC has been shown to have a significantly stronger impact than marketing generated content on consumer behaviour[65] 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.

---

[55] Report on Internet Safety Measures - OFCOM (2015)
[56] People, Power and Technology: The Tech Workers' View - DotEveryone (May 2019)
[57] Online Nation 2021 report - Ofcom 2021
[58] Platform Perceptions, Consumer Attitudes on Competition and Fairness in Online Platforms - CR Consumer Reports, Digital Lab (2020)
[59] Platform Perceptions, Consumer Attitudes on Competition and Fairness in Online Platforms - CR Consumer Reports, Digital Lab (2020)
[60] The spread of true and false news online - Vosoughi et al. (2018)
[61] Online Nation 2021 report - Ofcom (2021)
[62] Almost two-thirds of UK ad spend to be online by 2020 - Hammett (2019)
[63] Online Nation 2021 report - Ofcom (2021)
[64] Content, such as images, videos, text, and audio, that has been posted by users on online platforms.
[65] Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content - Goh et al. (2013)

25. **The legal incentive for firms (through potential legal liability) to address both illegal and legal but harmful harm is lacking.** 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 challenges of bringing such claims and the inevitable costs involved in legal action. On top of this, the existing legal framework for online harm 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.

26. **A clear, proportionate and predictable regulatory framework will encourage businesses to start up, grow and invest.** Many other countries are also planning to introduce online regulation. By acting first we will be able to provide certainty to platforms. There is an opportunity to set global standards, unlock investment and influence the global approach.

Wider international and regulatory context

Domestic context

27. **e-Commerce Directive**: The 2000 e-Commerce Directive (Directive 2000/31/EC) 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. The majority of eCommerce Directive was implemented into UK law via the Electronic Commerce Regulations 2002. The majority of these UK regulations have not changed at the end of the transition period. However, some provisions ceased to apply from 1 January 2021.

28. **Audiovisual Media Services Directive (AVMSD)**: In 2010, AVMSD expanded in scope to include Video on Demand services (such as Netflix). UK video on demand services (such as TikTok, OnlyFans or Vimeo) are required to take proportionate measures to ensure children are not normally able to access pornographic content. AVMSD 2020 (Directive (EU) 2018/1808) introduced rules for video sharing platforms (VSPs) for the first time. Last year, the government announced Ofcom as the national regulator for UK-established VSPs. The UK transposed the revised Directive through the AVMSD 2020 regulations which came into force on the 1st of November 2020. The revised AVMSD 2020 regulations place requirements on UK-established VSPs to protect all users from illegal content through taking appropriate measures. UK-established VSPs are also required to take measures to protect minors from harmful content. The regulations share broadly similar objectives to the OSB and will be superseded once the latter comes into force.

29. **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 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

30. 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. 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 harm.

31. **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.

32. **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.

33. **Australia**: The Australian Online Safety Bill 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 by requesting internet service providers (ISPs) block access to sites hosting seriously harmful content.

34. **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 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, would be responsible for conserving content for law enforcement, and for assessing risks around both tackling illegal content and breaching freedom of speech.

35. **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. 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.

36. **Multilateral collaboration:** Under the UK's Presidency of the G7, in April 2021, the G7 agreed on a set of Internet Safety Principles. This is significant as it is the first time that an approach to internet safety has been agreed in the G7. The Principles are broad in scope, allowing for both regulatory and non-regulatory approaches to increasing internet safety. The agreed text includes four underpinning principles that will inform approaches as well as four operational principles on safety technology, media literacy, child protection and youth participation where consensus for concrete action has existed. In addition to this, in March 2020, in collaboration with the Five Country Ministerial representatives of the US, Canada, Australia and New Zealand, the UK formally launched the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. These range from pledges to stop existing and new child sexual abuse material appearing on platforms, taking steps to stop the livestreaming of abuse, identify and stop grooming and predatory behaviour.

Policy objectives

37. The policy objectives are:

    - **to increase user safety online**: this will be achieved through reduced risk and incidence of specific online harms, especially with respect to children.

- **to preserve and enhance freedom of speech online:** this will be achieved through both reducing online harm which can lead to user disengagement and ensuring that the proposals do not result in 'over-blocking' and unjustified content removal - the OSB has strong safeguards for freedom of expression online.
- **to improve law enforcement's ability to tackle illegal content online**: this will be achieved through both a general reduction in illegal harm online and by making it easier for law enforcement to tackle identified illegal harm through increased transparency and reporting, improvements in safety technology, and effective regulatory oversight.
- **to improve users' ability to keep themselves safe online:** this will be achieved through greater platform transparency and a combination of non-regulatory support measures which focus on empowering users, such as media literacy initiatives.
- **to improve society's understanding of the online harm landscape**: this will be achieved through enhancing the amount and quality of information in relation to online harm that is available to government, industry and civil society.

38. These objectives will form the basis of monitoring and future evaluation.

Options considered

Summary of options

39. This impact assessment (IA) considers only one regulatory option in addition to a *do nothing* baseline; however, a range of options were considered as part of the consultation[66] and earlier policy development.

    - **Option 0 - do nothing**: The *do nothing* option would entail a continuation of platforms being liable for illegal content that they "host" only with no existing legal framework to tackle harm being caused to children or adults through online content and activity which is harmful but not illegal.

    - **Option 1 - online safety framework**: This option introduces 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. Duties are set out in primary legislation with details of how in-scope companies can fulfil their duty of care set in codes of practice.

40. **Option 1 is the government's preferred option** as it is likely to achieve reductions in online harm while maintaining a proportionate and risk-based approach. It will also aim to deliver a vibrant and competitive digital economy with high levels of user trust and confidence. Doing nothing has not provided sufficient incentive for platforms to reduce online harm.

Justification for the preferred option

Consideration of additional options

41. As part of the early policy development process, the Government considered a range of options to address the problem of online harm. In addition to considering a wide range of options throughout policy development, the Government assessed three distinct policy options against a *do nothing* counterfactual in its published consultation stage IA, these were:

    - **limited risk based scope**: A duty of care for UGC and activity addressing illegal harm, and safeguarding children from both illegal and harmful content and activity. Duties are set out in primary legislation and guidelines or codes of practice.

[66] Consultation outcome - Online Harms White Paper (DCMS & Home Office, 2019)

- **full risk based scope**: A duty of care for UGC 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.
- **uniformly applied safety duties**: Detailed safety duties setting out organisations' responsibilities in addressing illegal harm 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 categories of harm and organisations in scope.

42. The consultation stage IA provided an indication of the likely scale of impacts stemming from the three options and estimated an illustrative break even point for each:

Table 1: Estimates presented in the consultation stage IA

	Limited risk based scope	Full risk based scope (preferred)	Uniformly applied safety duties
Net present value	-£1,689m	-£2,118m	-£7,355m
Equivalent annual net direct cost to business (EANDCB)	£156m	£206m	£814m
Percentage reduction in [quantified] harms required to break even	3.1%	3.9%	13.5%

43. **Limited risk based scope** was discounted as it did not address the significant problem of legal but harmful content accessed by adults. While the break-even point for a **limited risk-based scope** was lower than the full risk-based scope, this was mainly due to the difficulty in quantifying legal harm in the baseline. There are many harms that fit into the legal but harmful category, and for which preliminary evidence suggests are significantly prevalent in the UK to have potentially large costs, but are yet to be quantified.[67] This is largely due to many of these harms emerging relatively recently, and so data and evidence on the impact of these harms remains sparse. These include many forms of online abuse, disinformation, content related to self-harm and suicide, childrens' access to pornographic content, and advocation of risky and dangerous behaviour.

44. The option assessing **uniformly applied safety duties** was discounted as it was estimated to result in significant impacts on business while not reflecting the variance of harm on different platforms and failing the proportionality test. **A full risk-based scope** was the Government's preferred option in the consultation stage IA and remains the preferred option at final stage (represented here as Option 1). A full description of the three options previously assessed can be found at page 21 of the consultation stage IA.[68]

Non-regulatory approaches

45. This IA does not specifically consider a non-regulatory option as an alternative to legislation. Self-regulation and voluntary approaches to tackle harm were considered as part of the long list policy development process, but given the wide-ranging and significant societal impacts of online harm, 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.

46. Alongside online safety legislation, the Government is pushing forward with a number of innovative non-regulatory online safety implementation measures. These measures are complementary to the legislation, vital to the success of the framework, and aim to create the optimal conditions for the legislation to be effective. Proposed measures include:

[67] Indicative figures for prevalence of particular types of harm, including some of the legal but harmful harms listed can be found in Ofcom's

[68] The Online Safety Bill - impact assessment (HMG, 2020)

- **Media literacy initiatives**: this includes taking steps to educate and empower users with the skills they need to stay safe online, by supporting organisations that undertake media literacy activity to do so in a more effective and wide-reaching way. Our recently published Online Media Literacy Strategy highlighted 6 key media literacy challenges we will be taking action to address including promoting better evaluation of the outcomes of media literacy initiatives, and building audience resilience to mis- and disinformation. Our accompanying first annual Online Media Literacy Action Plan set out eight government-led initiatives to support these objectives, including piloting work with civil society groups and training initiatives for teachers, support workers, librarians and community groups; establishing a stakeholder media literacy Taskforce; and improving the inclusivity of media literacy awareness through a communications campaign.
- **Child and adult online safety initiatives**: this will include national research projects to strengthen the evidence base on the prevalence and impact of online harms such as cyberbullying, anonymous abuse and racism impacting children, adults and vulnerable users groups.
- **Investment in the online safety technology sector**: initiatives to grow and support innovation in the online safety tech industry, to help accelerate development of the technical capabilities that will support online harm regulation as well as growing and raising the profile of the sector more generally, thereby creating more jobs. These include sector analysis and definition reports, export support, a challenge fund to generate innovative technological ways of detecting harm, the Online Safety Data Initiative (OSDI) to open up access to sensitive online harms data, and funding of an innovation network in partnership with InnovateUK. DCMS also supported the launch of the Online Safety Tech Industry Association (OSTIA) in March 2020 and continues to work closely with them to address challenges within the sector.
- **Age Assurance initiatives:** a programme of work to improve trust and innovation in the sector, and help drive the uptake of solutions in advance of the Online Safety Bill. These include a programme of research on the viability of different methods of age assurance and supporting the development of standards that will bring consistency to criteria including security, user privacy, and inclusion. This includes working with the British Standards Institute and the International Organization for Standardization to develop an international standard. DCMS also intends to work with industry to support greater transparency on the effectiveness of solutions and establish a consistent and trusted way for companies to provide evidence on the accuracy of solutions.
- **Safety by design initiatives**: a programme of research to improve understanding on the role of specific platform design choices on key types of harm, and work to produce and promote guidance on a safety by design approach and industry best practice. It also includes research to better understand the knowledge that key disciplines, including web developers and product designers, have on user safety to support targeted interventions to improve awareness and action on safety.

47. While this IA does not consider a separate non-regulatory option, the Government is committed to supporting businesses and users outside of planned legislation.

Development of the preferred option

48. The Government has engaged extensively with platforms, users, Parliament and civil society throughout the development of the online harms policy. The Online Safety Bill (OSB) as introduced is the result of an iterative and collaborative process of policy development. Starting in October 2017, the Department for Digital, Culture, Media & Sport (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 harm and the Government's role in supporting users.

49. The Online Harms White Paper (OHWP) was then 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. It described a new regulatory framework establishing a duty of care on platforms 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 UGC 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 harm simply moves and proliferates outside of the ambit of the new regulatory framework.

50. The open public consultation process received over 2,400 responses ranging from companies in the technology industry including large tech corporations and small and medium-sized enterprises, academics, think tanks, children's charities, rights groups, publishers, governmental organisations and individuals. In response to the consultation, the Government gave an indication of its direction of travel in a number of key areas in the OHWP - Initial Government Response[69], 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.

51. Following this, further work was undertaken to develop and refine policy with a number of important changes made. The full intended policy position was set out in the full government response[70] published in December 2020 along with confirmation that Ofcom would be named as the regulator. In May 2021, the draft OSB was presented to Parliament alongside a consultation stage IA and Parliament established a joint committee to conduct pre-legislative scrutiny.

52. Since the OHWP, and as a result of extensive engagement, there have been numerous changes to the policy. Some of the key changes include:

    - The assurance of robust safeguards for journalistic content to protect freedom of expression
    - Specific exemptions for low-risk services including reviews and comments on directly published content to minimise the overall burden on platforms, especially small low risk platforms
    - Certain categories of harmful content (for example, advertising) to be excluded from regulatory scope in order to prevent regulatory duplication
    - A refined definition of duty of care covering harm to individuals but not to society more broadly to provide more clarity for platforms
    - 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 harm; only the highest risk and highest reach organisations providing Category 1 services will have to take action in respect of adult users accessing legal but harmful content on their services to minimise burdens on business
    - The removal of the specific exemption for online fraud offences to protect users from a harm with significant realised impacts
    - Provisions to protect content of democrat importance to protect freedom of expression
    - A provision on pornographic provider content to ensure that the OSB prevents children from accessing non-user generated pornographic content
    - A duty on Category 1 and 2A platforms to implement systems and processes to minimise the publication and/or hosting of fraudulent advertisements
    - A duty on Category 1 platforms to offer optional user verification and provide user empowerment tools

Option 0 – do-nothing

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

54. **Alongside reporting to the platform, there are a number of other routes for individuals to report content they believe to be illegal.** For example, the IWF provides a mechanism for individuals to anonymously report online CSA 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.

[69] Initial Government Response to the Online Harms White Paper (HMG, 2020)
[70] Full Government Response to the Online Harms White Paper (HMG, 2020)

55. **The current systems, especially relating to legal but harmful content and activity rely on voluntary action by platforms.** 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.

56. **In principle, an individual can bring a claim for breach of contract (either in the local small claims court or the High Court) if they consider that a 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 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.

57. **An individual - who would not need to 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.

58. **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.

59. **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.

60. **Alternatively individuals have the opportunity to report harmful content and activity to the platform.** But it is entirely up to the platform as to how it will respond, and how effective that will be, as a means of redress.

61. **The legal incentive for firms (through potential legal liability) to address both illegal and legal but harmful harm 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 harm solely addresses illegal harm and not harm that is legal but harmful. It is consequently up to the individual platforms to voluntarily address legal but harmful content.

62. **Under the do-nothing option, platforms face perverse and competing incentives in relation to content moderation.** Harm such as misinformation have wide-ranging negative impacts on both the individual and 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,[71] there is the potential for perverse incentives to delay or incentivise against the removal of harmful content.

63. **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.** By 2024, internet advertising is expected to account for 70% of total UK advertising spend,[72] and figures for 2021 suggest that this spending is focussed heavily on a small number of large companies (with Google and Facebook alone accounting for 68.5% of total digital advertising spend).[73] It is unlikely that the volume and concentration of spend is significantly sensitive to 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.

[71] The spread of true and false news online (Vosoughi et al., 2018)
[72] Advertising spending in the United Kingdom 2021-2024 (Statsita, 2021)
[73] UK Digital Ad Spending 2021 (eMarketer, 2021)

64. **Public pressure can act as a driver of content moderation processes but this could ultimately lead to a delayed and reactive approach to addressing harm.** A study of VSPs highlighted that public pressure (as it relates to brand integrity) is a driver of investment in user safety measures.[74] While it is right for platforms to react to user sentiment, this leaves open the possibility that approaches are delayed and only reactive to harm which attracts media attention. Public pressure and a desire to maintain brand integrity is insufficient in ensuring a transparent and proactive approach to addressing harm.

Option 1 - online safety framework

65. Option 1 introduces a new regulatory framework establishing legal duties on companies to improve and protect the safety of their users online, overseen and enforced by Ofcom, an independent regulator.

Platforms in scope

66. The new regulatory framework will apply to:

- any service which hosts UGC which can be accessed by users in the UK; and/or
- any service that facilitates private or public interaction between service users, one or more of whom is in the UK; and
- search services; and
- any service which publishes pornographic content which can be accessed by users in the UK

67. In response to stakeholder views expressed through the public consultation, the Government incorporated the following exemptions for specific types of services:

- **'Low risk functionality' exemption**: The OSB exempts 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 platform, 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' UGC and enable interactions between members within that organisation. This encompasses online services which are used internally by 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 UGC on their platform. In practice, this takes out network infrastructure such as ISPs, 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 offered to businesses where again the business doesn't actually have control over specific pieces of content or activity.
- **Educational institutions**: 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. It also includes platforms like intranets and cloud storage systems, but also "edtech" platforms.
- **Email and telephony**: Email communication, voice-only calls and short messaging service (SMS)/multimedia messaging service (MMS) remain outside the scope of legislation.

68. Furthermore, business-to-customer interactions are not considered UGC 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.

69. Based on analysis conducted by Revealing Reality (RR) and explained in more detail later in the IA, the Government expects approximately 25,100 platforms to fall within scope of the online safety framework. The estimate presented in the previous IA was 24,000. This does not reflect any change in methodology but simply represents the inclusion of pornography providers and the first year of the appraisal period as

[74] Understanding how platforms with video-sharing capabilities protect users from harmful content online - (EY, 2021)

2024 with full compliance with the regime from 2025[75] (the number of businesses grows in line with average growth rate in firms - 3% between 2000-2020). Prior to the announcement of the above exemptions, 180,000 organisations were expected to fall within scope (this reflects a reduction of around 160,000 organisations).

Harmful content in scope

70. The OSB seeks to address the following broad categories of harmful online content:

    - **illegal UGC and activity** which is an offence under UK law - such as CSA, terrorism, hate crime and sale of illegal drugs and weapons; and
    - **legal but harmful UGC and activity** which may not amount to an offence, but which gives rise to a foreseeable risk of psychological and physical harm to adults and children - such as abuse or eating disorder content; and
    - **underage exposure to UGC and activity** which gives rise to a foreseeable risk of psychological and physical harm to children - such as underage access to pornography, violent content which is not appropriate for younger children.
    - **underage exposure to pornographic provider content which** is published and not user generated

71. The OSB does not seek to address UGC which gives rise to a foreseeable risk of harm to corporations and organisations and their interests (e.g. copyright offences, competition law). In addition, a number of categories of UGC and activity are specifically excluded from the scope of the OSB 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.

Categories of regulated services

72. To ensure proportionality, the online safety framework will establish differentiated expectations on companies in scope, with regard to different categories of harmful content and additional requirements outside of the core duty of care. The OSB creates different categories of regulated services, these are as follows:

    - user to user services meeting the Category 1 thresholds;
    - search services meeting the Category 2A thresholds;
    - user to user services meeting the Category 2B thresholds; and

73. Thresholds for these categories will be set out in secondary legislation; however, they will relate to a platform's number of users and its functionalities and therefore, the risk of harm on the platform. At a high level, Category 1 platforms are likely to be the highest risk and highest reach user to user platforms, such as a small group of the largest social media sites and pornography sites. The same principle applies to Category 2A but relates to the highest risk and highest reach search services, such as a small group of the largest online search engines. Category 2B services are expected to be high-risk, high-reach platforms but that may not necessarily meet the Category 1 threshold. Based on current policy intention, between 30-40 platforms are expected to be designated as either Category 1, 2A, or 2B.

74. In addition, on the basis of regulating pornographic provider content, pornography publishers that do not host UGC or enable P2P interaction will be in scope of the OSB. These platforms will only be required to comply with the provision on published pornography ("pornography provision") and will not be in scope of the core safety duties.

Core platform safety duties

75. The primary responsibility for each company in scope will be to take action to prevent UGC or activity on their services causing significant physical or psychological harm to individuals. The table below outlines which categories of regulated services are expected to comply with each of the core duties.

[75] In reality, different aspects of the Bill are likely to come into force over a period of time around the assumed analytical start date and the chosen appraisal period represents an analytical simplification. The Government's intention is to have the regime operational as soon as possible after Royal Assent, whilst ensuring the necessary preparations are completed effectively and services understand what is expected of them.

Table 2: Differentiated duties on in-scope companies

Duty	All UGC services	Category 1	Category 2A	Category 2B	Pornography publishers[76]
Risk assessment duty: to assess the level of risk on the platform	✔	✔	✔	✔	✕
Illegal duty: to put in place systems and processes to minimise and remove priority illegal content and to remove non-priority illegal content when identified through user reporting.	✔	✔	✔	✔	✕
Child safety duty: If the platform is likely to be accessed by children, to put in place systems and processes to protect children from harmful content.	✔	✔	✔	✔	✕
Legal but harmful duty: to address legal but harmful content accessed by adults, through enforcing a platform's own terms of service.	✕	✔	✕	✕	✕

76. To comply with these core duties they will complete an assessment of the risks associated with their services and take reasonable steps to reduce the risks of harm they have identified occurring. The steps a company needs to take will depend, for example, on the risk and severity of harm occurring, the number, age and profile of their users and the company's size and resources. Companies will fulfil their duty of care by putting in place systems and processes that improve user safety on their services. These systems and processes could include, for example, user tools and content moderation procedures.

77. Robust protections for freedom of expression have been built into the design of duties on companies. Companies will be required to consider users' rights, including freedom of expression online, both as part of their risk assessments and when they make decisions on what safety systems and processes to put in place on their services.

78. The Government will set out priority categories of legal but harmful content and activity in secondary legislation and identify priority categories of offences. This will focus companies', and the regulator's, efforts on the most harmful issues. Companies will still be required to tackle other relevant non-priority material on their services.

Additional requirements on platforms

79. All companies in scope will have a number of additional requirements beyond the core duties of care. The table below outlines which categories of regulated service are expected to comply with each of the additional requirements

Table 3: Additional requirements beyond the core duties of care

Duty	All UGC services	Category 1	Category 2A	Category 2B	Pornography publishers[77]
Pornography provision: to prevent children from accessing published pornographic content.	✔	✔	✔	✔	✔

[76] That do not host UGC or enable P2P interaction and are therefore not in scope of the core duties.
[77] That do not host UGC or enable P2P interaction and are therefore not in scope of the core duties.

User reporting: to provide mechanisms to allow users to report harmful content or activity and to appeal the takedown of their content.	✔	✔	✔	✔	×
CSA content: If the platform is a UK platform or is a non-UK platform that does not already report, to report identified online CSA.	✔	✔	✔	✔	×
Transparency: to publish reports containing information about the steps they are taking to tackle online harm on those services.	×	✔	✔	✔	×
Fraudulent advertising: to minimise the publication and/or hosting of fraudulent advertising.	×	✔	✔	×	×
User empowerment: to offer optional user identity verification and user empowerment tools to give users more control over their online experience.	×	✔	×	×	×
FoE and privacy: to produce freedom of expression and privacy impact assessments	×	✔	×	×	×
Protected content: to protect journalistic content and content of democratic importance	×	✔	×	×	×

80. Option 1 includes a specific provision which requires pornography publishers to prevent children from accessing published pornographic content (i.e. non user-generated content). The pornography provision does not form part of the core child safety duties, but will be enforced by Ofcom with providers being subject to the same enforcement measures as other in-scope services. The pornography provision does not capture user-to-user content or search results presented on a search service, as the draft Online Safety Bill regulates these separately (under the Bill's core duties). Platforms in scope of the Bill's core duties which also carry published (i.e. non user-generated) pornographic content would be subject to both the wider provisions in the draft Bill for user-to-user services and the pornography provision. The pornography provision will deliver an equivalent outcome to the duties for user-generated pornography, in preventing children's access to pornographic content.

81. In addition, all in scope platforms in scope of the core duties will have to provide mechanisms to allow users to report harmful content or activity and to appeal the takedown of their content. Users must be able to report harm when it does occur and seek redress. They must also be able to challenge wrongful takedown and raise concerns about companies' compliance with their duties. All companies in scope will have a specific legal duty to have effective and accessible reporting and redress mechanisms. This will cover harmful content and activity, infringement of rights (such as over-takedown), or broader concerns about a company's compliance with its regulatory duties. Expectations on companies will be risk-based and proportionate, and will correspond to the types of content and activity which different services are required to address. For example, the smallest and lowest risk companies might need to give only a contact email address, while larger companies offering higher-risk functionalities will be expected to provide a fuller suite of measures.

82. The OSB also introduces a legal requirement on technology platforms to report online CSA content. Introducing a CSA reporting requirement on UK (and some non-UK) platforms will ensure that they 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. This requirement will apply

differently to platforms depending on where they are based, which is different from the approach being taken within the regulatory regime, where duties will apply to all in-scope services that have UK users. UK platforms (those that provide services from within the UK) will be required to report all identified CSA (all CSA offences set out in the OSB) to a new designated body. Platforms providing services from outside of the UK will only have to report identified CSA offences that are perpetrated by a UK user, and only if they do not already report CSA.

83. The remaining additional requirements apply only to a small group of high-risk and high-reach services. Category 1, 2A and 2B platforms will be required to publish transparency reports containing information about the steps they are taking to tackle online harm on those services. Option 1 does provide the Secretary of State with a power to expand the scope of transparency reporting beyond Category 1, 2A and 2B platforms 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. Category 1 and 2A platforms will also be required to minimise the risk of fraudulent advertising appearing on their platform. While specific steps will be set out in future codes of practice (subject to consultation and IAs), this will likely include conducting some form of additional checks on advertisers, sharing information on known fraudulent advertisers, and removing fraudulent adverts when reported by users.

84. An even smaller group of the highest risk and highest reach user-to-user platforms only will have a number of further requirements with which to comply. These include implementing optional user empowerment tools which are likely to involve allowing users to verify their identity and giving users the ability to filter the type of content they see on those platforms. They will also need to assess their impact on freedom of expression and privacy and publish this in the form of an impact assessment.

Preferred option and implementation plan

85. **Option 1** is the Government's preferred option and is the result of extensive engagement with platforms, Parliament, civil society organisations, and wider society. The preferred option is risk-based and proportionate and is expected to achieve the stated policy objectives. Importantly, the preferred option does not place undue burdens on platforms where there is a low, or no, risk of harm.

86. The OSB is being introduced in 2022 and the Government expects passage to take 10-12 months, which means Royal Assent is expected in 2023. These timelines are estimates, and are subject to parliamentary time. The Government's intention is to have the regime operational as soon as possible after Royal Assent, whilst ensuring the necessary preparations are completed effectively and services understand what is expected of them. The Government is already working closely with Ofcom to ensure that the implementation period that will be necessary following passage of the legislation is as short as possible.

87. 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.

Appraising the preferred option

Approach to appraisal

88. At this primary stage, it is not possible to predict with certainty the actions of Ofcom or the steps platforms may take to ensure they are compliant with the regulation. While the OSB sets out a duty of care, the specific requirements and the actions platforms can take to comply will be set out in codes of practice laid by Ofcom and where necessary secondary legislation. Given that specific requirements are unknown at this stage, costs and benefits included here are largely illustrative and aim to indicate the potential scale or nature of impacts of the whole policy (scenario 2 in the Regulatory Policy Committee's

(RPC's) primary legislation guidance).[78] All future codes of practice will be subject to an IA, including an assessment of the impacts on small and micro-businesses (SMBs) and innovation (IA requirements on Ofcom under the OSB go beyond current regulator requirements under the Small Business, Enterprise and Employment Act 2015).[79]

89. While it is not possible at this stage to provide a fully monetised appraisal of the policy or a verifiable assessment of the EANDCB, every effort is made to provide an indication of the likely scale of impact of the whole policy (including future codes) through presenting illustrative monetised costs, proxied impacts from similar policies, and comprehensive qualitative analysis.

90. For this IA's assessment of potential benefits, it is not possible to develop a precise estimate of the reduction in online harm that will be achieved by the preferred option. Instead, this IA attempts to quantify the economic cost of online harm under a *do nothing* counterfactual and conducts both break even analysis and scenario analysis based on a range of illustrative harm reduction scenarios.

91. While timelines are dependent on external factors, for appraisal purposes, this IA uses a ten-year appraisal period running from 2024. Familiarisation costs are assumed to be incurred in the first year of the appraisal period with full compliance from 2025. This approach is an analytical simplification - in reality, codes of practice are likely to be staggered each allowing time for businesses to ensure compliance and are unlikely to fall in line with calendar years. All impacts are presented in 2019 prices and 2020 present value base year.

Main sources of evidence

92. This final stage IA supporting the OSB, draws on a number of evidence sources to attempt to provide an indication of the likely scale of impacts.

- **RR research**:[80] In 2020, DCMS commissioned consultancy firm RR to estimate the number of organisations in scope of the framework and to determine the likely incremental costs of compliance. More details on the methodology are included later in this IA, and the results form the basis of our current estimates
- **AVMSD research:**[81] DCMS commissioned consultants from EY to research the measures that VSPs take to protect users online ahead of the implementation of the rules for VSPs under AVMSD. The Directive sets requirements on VSPs (e.g. YouTube) to protect users from harm. Ofcom is the regulator for UK-established VSPs and therefore, actions taken and costs incurred by in-scope businesses represent another reasonable proxy for the costs of the OSB. Note that qualitative and quantitative evidence was collected from platforms within and outside of the UK's jurisdiction.
- **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 Germany's NetzDG which is 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 OSB, both address (or aim to address) online harm 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 harm.
- **Business engagement**: Since the publication of the OHWP, DCMS has engaged extensively with affected platforms. Engagement includes a series of bilaterals, roundtables, and a cost survey. In addition to a number of bilaterals with in-scope platforms, DCMS has held a number of

[78] RPC case histories – primary legislation IAs, August 2019
[79] Small Business, Enterprise and Employment Act 2015
[80] Not yet published - will be published soon
[81] Understanding how platforms with video-sharing capabilities protect users from harmful content online (EY, 2021)

roundtables with industry on the topic of compliance costs and issues relevant to small and medium sized enterprises. In addition, following the publication of the full government response, DCMS sent cost surveys to a sample of 36 platforms to understand in greater detail how they 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. The results of this engagement is discussed throughout but is mainly qualitative given that platforms were largely unable to provide cost information without knowing the content of future codes. A selection of UK-focussed AV providers were also engaged with a cost survey, the results of this can be found in the age assurance cost section.

- **Ofcom call for evidence on VSPs**:[82] Ofcom published their call for evidence on VSP regulation on 20 July 2020. There were 39 non-confidential responses which included social media platforms, platforms with video sharing capabilities, public sector institutions and individuals. The findings from the call for evidence were used to inform the development of the draft and final guidance on the VSP regime. The call for evidence was divided into two parts: i) Queries for industry which included questions on the services provided and in particular on the mechanisms for keeping users safe online; ii) Questions for all stakeholders which included queries on how the design of the VSP regulatory regime can best keep users safe online.

Costs and benefits

Baseline

93. Evidence on the current level of harm mitigation under the baseline is limited. The systems and processes platforms have in place vary significantly across platforms, as does spend on user safety. For some platforms keeping users safe online is part of the organisation's ethos and for others activities such as content moderation is much lower in their priorities.

94. 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. This was supported in engagement with stakeholders. The vast majority of platforms engaged already conduct risk assessments, set terms of service and acceptable use policies, conduct both human and automated moderation, allow users to report harm, and have systems to handle complaints.

95. RR research also found that different types of mitigation are implemented to varying degrees. For instance, while automated moderation is used throughout, the complexity and tailoring of this to the specific platform varies. For example, a low-risk organisation interviewed uses 'off the shelf' automated moderation to detect spam, whereas a high-risk organisation uses their own bespoke automated software tailored to detect specific harm present on their site. Findings from the RR research on high variability in current mitigations was further corroborated by EY's AVMSD research of VSPs. 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.

96. Most organisations are already investing in protecting their users in the absence of regulation and platforms expect this investment to continue to increase over time. EY's AVMSD research found that total annual expenditure on measures to protect users from harmful content ranged from hundreds of pounds for the very smallest platforms to over £1.5bn for the largest platforms.

97. Organisations and online platforms have also dedicated significant resources to specifically tackling online child sexual exploitation and abuse (CSEA). In March 2020, following a Five Country ministerial meeting between the UK, US, Australia, Canada and New Zealand, the UK launched the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse to further promote and enhance this

[82] Consultation on guidance for VSP providers on measures to protect users from harmful material (Ofcom, 2021)

work. These principles set out a consistent and high-level framework for industry actors, aiming to co-ordinate the approach to tackling online CSEA globally, outside of formal regulation. The promotion of these principles has been supported by the WePROTECT Global Alliance across 97 governments, 25 technology companies and 30 civil society organisations.

98. While accurate evidence does not exist 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 Option 1, i.e. only the incremental costs of regulation have been included.

99. As outlined in the above rationale for intervention, many categories of online harm 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 harm results in a societal cost of at least £136 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 subset of harm that this IA was able to quantify.

Summary of impacts

100. All impacts are assessed over a 10-year appraisal period starting from the date of implementation. 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. Given the uncertainty around future requirements, costs and benefits are illustrative and attempt to provide an indication of the likely scale of impact from primary legislation, related secondary, and future codes of practice. Three scenarios are set out in Table 4: a low, central and high scenario. Details and assumptions underpinning these scenarios are outlined in further detail in each of the accompanying cost sections that follow the table.

Table 4: Summary of impacts

Impact	Treatment	Low	Central	High
Reading and understanding the regulations	Cost to business (monetised)	£9.6 million	£12.0 million	£17.5 million
Ensuring users are able to report harm	Cost to business (monetised)	£17.7 million	£23.1 million	£33.8 million
Updating terms of service	Cost to business (monetised)	£17.8 million	£23.1 million	£33.6 million
Conducting risk assessments	Cost to business (monetised)	£17.5 million	£33.1 million	£48.7 million
Undertaking additional content moderation	Cost to business (monetised)	£1,319.1 million	£1,902.6 million	£2,486.2 million
Employing age assurance technology	Cost to society (monetised)	£17.9 million	£35.8 million	£89.6 million
Transparency reporting	Cost to business (monetised)	£0.8 million	£6.3 million	£10.3 million

Fraudulent advertising duty	Cost to business (monetised)	£64.6 million	£145.8 million	£226.7 million
User verification and empowerment duties	Cost to business (partially monetised)	£8.8 million	£11.9 million	£13.7 million
FoE and privacy IAs	Cost to business (monetised)	£1.1 million	£2.7 million	£11.5 million
Reporting online CSA to designated body	Cost to business (non-monetised)	n/a	n/a	n/a
Industry fees	Cost to business (monetised)	£313.9 million	£313.9 million	£313.9 million
Enforcement action (fines and business disruption measures)	Cost to business (non-monetised)	n/a	n/a	n/a
Justice impacts	Cost to government (monetised)	£0.3 million	£0.3 million	£0.3 million
Wider impacts (freedom of expression, privacy, competition, innovation, trade)	Cost to society (non-monetised)	n/a	n/a	n/a
Reduced prevalence of online harm	Benefit to society (non-monetised)	Break-even: 1.5%	Break-even: 2.1%	Break-even: 2.7%

Approach to business costs

Number of platforms in scope

101. Previous estimates for the number of platforms in scope were not challenged in response to the consultation stage IA. Given the wide-ranging scope of Option 1 and the lack of granular data, it is difficult to determine with precision the number of affected organisations. However, this IA considers the sampling approach used in the previous IA and explained below to be the most robust existing evidence on in-scope firms and the methodology therefore, remains unchanged. The number of affected platforms (and Civil Society Organisations, CSOs) within scope of the regulations is estimated to be around 25,100[83] in the first year of the appraisal period. This equates to between 0.3-0.4% of all UK businesses.

102. To determine the number of platforms in scope, RR extracted a stratified sample of 500 organisations from the Inter-Departmental Business Register (IDBR).[84] The sample consisted of 100

[83] The exact estimate is 25,051

[84] A comprehensive list of UK businesses used by the government for statistical purposes.

randomly selected businesses in each of the following size categories (sole traders, micro (not including sole traders), small, medium and large).[85] 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 ±2.6 to 4.4 percentage points). Additionally, every organisation within the sample had to be manually reviewed and categorised according to features that could be considered as hosting UGC or enabling P2P interaction. Findings were then extrapolated using the Department for Business, Energy and Industrial Strategy's (BEIS) Business Population Estimates (BPE)[86] to estimate the total number of in-scope platforms 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 initial sampling and extrapolation resulted in an estimate of approximately 18,000 in-scope platforms.

103. Option 1 will apply to CSOs as well as businesses. While the IDBR (from which the original sample was taken) does include CSOs, BEIS' BPE does not. To address this, findings from the original sample were further extrapolated using data on CSOs in the UK Civil Society Almanac[87] and around 550 CSOs were added to the estimates. This is a reasonably reliable methodology for determining the number of CSOs in scope; however, it does have limitations:

- 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.
- The estimate of CSOs in scope is likely an overestimate. When CSOs do utilise UGC this is mainly through third-party sites like Facebook, Twitter, and Youtube who themselves would be in scope rather than the CSO.
- 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.[88]

104. In line with the consultation stage IA, 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 harm occurring on the platform). There has been no evidence submitted as part of pre-legislative scrutiny or in response to the consultation stage IA to suggest that this assumption should be revised. In addition, Option 1 is functionality based and is sector agnostic. An in scope CSO with the same risk profile as an in scope business would incur the same costs.[89]

105. Acknowledging the potential for gaps in the random sample (for example the lack of in-scope small platforms), 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 platforms 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 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 platforms 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 to account for an implementation date of 2024. For modelling purposes, this growth rate continues throughout the appraisal period.

106. Steps taken so far focus on determining the number of platforms in scope based on whether they host UGC, enable P2P interaction or are search engines - these platforms are in scope of the core duties and are the main regulated entities. The pornography provision is an additional requirement on pornography publishers to prevent children from accessing non-user-generated pornographic content, regardless of whether they are in scope of the core duties. As highlighted by a study of children's access to pornography, explicit adult content can be found on a variety of platforms, including social media sites,

[85] The definition is in line with SBEE Act.

[86] An estimate of the total number of private sector businesses in the UK at the start of each year, with their associated employment and turnover.

[87] UK Civil Society Almanac 2020 - NCVO

[88] CSOs are matched to the business size categories based on average revenue by business size as presented in BEIS' BPE.

[89] Of course, requirements would be proportionate to both risk and resources available.

VSPs, search engines, chat sites, and dedicated pornography sites.[90] The majority of these types of platforms are already in scope of the core duties as user-user services and will be captured in the estimates above.

107. Even though many of the most visited pornography sites and sources are in scope of the Bill's core duties, an important proportion of dedicated pornography sites are not as they do not host UGC or enable P2P interaction and so are only in scope of the pornography provision. To illustrate this, an assessment of the top 200 pornography sites found that 36% of sites (or 72 sites) and 16% of traffic was to sites outside of the scope of the Bill's core duties. The pornography provision will ensure these sites protect children from harmful content.

108. Given the nature of the industry, evidence on the number of pornography publishers and the location of their economic activity is limited, it is difficult to determine with certainty the number of UK-based pornography publishers that do not host UGC or enable P2P interaction. Based on an assessment of the top 200 pornography sites most popular with UK users, the BBFC found that only four were based in the UK (out of 126 for which this information was available). The vast majority of pornography sites are based in the US and, even there, industry reports on the US market put the number of businesses operating pornographic websites at only 89.[91] The same report estimates that a single organisation, namely MindGeek, holds an 80% market share and owns many of the most popular sites. One report - although conducted in 2013 - estimates that 60% of pornography sites are hosted in the US, compared with 7% in the UK.[92] Using sites as a proxy for the number of businesses based in each country and comparing the US to the UK, this would suggest that the number of UK-based pornography publishers in 2020 is likely be around 10, broadly in line with low number of sites based in the UK from BBFC's research on country of origin. Uplifted by the average annual growth in the business population to the first year of appraisal, this impact assessment conservatively estimates an additional 11 UK-based businesses in scope as a result of the pornography provision.[93] The number of businesses does not reflect the number of pornography sites, as each business is likely to operate multiple sites - as is the case in the US market.[94]

109. Following the above steps, the final estimate for the number of in-scope platforms is approximately 25,000 organisations.

Table 5: Steps to attain an estimate for the number of in-scope platforms

	Micro	Small	Medium	Large	Running total
Percentage in-scope within sample	0.3 %[95]	0 %	2 %	4 %	-
Number of in-scope platforms within UK economy (nearest hundred)	17,100	0	800	400	18,300
Number of in-scope CSOs within UK economy[96] (nearest hundred)	400	0	100	<100	18,900
Accounting for gaps in sample with known types of platform	-	~1,000	~2,000	-	21,600

[90] Another potential source of online pornography are UK-based video-on-demand (VoD) sites. The impact assessment for Part 3 of the Digital Economy Act estimated that there were around 100 of these based in the UK. More recent estimates suggest around 150, with 40-50 of these being adult services. VoDs are not in scope of this Bill as they are already subject to existing regulation (the UK's transposition of the Audiovisual Media Services Directive) which includes protecting children from pornography as well as wider duties related to product placement, sponsorship, and incitement to racial hatred.

[91] Adult & Pornographic Websites Industry in the US - Market Research Report (IBIS World, 2020)

[92] Top 10 adult website host countries (Metacert, 2013)

[93] Of course, some of these businesses are likely to have already been captured as user-user services but this impact assessment conservatively assumes that all are additional rather than a proportion. As these businesses are not in scope of the core duties, there is no risk of double counting costs based on this conservative approach.

[94] The number of businesses does not reflect the number of pornography sites, as each business is likely to operate multiple sites - as is the case in the US market.

[95] Weighted data combining 0 employee and 1-9 employee bands

[96] Note the size of CSOs is determined by annual revenue in line with appraisal practice in this area.

Number of non-UGC pornography publishers[97]	9	1	0	0	10
Number of in-scope platforms uplifted to 2024[98] (nearest hundred)	20,200	1,200	2,900	700	**25,100**

110. The methodology described above was conducted both before and after the Government announced a list of exemptions for specific types of services in December 2020. Before the exemptions, it was estimated that around 3% of all UK businesses would have been in scope, equating to approximately 180,000 platforms. The exemptions therefore removed approximately 155,000 platforms from the scope, mostly SMBs exempted by the low risk functionality exemption. This IA conducts sensitivity analysis on the number of businesses in scope in the risks and sensitivity section.

Risk categorisation of platforms

111. Option 1 is risk-based which means that there are differentiated expectations on companies in scope with regard to different categories of harmful content and the additional requirements outside of the core duty of care (see Table 2). In addition, even within the differentiated platform duties, expectations on platforms will differ depending on the risk of harm on their platform and the resources available to the platform. For example, while every platform will be required to assess the risk of illegal harm on their platform, the level of detail required and the steps they have to take in producing these risk assessments will vary greatly. This approach ensures proportionality both in the differentiated duties and in the specific way in which platforms can comply with codes. Platforms which offer services with the lowest risk of online harm will face the lowest regulatory burdens and platforms offering high-risk services will be required to take the most action.

112. Given that the specific way in which platforms can comply will be set out in future codes of practice, it is not possible to know exactly what they will do. However, to reflect this proportionality in the analysis of businesses' impacts, the Government commissioned the production of an organisation categorization framework (OCF) to split platforms into three risk tiers (low, mid and high) which helps with estimating the type of likely actions they would take in complying with Option 1. The OCF was developed using extensive desk research and interviews with experts, such as the IWF, Childnet, and Internet Matters.

113. The OCF first identified all factors that could define whether or not an organisation could fall in scope of the OSB and factors that could affect its ability to tackle online harm. The two primary categorisation criteria incorporated into the OCF were 'features' and platform size (as measured by the number of employees). There were 41 features that enabled users to share or discover UGC or enable peer-to-peer (P2P) interaction assessed as part of developing the OCF. These included features such as the ability to livestream, share content that exists on the platform, like and dislike content, group message, video call, post comments under content, geo-tag, and display a feed of UGC. The OCF was used for research purposes only and is not directly related to the contents of the OSB or a checklist by which platforms can determine whether or not they are in scope. Instead, it is a set of criteria which enables manual assessment of sample platforms.

114. The categorisation of in-scope platforms 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 likely to be broadly 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'). Scoring based on the OCF indicated that the majority of the around 25,100 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 0.001% are estimated to meet the Category 1 and 2A thresholds

[97] The proportion of additional pornography publishers estimated to fall within each size category is based on business demographics within creative industries. DCMS Sectors Economic Estimates 2019: Business Demographics (DCMS)

[98] Start of the appraisal period and expected date of implementation

(additional requirements on the largest and highest risk platforms, based on policy intention this is expected to be around 20 platforms).

115. Platforms in scope will vary greatly as will the way in which they offer functionality that allows UGC and P2P interaction. Table 6 illustratively provides some examples of the types of organisations that could fall within each risk category:

Table 6: Example types of organisations within risk categories

Risk tier	Example features within sample	Example organisations
Low risk	● Comments sections (for UGC) ● Ability to like content	● Retail websites (that are not out of scope due to the limited functionality exemption) ● Blogging platform
Mid risk	● Ability to post content ● Message someone you know or have friended	● Forums ● Dating sites ● Online gaming
High risk	● Feed of UGC ● Live Streaming ● Ability to contact unknown users	● Social media companies ● Large search engines ● Streaming services

116. It should be noted that Table 6 is illustrative and used for analytical purposes only. It presents example types of organisations that may fall within the different risk categories based on current understanding of the types of functionality present on these platforms.

Development of platform actions

117. It is difficult at this stage to estimate with certainty the steps platforms will take and the costs they will incur as a result of complying with the OSB. This is because:

- Option 1 establishes differentiated expectations on companies in scope with regard to different categories of harmful content and the additional requirements outside of the core duty of care. Thresholds for Category 1, 2A and 2B will be set out in secondary legislation and therefore, it is unclear at this stage which organisations they will apply to.
- Option 1 is proportionate even within duties, and expectations will vary greatly between for example small low risk platforms and large high risk platforms. Specific steps in-scope platforms can take will be outlined in future codes of practice laid by Ofcom (themselves subject to IAs).
- Companies will need to comply with the codes; however, if preferred, they will also be able to demonstrate to the regulator that an alternative approach is equally effective.
- Even while some aspects of the OSB will clearly result in specific actions such as conducting risk assessments or transparency reporting (for Category 1, 2A and 2B), the steps platforms can take and the information required in these will not be set out until future codes of practice.
- The high-level duties related to illegal content, legal but harmful content, and protecting children set out at primary stage legislation are not prescriptive and therefore, any attempt to estimate the specific actions taken by platforms is by definition speculative.

118. A common theme of the Government's engagement with in-scope platforms is that they are unable at this stage to provide reasonable estimates of costs or even actions likely to be taken to comply. This is to be expected at this stage and following introduction, Ofcom will begin a series of consultations with industry on codes of practice and will produce IAs to determine the costs to platforms.

119. Given the uncertainties at the primary stage, this IA develops a plausible set of actions platforms may take based on estimates of the size and risk of harm on the platform. These include:

- **Reading and understanding the regulations (familiarisation costs)** - this includes both primary legislation and related secondary, and future statutory codes of practice

- **Ensuring users are able to report harm** - this relates to the mechanism through which users can report harm and could be as simple as a visible email address (already a statutory requirement) or a system which can triage large volumes of reports.
- **Updating terms of service** - evidence discussed below suggests that this is a business-as-usual activity for in-scope platforms. However, platforms may decide to assess and update their terms of service in response to future codes of practice.
- **Conducting risk assessments** - this relates to the requirement to carry out an illegal content risk assessment and 'if likely to be accessed by children' to carry out a children's risk assessment. For Category 1 platforms, as part of this they will also have to assess the risk of legal but harmful content.
- **Undertaking additional content moderation** - the OSB does not require additional content moderation; however, it is likely that platforms will increase resources in this area to comply with the duties.
- **Employing age assurance technology** - in complying with the child safety duties, some higher risk platforms are likely to adopt age assurance (and specifically age verification) technologies.
- **Transparency reporting** - this relates to producing annual published reports on platform harm and related actions taken by the platform.
- **Fraudulent advertising duty (customer due diligence)** - as part of complying with the fraudulent advertising duty, it is likely that in-scope platforms will conduct CDD (customer due diligence) on advertisers.
- **User verification and empowerment duties** - this relates to the requirement on large social media platforms to offer optional user verification and provide user empowerment tools.
- **Assessing impacts on freedom of expression and privacy** - this relates to publishing an assessment of impacts on freedom of expression and privacy and keeping this updated.
- **Reporting online CSA to designated body** - this refers to the cost of reporting identified CSA content to the relevant designated body.

120. Cost estimates for this plausible set of platform actions is based on evidence provided by platforms, proxied from similar regulations, or based on reasonable assumptions of time requirements and standard appraisal practice.

Costs to business

121. For appraisal purposes, it is assumed that legislation enters into force in 2024. The first year is assumed to be a transition year giving platforms time to prepare for compliance based on the specific details set out in codes of practice and secondary legislation. This IA assumes that platforms will incur familiarisation costs and transition costs in the first year but will not incur compliance costs until year two. This is a simplified assumption for analytical purposes only - in reality, the codes of practice will be staggered and platforms will ensure compliance across a number of years.

Familiarisation costs

Requirements

122. In-scope platforms[99] will be expected to familiarise themselves with the regulations which includes understanding which aspects of the safety duties apply to them and what steps they must take to ensure compliance.

Cost estimates

123. Platforms are expected to incur the following costs associated with familiarisation:

- **Initial familiarisation**: while only in-scope platforms are required to familiarise themselves some, who think they could potentially be in scope,[100] under a broad interpretation of the regulations, may have to read the regulations - even if only to determine that they were out of scope.

[99] Including pornography providers that are in scope of the pornography provision but not the core duties. These platforms are expected to incur full familiarisation costs.

[100] Those which offer online services with any features that could be considered in-scope such as posting, sharing, reacting to content, messaging, calling, commenting, tagging, discovering or seeing UGC.

- **Potential legal advice for SMBs**: in-scope SMBs may require legal advice to clarify aspects of scope and which parts of the OSB apply to them
- **Secondary familiarisation**: Beyond the initial familiarisation, actual in-scope platforms are expected to spend more time reading the regulations
- **Dissemination of information**: medium and large in-scope platforms are expected to disseminate the information across a proportion of their staff

124. For initial familiarisation, based on RR research, there are approximately 180,000 platforms that could be considered potentially in scope. For initial familiarisation, it is estimated that between 20%-50% of all platforms potentially in-scope would read the regulations (25% in the central scenario) - this is approximately 20,000 out-of-scope platforms incurring costs of familiarisation. As with other regulations, it is very difficult to predict with certainty how many firms outside of scope would incur costs of familiarisation - evidence for this within the context of online harms is extremely limited. The assumed range merely represents a conservative estimate to provide an indication of the likely scale of impact on out-of-scope platforms. 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 £20.62 is expected to read the regulations within each business (all wages in this IA come from the Annual Survey of Hours and Earnings[101] and are uplifted by 22% to account for non-wage labour costs). The explanatory notes are approximately 52,000 words and would therefore take just over four hours based on a reading speed of 200 words per minute.[102] This results in the cost of initial familiarisation of between £3.2 million and £8.0 million (central, £4.0 million).

125. In addition, the central estimate includes one hour of legal advice for every in-scope SMB. Legal advice is not included in the low estimate and rises to two hours for every in-scope SMB in the high estimate. The inclusion of legal advice represents the potential need to confirm whether a platform does fall within scope and to advise on which aspects of the OSB are likely to affect them. While many SMBs may not require this, some will likely seek more extensive legal advice by assuming one hour for every firm this IA attempts to capture the total cost rather than provide an accurate per platform estimate.[103] This IA estimates the cost of legal advice to be between £0 and £1.7 million (central, £0.8 million).

126. For secondary familiarisation, in-scope platforms are expected to spend more time reading the regulations. For these (around 25,000), another member of staff in micro-platforms (rising to 2, 5, and 10 for small, medium and large platforms respectively) is expected to read the legislation's explanatory notes. For medium and large platforms, these staff are expected to be regulatory professionals whereas wage estimates for Chief Executives are used for in-scope SMBs. Secondary familiarisation is expected to result in costs of £5.7 million - this is uniform across the range of estimates.

127. Finally, for medium and large in-scope platforms, costs are expected to be incurred through disseminating the information across a proportion of their staff. While it is unclear what exact proportion of staff will need to be made aware of the regulations, this IA estimates that between 5%-20% of staff within in-scope medium and large platforms will spend 30 minutes familiarising themselves (10% in the central scenario). This could be through a staff meeting or engaging with a summary email. Dissemination is expected to result in costs of between £0.7 million and £2.1 million (central, £1.4 million).

128. Following the methodology noted above, familiarisation costs are estimated to total between £9.6 million and £17.5 million (central, £12.0 million).

129. It should be noted that costs estimated above cover only familiarisation of the primary legislation. There will be additional costs to platforms 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 they will have to familiarise themselves with in order to comply. However, by using Ofcom's Electronic Communications Code[104] as a proxy this IA can provide an indication of the likely scale of these impacts for one particular code.

[101] Annual Survey of Hours and Earnings (ONS)
[102] Business Impact Target Appraisal Guidance - BEIS
[103] The wage of a legal professional is used here.
[104] Electronic Communications Code - Ofcom

130. 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 platforms would have to read between 116-244 pages - on the assumption that all sections were relevant to that particular platform. 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). To illustrate what this could mean in the context of the OSB, using the wage of a regulatory professional for medium and large platforms and a Chief Executive for SMBs, the total cost could range from £6.3 million to £9.2 million.

131. Previous estimates for familiarisation costs were not specifically challenged in response to the consultation IA. However, based on the qualitative evidence from engagement with in-scope platforms, the above approach reflects the following changes to previous estimates:

- The individual(s) within SMBs expected to familiarise themselves with the regulation has been changed from regulatory professionals to Chief Executives. While use of regulatory professional wages was only a proxy, this now better reflects that owners of smaller platforms are likely to be the ones who conduct familiarisation, a point noted by SMBs engaged and advised by the RPC.
- The addition of potential legal advice for in-scope SMBs.

132. Table 7 outlines the range of expected costs associated with reading and understanding the regulations:

Table 7: Reading and understanding the regulations (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Reading and understanding the regulations	£9.6 million	£12.0 million	£17.5 million

Transition costs

133. Table 8 sets out the total transition costs across the policy options. Details on how these costs have been estimated is below.

Table 8: Transition costs (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Transition costs	£35.5 million	£46.1 million	£67.4 million

134. Platforms are expected to incur the following transition costs:

- **Ensuring users are able to report harm** - this relates to the mechanism through which users can report harm and could be as simple as a visible email address (already a statutory requirement) or a system which can triage large volumes of reports.
- **Updating terms of service** - platforms may decide to assess and update their terms of service in response to future codes of practice.

Ensuring users are able to report harm

Requirements

135. Under the framework, platforms will be expected to accommodate user reporting of harm and provide an avenue for user redress (challenge of content removal). User reporting and redress mechanisms are expected to vary across platforms. For example, for the smallest lowest-risk, they may only be required to have an email address visible on their service (already a legal requirement under the Electronic Commerce Regulations 2002[105]) while high risk platforms may require reporting mechanisms which can handle and triage larger volumes of reporting.

Baseline

136. All available evidence suggests that the majority of in-scope platforms already allow users to report harm. All respondents to Ofcom's VSP consultation allowed users to report harmful content with mechanisms ranging from three-dot icons to flagging buttons near the content. Through interviews with a sample of in-scope platforms, 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 harm on their platforms. Many of these also tailored the options in their reporting functions to represent the categories of harm commonly reported on their sites, and to enable them to better triage reports to ensure they dealt with the high priority categories of harm first. 100% of respondents to DCMS stakeholder survey (out of 8 that answered the question) had reporting mechanisms in place and similarly, the AVMSD research found that most platforms allowed users to flag content for review.

Cost estimates

137. Based on all available evidence, this IA expects the vast majority of platforms to already allow user reporting. In line with Ofcom's findings in the context of the VSP regime, any costs are expected to be minimal, incremental, and relate to staff time[106] and ensuring reporting mechanisms remain fit for purpose, for example, simply repositioning of the organisations' email address for low risk platforms or minimally revising the triage functionality for higher risk platforms. This IA does not expect platforms to have to undergo significant redesign of online services to comply with the reporting requirement.

138. 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 (rising to 2, 4 and 6 for small, medium and large respectively).
- Mid risk platforms: 2 hours of programmer time for micro (rising to 4, 6 and 8 for small medium and large respectively)
- High risk platforms: 8 hours of programmer time for micro (rising to 12, 16 and 20 for small medium and large respectively)

139. In addition to programmer time, for each in-scope platform, one hour of Chief Executive/Senior Official time is estimated for sign-off of the changes.

140. Previous estimates for user reporting costs were not specifically challenged in response to the consultation IA and the approach remains broadly the same. However, this IA allows for the possibility that a small number of in-scope firms in the baseline may not currently allow user reporting in any form. Only one platform provided estimated costs of £1,000 per year for their user reporting function. This IA conservatively assumes that between 5% - 20% of in-scope firms across low and mid risk platforms[107] may have to develop a user reporting mechanism rather than just make incremental changes (10% in the central scenario). In the absence of evidence on cost differentials across risk and size categories, these platforms are expected to incur costs of £1,000. The above approach results in a total cost of implementing or revising user reporting mechanisms in the first year of between £3.3 million - £6.3

[105] The Electronic Commerce (EC Directive) Regulations 2002

[106] While there will be additional costs related to operating the user reporting system, this is considered as a compliance cost under additional content moderation.

[107] Evidence suggests coverage in high risk platforms is universal and costs are expected to be incremental only.

million (central, £5.2 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. Table 9 outlines the range of expected costs associated with ensuring users are able to report harm:

Table 9: Ensuring users are able to report harm (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Ensuring users are able to report harm	£17.7 million	£23.1 million	£33.8 million

Updating terms of service

Requirements

142. Under Option 1, all companies will be required to set terms of service for illegal content and, if relevant, protecting children. In addition, Category 1 organisations will be required to set terms of service in relation to legal but harmful content.[108]

Baseline

143. Available evidence from AVMSD research and platform engagement indicates that terms of service are already widespread under the baseline. AVMSD research found that the most commonly implemented user-safety measure was 'acceptable use policies' which large and medium sized platforms in the sample[109] considered to be fully functional at addressing critical risks. In addition all respondents to a survey of stakeholders already had terms of service (out of 8 that responded to the question). In addition, nearly all respondents to Ofcom's VSP call for evidence[110] had terms and conditions which prohibited the specific categories of harmful material under the VSP framework.

144. In addition, changes to terms of service is a business-as-usual activity undertaken by platforms. AVMSD research indicated that platforms regularly update these policies in response to their users. This was supported by respondents to Ofcom's VSP call for evidence with many platforms indicating that they regularly review and update their terms and conditions. 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 platforms are illustratively expected to incur some incremental costs associated with assessing their own terms of service and revising them to reflect the regulator's code of practice.

Cost estimates

145. Based on an assessment of 14 of the most popular online services' terms of service,[111] they range in length from 2,451 words to 15,260[112] with an average length of 5,976. It is estimated that 1.5 hours will be spent initially on reading, assessing, and making the changes. One member of staff (one senior official at a wage of £38.96 for SMBs and one regulatory professional at a wage of £20.62 for medium and large platforms) is expected to read and assess the current terms of service and make the necessary changes. In addition, we expect businesses to potentially require between 1-4 hours of legal advice[113] (2 hours in the central scenario). Finally, this IA estimates one hour of Chief Executive / Senior

[108] 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.

[109] 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.

[110] Consultation on guidance for VSP providers on measures to protect users from harmful material (Ofcom, 2021)

[111] These include some of the most popular services such as Facebook, Instagram, Twitter and TikTok.

[112] Visualizing the Length of the Fine Print, for 14 Popular Apps - visual capitalist (April 2020)

[113] Assumed to be given here by a legal professional at a wage of £39.23

Official time for sign-off of the changes is included in the estimates. To reflect the potential need for ongoing updates, this cost is expected to be incurred each year but reduced by 50% from the second year onwards.

146. The table below outlines the range of expected costs associated with updating terms of service:

Table 10: Updating terms of service (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Updating terms of service	£17.8 million	£23.1 million	£33.6 million

147. Previous estimates for updating terms of service were not specifically challenged in response to the consultation IA and the approach remains broadly the same. However, based on the qualitative evidence from engagement with in-scope platforms related to the need for legal advice, these estimates include legal advice for all platforms rather than just medium and large as estimated previously.

Compliance costs

148. For appraisal purposes, it is assumed that legislation enters into force in 2024 and platforms are expected to comply with the codes from 2025. This IA therefore assumes that compliance costs will begin from the second year of the appraisal period. The table below sets out the total compliance costs across the policy options. Details on how these costs have been estimated is below.

Table 11: Total compliance costs (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Compliance costs	£1,427.2 million	£2,132.2 million	£2,869.8 million

149. Platforms are expected to incur the following costs associated with compliance:

- **Conducting risk assessments** - this relates to the requirement to carry out an illegal content risk assessment and 'if likely to be accessed by children' to carry out a children's risk assessment. For Category 1 platforms, as part of this they will also have to assess the risk of legal but harmful content.
- **Undertaking additional content moderation** - the OSB does not require additional content moderation; however, it is likely that platforms will increase resources in this area to comply with the duties.
- **Employing age assurance technology** - in complying with the child safety duties, some higher risk platforms are likely to adopt age assurance (and specifically age verification) technologies.
- **Transparency reporting** - this relates to Category 1 platforms producing annual published reports on platform harm and related actions taken.
- **Fraudulent advertising duty** - as part of complying with the fraudulent advertising duty, it is likely that in-scope platforms will conduct CDD on advertisers.
- **User verification and empowerment duties** - this relates to the requirement on large social media platforms to offer optional user verification and provide user empowerment tools.
- **Assessing impacts on freedom of expression and privacy** - this relates to publishing an assessment of impacts on freedom of expression and privacy and keeping this updated.
- **Reporting online CSA to designated body** - this refers to the cost of reporting identified CSA content to the relevant designated body.

Conducting risk assessments

Requirements

150. All platforms in scope will be required to produce a risk assessment. Platforms will be expected to assess risks corresponding to the type of content and activity a platform is required to address. In practice, this means the vast majority will only be required to assess risks related to illegal content and activity and - if likely to be accessed by children - content and activity which is harmful to children. There is an additional requirement on Category 1 services to assess risks related to legal but harmful content and activity accessed by adults.

Baseline

151. From engagement with industry, under the baseline, many (especially higher risk platforms) already conduct internal risks assessments. Platforms use these risk assessments to prioritise user safety resources and to ensure emerging risks are identified. In addition, while not an explicit requirement of the VSP regime, Ofcom already strongly encourages platforms in its guidance[114] to assess the level of risk on the platform, and notes that this will form part of the European Commission's Proposal for a Regulation on a Single Market For Digital Services.

Cost estimates

152. Risk assessment cost information provided by platforms is limited. Only two provided the cost of producing a risk assessment but these both related to risk assessments they already produced and ranged from £2,500 to £10,000. Given that many organisations already produce these, this figure would overestimate the incremental cost. Based on qualitative evidence provided by platforms, the cost to business and effectiveness of risk assessments are likely to depend on:

- **Expectations on platforms**: that is the need to minimise the administrative burden on platforms required to assess risk across multiple duties.
- **Focus of risk assessment**: risk assessments need to consider the range of measures a platform has in place in relation to its specific risks. Some measures will be more important to some platforms than others, depending on the type of content they host and whether they are likely to be accessed by children.
- **Alignment with international and domestic regulations**: the need to ensure expectations on platforms align with current risk assessment practices which are conducted in compliance with other relevant regulations, for example AVMSD, Digital Services Act, and others.

153. Ofcom will set out the steps platforms can take to comply with the requirement to produce risk assessments in future codes of practice. Given that cost information is limited in the context of risk assessment, estimates presented in the previous IA are retained to provide an indication of the likely scale of impact at this stage. These are based on estimates from the Networks and Information Systems Regulations 2018 (NIS)[115] used as a proxy for the cost of producing an online harm risk assessment (or revising an existing one).

154. In order to estimate the expected incremental 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 £39.23) and 2 hours for a senior manager (at a wage of £20.92) for micro and small platforms, rising to 5 and 7 for medium sized platforms and 10 and 14 for large platforms respectively. In addition, in line with the possibility that some - while rare - may not currently assess risks on their platforms, this IA illustratively estimates that between 0%-5% of in-scope mid risk and high risk platforms (2.5% in the central scenario) may incur costs as large as those provided by platforms in the context of risk assessments they already produce. While we do not have comprehensive evidence of baseline risk assessment practices (given the scope of the regulation), all available evidence suggests some element of assessing risks on platforms is widespread across platforms. Based on the limited cost information available, this IA estimates these platforms could incur

[114] Video-sharing platform guidance (Ofcom, 2021)
[115] The Network and Information Systems Regulation 2018 - DCMS (April 2018)

costs of £6,250 (or the midpoint of the range provided above). As the cost of producing a risk assessment is likely to reduce once a platform has reported for the first time, the cost is incurred each year for all businesses but expected to reduce by 50% from the second year of compliance onwards.

155. Table 12 outlines the range of expected costs associated with assessing the risk of harm on the platform:

Table 12: Risk assessments

	Low	Central	High
Option 1: Risk assessments	£17.5 million	£33.1 million	£48.7 million

Undertaking additional content moderation

Requirements

156. The core duties of care require all in-scope platforms to put in place systems and processes to address illegal content and - if likely to be accessed by children - to protect children from harm, both illegal and legal but harmful. There is an additional duty on Category 1 organisations to address legal but harmful content likely to be accessed by adults through enforcing their terms of service. To protect freedom of expression and privacy, in fulfilling their safety duties, Category 1 platforms will have to put in place clear policies to protect journalistic content and content of democratic importance.

157. While platforms will fulfil their safety duties in many different ways, Option 1 is expected to result in some platforms requiring additional content moderation. This could be through hiring additional human content moderators, employing automated content moderation systems, or a combination of both. As with other aspects of Option 1, requirements on in-scope platforms will be proportionate and risk-based with the largest highest risk platforms expected to do more than the smallest lowest risk platforms.

Baseline

158. The vast majority of organisations in scope will already be taking some action to reduce the risk of online harm on their services. Many of the largest platforms already employ large teams of content moderators and operate sophisticated automated moderation systems - Facebook for example, employs over 15,000 human moderators in the US[116] and reportedly employed an additional 125 moderators in response to Germany's NetzDG.[117] Both RR's research and EY's assessment of VSPs demonstrated that resources spent on moderation activities in the baseline vary greatly from hundreds of pounds for the smallest lowest risk platforms to over £1 billion for the largest platforms.

159. RR's research found that some organisations consider it unlikely that the regulation will result in significant incremental costs. This is because of increasing user expectations over the safety of online communities, requirements set by advertisers and third-party suppliers, and to remain competitive in the industry. In support of this, EY's research on VSPs also found that most platforms in their study indicated that compliance with AVMSD (which includes many similar principles to the OSB) was not expected to result in incremental investment.

Cost estimates

160. Compliance costs related to potential additional content moderation will depend in full on the specific requirements set out in future codes of practice. At this primary stage, this IA is only able to provide an indication of the likely scale of impacts. Ofcom will consult on future codes and produce IAs once the specific requirements are set. The vast majority of platforms engaged are unable at this stage to provide estimated costs associated with potential additional content moderation until they know what they will be required to do. On this basis, given that previous estimates for content moderation were not challenged in response to the previous IA, the approach remains broadly the same and is considered a reasonable indication of scale of potential future costs.

[116] Facebook content moderators paid to work from home (BBC, 2020)
[117] NetzDG Transparency Report (Facebook, 2020)

161. RR interviewed a sample of in-scope platforms[118] to determine: their current practices and processes to mitigate the risks of online harm 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.

162. A strategic sample 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.

163. 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. Based on findings from the interviews, the percentage of in-scope platforms requiring extra spend on content moderation is conservatively estimated to be between 20%-30% of high risk in-scope organisations (25% in the central scenario) and between 5%-15% of medium and large mid-risk organisations (10% in the central scenario).

164. Since the previous IA, fraud facilitated through UGC is now in scope of the OSB and in-scope platforms will be required to address fraud as part of their safety duties. Evidence is limited on the extent to which baseline systems and processes (and those implemented in response to the online safety framework) are harm specific. Expected user safety investment such as human moderators, automated moderation, user reporting functions, and risk assessments are expected to enable in-scope platforms to address the full range of online harm and the removal of the exemption of user-generated fraud is unlikely to significantly affect costs for platforms already expected to incur costs associated with additional content moderation. In addition, the OSB does not provide a full list of categories of harm in-scope and therefore, current estimates provided by platforms are based on a broad understanding of the types of harm likely to be considered in-scope of the framework. However, for certain types of platforms, such as some large high-risk dating sites and ecommerce sites, the inclusion of fraud is likely to result in them incurring moderation costs, especially if they did not expect to prior to its inclusion. To reflect this within the estimates, under the high estimate, the percentage of platforms requiring additional content moderation is increased to 15% of mid-risk firms and 30% of high-risk firms. The estimates, including the central estimate, are highly uncertain given that it was not possible to interview a representative sample of in-scope platforms given the scope of the regulations.

165. 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[119] (the lowest estimate) to 15%[120] (the highest). These estimates were provided in the context of platforms' interpretation of the OHWP, that is, the cost of additional content moderation for platforms required to address all categories of harm 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 categories of harm). Turnover estimates used come from average turnover by business size band in BEIS' BPE. Sensitivity analysis is conducted on the full range of estimates provided by businesses.

166. For non-Category 1 platforms - those not required to address legal but harmful content accessed by adults - costs are expected to be lower than those incurred by Category 1 platforms. To calculate the cost to these, data from a number of large social media platforms' transparency reports on the volume of actioned content (content which was removed or minimised due to breaking the 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:

[118] Under the policy position as set out in the OHWP and not the subsequent exemptions.

[119] 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.

[120] 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.

167. Using the volume of actioned content in each category, an approximate percentage split of illegal vs legal but harmful actioned content was estimated. Four social media platforms' reports were assessed[121] and 2020 data was used. 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. For these categories, the volume of content actioned was split equally between illegal and harmful.[122]
- It is not clear that the four social media platforms' transparency reports are representative of the wider sample of in-scope businesses and are likely to be designated as Category 1 platforms. It may be the case that legal but harmful content represents a smaller proportion of overall harmful content on platforms not designated Category 1 or vice versa.

168. The percentage of actioned content in categories assessed as being likely illegal ranged from 14%-36%.[123] To reflect the costs to platforms not designated as Category 1 (those which are not required to address harmful content accessed by adults), given the ranges above, this IA estimates that the relative costs to these platforms would be approximately 25% of the relative costs to Category 1 platforms or 1.9% of turnover. Sensitivity analysis is conducted on the full range of estimates in the risks and sensitivity section.

169. Table 13 outlines the range of expected costs associated with additional content moderation:

Table 13: Additional content moderation (2019 prices, 2020 base year - 10 year PV)

	Low	Central	High
Option 1: Additional content moderation	£1,319.1 million	£1,902.6 million	£2,486.2 million

170. Three platforms provided cost information as part of DCMS' 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 the AVMSD research varied widely from hundreds of pounds for the smallest platforms to £1.5 billion for the largest VSP. With the exception of a handful of the largest and highest risk businesses, for those expected to undertake additional content moderation, the per platform costs under the central scenario above would represent a doubling or more of current content moderation costs which is likely to be significantly conservative and potentially an overestimate. While the range of estimates only reflect the percentage of platforms expected to incur costs, the per platform cost - in terms of percentage of revenue - is also tested in the sensitivity section.

[121] Facebook, Instagram, Twitter, and Snapchat

[122] Split categories include Facebook, Instagram, and Snapchat's hate speech category and Twitter's violence category which includes both threats of violence and glorification of violence.

[123] The previous IA used the same methodology on 2019 data which resulted in comparable findings with content categories assessed as likely illegal ranging from 15%-33%.

Employing age assurance technology

Requirements

171. While specific steps platforms can take to comply will be laid out in future codes of practice and regulator guidance, it is clear that some platforms will need to implement age assurance technologies. This may be as a result of complying with the core child safety duties or the pornography provision.

172. Age assurance refers to any method to establish the age of a user online. Age verification is one type of age assurance method, which provides the highest level of confidence in the age of a user. It commonly relies on officially provided data or hard identifiers, such as a credit card or passport. For this reason, it is best suited to 18 years+ services and content, rather than providing access for children who often do not have suitable documents. Other solutions that provide a lower level of confidence in the age of a user are referred to as 'age estimation'. These solutions are commonly AI based approaches that use biometric or behavioural data. This set of technologies is more nascent and the process of establishing the age or age-group of an under-18 years of age user is more complex. However, the technology has strong potential and is rapidly developing. Age estimation is more suitable to the under 18 year-old space as solutions do not rely on documentation, which many children do not have access to. Age assurance technologies are important tools that enable companies to take steps to protect children from online harm, including both legal but harmful and illegal content and activity, for example, protecting children from grooming.

173. The specific child safety duties in the OSB apply to platforms which are "likely to be accessed by children". This approach has been established by the legislation underpinning the Information Commissioner's Office's (ICO) 'Age Appropriate Design Code' (section 123 of the Data Protection Act 2018) with regards to protecting children's data. Consistency across regulations reduces additional burdens on businesses, many of whom will already have taken steps to comply with the Age Appropriate Design Code. Whilst the legislation is technology neutral, some high risk services which are likely to be accessed by children will be required to know the age of their users to provide them with appropriate protections, and therefore may choose to implement age assurance technologies[124] to do this. In addition, as part of the pornography provision, services that publish pornography will likely have to verify the age of their users to ensure that children are not able to access this type of content.

174. Without the pornography provision, platforms in scope of the core child safety duties would only be required to protect children from user-generated pornographic content. With the addition of the pornography provision, the intention is to minimise potential gaps in regulatory coverage and bring into scope children's access to non-user-generated pornographic content. This impact assessment focuses on the outcome, namely the implementation of age assurance technologies resulting from the OSB in its entirety, regardless of whether it is a result of the core duties or the published pornography provision. In its development of future codes of practice and regulator guidance, Ofcom will further consider the separate but related impact on businesses in scope of both the core child safety duties and published pornography provision.

175. Pornography is hosted on a range of platforms. Some of these platforms are user-user services and in scope of the core duties. Others will only be in scope of the OSB as a result of the pornography provision. While there is no definitive study, the BBFC estimates that there are between 4-5 million dedicated pornographic websites accessible in the UK. However, the number of businesses that this represents is much lower as companies often operate multiple sites. Further, the number of UK-based businesses that this represents is even lower, with the majority of sites operated by companies outside of the UK - the vast majority being in the US. Table 14 outlines the main types of platforms with the potential to currently host or publish pornography. It is important to note that not all platforms within each category will host or publish pornography and many will explicitly prohibit it as part of their terms of services. Table 14 should be viewed as a conservative upper bound estimate of platforms with the potential to host pornography (and therefore, the potential to implement age assurance technologies in response to the OSB):

[124] 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.

Table 14: UK based platforms with the potential to incur age assurance costs

Type of platform	Number of UK-based businesses 2021	
	Total	**Potential to incur age assurance costs**
Social media platforms: The vast majority, if not all pornographic content on social media platforms is user-generated and in scope of the core duties.	213[125]	213 This is conservative as only those that host pornography and/or are likely to be accessed by children have the potential to incur costs. Many prohibit this content as part of their terms of service.
Search engines: These platforms are in scope of the core duties.	1,366[126]	1,366 This is conservative as only those that host pornography and/or are likely to be accessed by children have the potential to incur costs.
VSPs: Subject to a combination of the core duties and the pornography provision depending on the the type of pornographic content (user-generated vs non-user generated)	18[127]	0 VSPs that host pornography are already required to prevent children from accessing sexually explicit content under AVMSD.
Dedicated pornography providers: Subject to a combination of the core duties and the pornography provision depending on the the type of pornographic content (user-generated vs non-user generated)	11	11 See 'Platforms in scope' for further details on estimating the number of UK-based pornography providers.
Image sharing platforms: The vast majority if not all pornographic content on image sharing platforms is user-generated and in scope of the core duties.	Unknown	Unknown There is no definitive data on the number of UK-based image sharing sites. It is reasonable to assume this number is low when considering only UK-based businesses.
VoD platforms: These are **not** in scope of the OSB, and pornographic content on these sites will continue to be regulated under the video on demand regime.	c.150	0 VoDs are not in scope of the OSB.
Total number of in-scope platforms that could potentially incur some amount of age assurance costs in 2024[128]		1,736

[125] Social Media Platforms in the UK - Market Research Report (IBIS World, 2021)
[126] Search Engines in the UK - Market Research Report (IBIS World, 2021)
[127] Notified video-sharing platforms (Ofcom, 2022)
[128] In line with the rest of this impact assessment, the number of potential platforms is uplifted by the average growth of UK businesses (3%) for an implementation date of 2024.

Baseline

176. In relation to user-user platforms in scope of the child safety duties, of those self-declared as likely to be accessed by children in a survey of stakeholders, three out of four that answered the question said that they already employed age verification. 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, such as self-declared age, which on its own is unlikely to be considered an appropriate child safety measure. While most platforms 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.

177. When it comes to dedicated pornography sites, evidence suggests that age assurance technologies (and in particular age verification) are rare and that children can easily access pornographic content on these sites. BBFC research in 2020 indicated that of the top 200 pornography sites (which together account for over three quarters of UK traffic to adult sites), only 4.5% have existing mechanisms in place that may prevent, deter or delay children accessing the site before displaying any pornographic content. Even these measures, which included having to sign up or register payment details, are not significantly robust given that children as young as eleven may have their own debit card - no sites in the top 200 required credit card only payments. 14.5% of the the top 200 sites have a pop up warning indicating that access is reserved for over-18s but this can easily be ignored by children that are intentional viewers of pornography. Based on BBFC engagement with the adult industry, the current lack of age assurance - even when the industry has stated its willingness to adopt these technologies - appears to be the result of competitive concerns and the potential commercial impact if this requirement is not mandatory across all services. It is therefore important that the child safety duties and pornography provision together apply to all pornographic content accessible to UK users.

Cost estimates

178. It is not possible at this stage to fully monetise the impact of the potential employment of age assurance solutions by some platforms in scope of the child safety duties and pornography provisions under Option 1. This is because:

- The platforms required to employ age assurance controls and the type of controls required will be set out in future codes of practice and regulator guidance (themselves subject to IAs).
- Types of age assurance solutions, their accuracy and their availability are rapidly evolving. The government and industry expect technology to greatly improve in this market and there are significant opportunities for cost reductions between now and implementation of the OSB.
- Different platforms are expected to take different approaches to meeting their duties under the OSB. For example, even within those likely to implement age assurance, some larger platforms, in particular the largest social media platforms, may develop in-house solutions while smaller platforms could employ off-the-shelf solutions which are cost effective and readily available. In addition, it may be the case that costs instead fall on the user. Evidence from the BBFC's engagement with industry suggests that the majority of pornography sites were expected to use certified third-party solutions to minimise the risk of privacy concerns. Some of the larger pornography platforms have founded their own solutions but these are run as separate businesses (and still considered third-party).
- There are also solutions offered to both companies and users at no price but may contain advertisements[129] as a means to create revenue for the age assurance provider or include a number of monthly free checks before paying a monthly subscription.[130]

179. While it is difficult at this stage to provide an accurate assessment of direct business costs, this IA presents a comprehensive indication of the likely scale based on two separate approaches, namely presenting individual platform costs based on an industry pricing survey conducted in January 2022 and top-down user-modelling scenarios.

[129] https://ageverify.com/

[130] https://www.1account.net/business-demo

180. To better understand individual platform costs, a selection of UK-facing providers of third-party age verification solutions were engaged through a survey distributed by the Age Verification Providers Association (the UK's industry body for age assurance providers). Illustrative costs were provided on the basis of a number of example platform scenarios. These costs should only be considered as providing an indication of the likely scale of costs and the actual price paid will be the result of standard business negotiations between platforms and third-party services. In reality, costs will depend on a number of factors and nuances not captured in the below example scenarios. Table 15 sets out the findings of this engagement with industry.

Table 15: Illustrative platform scenarios

Scenario	Description
Platform A	● 25,000 unique monthly UK users. ● 1% of users are assumed to be new each month and have not verified their age previously. ● 180,000 total monthly visits.
Platform B	● 100,000 unique monthly UK users. ● 1% of users are assumed to be new each month and have not verified their age previously. ● 730,000 total monthly visits.
Platform C	● 1 million unique monthly UK users. ● 1% of users are assumed to be new each month and have not verified their age previously. ● 7.3 million total monthly visits.
Platform D	● 4 million unique monthly UK users. ● 1% of users are assumed to be new each month and have not verified their age previously. ● 29.2 million total monthly visits.

181. The above platform scenarios are illustrative only and range from what would be considered a relatively small platform to a relatively large platform.

182. **Per check costs**: Costs per check ranged from less than 1p to more than £1. The large range reflects the variety of approaches and methods available to platforms. The only criteria given within the illustrative platform scenarios was that the approach should be able to determine whether a user is over 18 and meet standards defined by the British Standard Institute (PAS 1296:2018). Even within this criteria, AV providers offer an extensive range of approaches depending on regulatory requirements. While there is a large range, the majority of per check costs provided were 10p or lower per age check with the upper bound of the range reflecting a suite of different approaches. Some estimates did reduce as volumes increased with some providers per check cost lower for Platform D than for Platform A for example; however, others remained consistent throughout.

183. **Monthly costs**: AV providers were also asked to estimate the monthly costs for each illustrative platform based on per check costs or any other monthly pricing option. Information provided here was even more dependent on regulatory requirements and approaches taken by platforms. For example, costs depend on whether the platform would verify users each time they access the site or only new users. Some platforms provided monthly costs based on the per check costs outlined above, that is to say first month costs would include verifying the existing user base and from month two onwards, only new users are verified. Across a 12 month period, monthly costs provided for Platform A averaged just over £600, rising to just over £1,800 for Platform B. Monthly costs were estimated to be between £10,000 and £40,000 for Platform C and between £30,000 and £90,000 for Platform D.

184. There may be additional costs (not captured above) of integrating age verification solutions within each in-scope platform. Many third-party providers offer support packages to businesses with step by step instructions and developer support. As part of Yoti's submission to Ofcom's call for evidence on the VSP regime, it noted that it takes approximately half a day for a digital platform to integrate with Yoti's

backend system.[131] Assuming between one and three developers are required and based on median developer wages, this could result in additional platform costs of between £108 and £324.

185. Costs provided by AV providers should be treated with caution as:

- Technology in the age verification market is moving quickly and the industry expects significant improvements in accuracy and reductions in cost in the short to medium term.
- There are significant movements towards interoperability with solutions that can work across a number of platforms. While this is not an established approach yet it is something that the government is supporting through its work on standards, including the Digital Identity and Attributes Trust Framework, which will support interoperable solutions to function. As such it is possible that platforms would not need to establish the age of every user as many will have had their age verified previously.
- The platform scenarios presented to industry are by nature static and artificial. Actual costs will reflect the outcomes of standard business negotiations between platforms and third-party providers.
- The AVPA noted that some AV providers would likely offer heavily discounted fees for smaller clients and start-ups.

186. Given all of the uncertainties and limited data, it is not possible at this stage to monetise the direct cost to UK-based businesses. However, it is possible to demonstrate the totality of economic impacts by taking a top-down user-based approach.

187. Estimates from Ofcom and Revealing Reality for the proportion of adults[132] and children aged between 11-17 years old[133] that intentionally access pornography are applied to 2020 ONS population data.[134] Population is assumed to grow in line with average population growth between 2000 and 2020 (growth rate of 0.65% per year).[135] This modelling estimates that with an implementation date of 2024, there will be on average approximately 27.2 million unique adults and 1.6 million unique children intentionally accessing pornography each year across a ten-year appraisal period.

188. While there is no data on the average number of pornography sites visited by each unique user, evidence from VSPs more generally suggest that people tend to use a limited number of platforms to view videos.[136] On this basis, this impact assessment conservatively estimates that individuals accessing pornographic sites do so on average on five separate sites. Verification of age is assumed to last for 12 months before a user is asked to complete the process again. AV providers that were engaged as part of this impact assessment noted that this would be a decision for the platform and depend on the regulations at the time. Users may be provided with a 'token or credential' and only be verified once across the period whereas it is also possible to verify a user every time they access the site.

189. As noted earlier, there are movements in the age verification industry towards interoperability where, once verified on one site, a user would not need to be verified again even when accessing a different site. The possible levels of interoperability in the age verification market is represented in the low, mid and high cost estimates. The low estimate assumes complete interoperability (where verification is required only once), the mid estimate assumes moderate interoperability with each unique user undergoing verification twice across the five sites, and the high estimate assumes no interoperability.[137]

190. Based on per check costs provided by AV providers, a cost of 10p per check is used resulting in total costs of between **£17.9 million and £89.6 million (central estimate = £35.8 million)** in present value terms across the ten year appraisal period. It is important to note that this user-based modelling represents total costs to online platforms, including platforms based outside the UK and platforms

[131] Yoti response - Ofcom's call for evidence
[132] Online Nation 2021 report (Ofcom)
[133] Young People, Pornography and Age Verification (Revealing Reality, 2020)
[134] ONS Population Estimates (ONS, 2020)
[135] Population growth - United Kingdom (World Bank, 2020)
[136] Understanding how platforms with video sharing capabilities protect users from harmful content online (EY, 2021)
[137] Interoperability in the model is varied by decreasing the average number of sites visited by 5 in the high estimate to 2 in the mid estimate and 1 in the low estimate. This reflects the number of times a user requires verification across the five separate sites they use to access pornography.

operated by individuals as opposed to businesses. While it is not possible to estimate the direct cost to UK-based businesses only, it will be much lower than estimated here given the geographic distribution of pornography providers. For this reason, these estimates are included in the illustrative NPSV but are not included in the illustrative EANDCB. Further work will be done to refine business costs as part of Ofcom's development of future codes of practice and regulator guidance, including consultation with industry.

Table 16: Employing age assurance technology (2019 prices, 2020 base year - 10 year PV)

	Low	Central	High
Option 1: Age assurance	£17.9 million	£35.8 million	£89.6 million

Transparency reporting

Requirements

191. Option 1 requires platforms to produce annual transparency reports if they are designated as Category 1 (highest risk and highest reach user to user platforms), Category 2A (highest risk and highest reach search services) or Category 2B (high-risk, high-reach platforms but that may not necessarily meet the Category 1 threshold). Thresholds will be set out in secondary legislation and will be based on factors including a platform's number of users and its functionalities. While it is not clear how many platforms will be designated, based on policy intention, this IA estimates that between 30-40 platforms will be required to produce transparency reports.[138] 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 platforms who are required to report. The specific information that they will need to include, will be left to the regulator and will differ between platforms.

Baseline

192. Based on available baseline evidence, many large high-risk platforms already produce transparency reports. Three out of four large high-risk platforms that responded to DCMS' stakeholder survey already produced these, and it is clear from subsequent engagement that many do (through NetzDG requirements for example or just best practice). The vast majority of major social media companies already produce these, including granular data on harm, content removal, and content reinstated following challenges (see Facebook, Youtube, Instagram, Twitter and others).

Cost estimates

193. Estimates presented in the previous IA were not challenged with cost evidence supplied by platforms, and they still represent a reasonable estimate for the incremental cost of transparency reporting - that is the cost of potential revisions to existing reporting practises. However, qualitative evidence from recent engagement with in-scope platforms highlights some of the key cost drivers that will influence the scale of the regulatory burden, these are:

- **Alignment with international regulations**: the more that reporting requirements align with other international regulations (both current and planned) the less burdensome this will be for platforms.
- **Alignment with current reporting practises**: as above, the more these requirements align with current transparency reports produced by platforms the lower the cost.
- **Flexibility in terms of metrics presented**: one platform engaged noted that the cost of reporting is trivial compared to the cost of collecting data not currently collected. The right balance between flexibility for platforms and ensuring important metrics are presented (potentially in different ways by different platforms) is key to minimising costs.
- **Engagement between Ofcom and platforms**: year-on-year changes in key metrics presented in transparency reports could be due to external factors rather than solely changes in the level of

[138] For costs, the midpoint of the range is taken.

harm. It will be important for Ofcom to work closely with platforms to understand the information presented and external trends.

194. 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).[139] Estimates provided for NetzDG are a reasonable proxy for the transparency reporting requirements under the OSB. The cost of this activity is likely to be front-loaded, especially for platforms without appropriate systems already in place - to reflect this, the cost of transparency reports is expected to reduce by 50% from year 2 onwards.[140]

195. While the central estimate remains unchanged since the previous IA, Table 17 outlines some additional reporting costs gathered from other UK reporting requirements. Estimates below related to corporate governance reform and climate-related financial disclosures by publicly quoted companies form the low and high estimates as - outside of Germany's Network Enforcement Act which is the most analogous - they are most similar to reporting requirements under the online safety framework in terms of the focus on data and metrics.

Table 17: Comparison of reporting costs

Reporting requirement	Estimated costs per business
Germany's Network Enforcement Act (2017) *Requirement to report quarterly in German on their efforts to tackle illegal harm, including complaints and performance data*	£45,000 annual cost of reporting
Minimum implementation of the EU Non-Financial Reporting Directive for public interest entities with over 500 employees (2016) *Costs of reporting on anti-bribery and corruption matters.*	£951 first year costs with ongoing costs of £455
Mandating climate-related financial disclosures by publicly quoted companies, large private companies and Limited Liability Partnerships (2021) *Requires in-scope companies to report on metrics and targets used to assess and manage climate related risks and includes publishing as part of their annual report.*	£73,700 first year costs with ongoing costs of £56,800[141]
Climate Change Risk – Governance and Disclosure (TCFD) Requirements (2021) *Requirement on pension schemes in scope to publish a Task Force on Climate-related Financial Disclosures (TCFD) report.*	£3,750 first year costs with ongoing costs of £3,375
Corporate Governance Reform (2018)	£5,688 annually[142]

[139] Act improving law enforcement on social networks [Netzdurchführungsgesetz – NetzDG] - European Commission (2017)

[140] If the information required from platforms under the reporting requirements is changed frequently throughout the appraisal period, it is possible that costs could increase back to year 1 estimates.

[141] This includes both the metrics and targets aspect and signposting which are analogous to the kind of information required under the Option 1.

[142] This includes data collection, presentation and board discussion, and sign-off at the committee level.

Requirement in-scope companies to report on pay ratio.	
Payment Reporting Requirement (2016) *Requirement to report on payment information, including late payments*	£1,270 first year costs and £1,012 each year after[143]

196. Costs presented in the previous IA (proxied from NetzDG) remain the most reasonable and analogous. However, low and high estimates have been updated to include the full range of proxy costs. Table 18 outlines the range of expected costs associated with transparency reporting:

Table 18: Transparency reporting (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: Transparency reporting	£0.8 million	£6.3 million	£10.3 million

Fraudulent advertising duty

Requirements

197. Option 1 places an additional advertising duty on Category 1 and 2A platforms to implement systems and processes to minimise the risk that they publish and/or host fraudulent advertisements. While the exact steps businesses can take will be set out in future codes of practice (subject to consultation and impact assessments), this duty will result in these platforms being required to implement more comprehensive fraud prevention measures. In line with the rest of Option 1, the small number of platforms in scope of this duty (c.20) are likely to ensure compliance in a variety of ways depending on the risk of fraudulent advertising on their platform and any anti-fraud measures currently in place. Potential processes that these platforms could take include some form of increased customer due diligence (CDD), such as know your client (KYC) checks, credit checks, and sharing information on known fraudulent advertisers. They will also need to ensure that users can easily report fraudulent adverts and take appropriate action on receiving these reports.

Baseline

198. The digital advertising market is largely controlled by two platforms, namely Facebook and Google together accounting for 80% of all spending on search and display advertising. Based on desk research of large social media sites and search services, current baseline coverage of anti-fraud measures and advertiser verification is mixed. Some platforms do not verify advertisers and instead focus on advertisement curation and ensuring that they are not in breach of the sites' terms of service. Other platforms have very light touch signup requirements, such as verifying an advertiser's email address or website and potentially payment details. Many platforms operate optional verification for businesses wanting to advertise, where businesses are encouraged to undergo some form of due diligence to appeal to customers. Where platforms currently mandate advertiser verification, this is largely focussed on advertising related to social issues, elections and politics. Advertisers wanting to post content on these issues are required to provide valid identification and comply with a number of rules, including adding disclaimers to adverts and the sources of funding.

199. Facebook - the second largest player in the online advertising market with over 50% share of the display market - verifies political advertisers but has not announced plans to extend this to all

[143] These include reporting costs minus familiarisation costs

advertisers. Facebook along with Twitter and Microsoft recently announced that they would only host advertisements for financial products from companies that are authorised by the FCA.[144] This covers some of the types of measures that Ofcom will expect from in-scope platforms. These measures were likely introduced due to pressure from government and consumers and in anticipation of likely upcoming legislation. In 2019, Facebook also took a number of fraudulent advertisers to court for violating advertising policies and for defrauding individuals and tricking them into installing malware. In addition, Facebook - like the vast majority of social media sites and search services - allows users to report fraudulent adverts.[145]

200. In 2018, Google announced a new identity verification policy for political advertisers requiring them to provide government-issued identification and source of funds. In 2020, Google announced that it would extend this programme to all advertisers on its platform. Advertisers will need to submit to Google personal identification, business incorporation documents or other information that proves who they are and the country in which they operate. Additionally, in line with plans from other large platforms, Google verifies all UK advertisers that wish to post financial services related adverts of any kind and requires that they are authorised by the FCA. This is important as Google alone represents 90% of the search advertising market and is by far the single largest platform in the online advertising space.

201. It is not possible at this primary stage to discount platforms mentioned above from incurring potential business costs. While many of the current and planned measures are in line with actions businesses are likely to take to comply with Option 1's advertising duty, it is not clear how effective they are and companies will likely be required to go further by, for example, tackling broader categories of fraud beyond financial. Ofcom will ultimately consult platforms, assess current baseline measures, and determine the steps businesses can take to comply.

202. The Advertising Standards Authority (ASA) - the UK's independent advertising regulator[146] - has partnered with major online platforms to address fraudulent advertising. The ASA has introduced the Scam Ad Alert system which allows users to report fraudulent adverts. Once reported, the ASA works with online platforms to take fraudulent adverts down and to stop similar adverts appearing. In the first six months of the Scam Ad Alert system, the ASA received 1,274 reports resulting in 121 alerts being sent to online platforms. Given the lack of robust data, it is difficult to determine long term trends and therefore it is not possible to fully evaluate the current self-regulatory system. However, it is clear that fraudulent adverts are still widespread online and result in significant financial (and non-financial) loss to victims. Platforms are currently taking voluntary measures in this space but it is not clear how effective these are or whether further action is necessary. On this basis, the government has determined that a specific advertising duty on Category 1 and 2A platforms to ensure regulatory oversight of anti-fraud measures is necessary to mitigate wide scale economic losses.

Cost estimates

203. It is not clear at this primary stage what platforms will be required to do in response to the advertising duty. Option 1 sets out necessarily high-level duties on platforms and Ofcom will work with industry to assess the impact of measures it deems appropriate for compliance, including a full assessment of the impact on small and micro businesses (who themselves will not be in scope of the advertising duty but may be affected by it). At this stage, this impact assessment draws on a range of evidence sources to provide an indication of the likely scale of impact.

204. There is likely to be a range of potential measures that platforms could introduce to comply with this duty. For example, it could include verifying advertisers, credit checks, sharing information on known

[144] Tech giants agree to only publish ads of FCA-authorised firms (International advisor, 2021)
[145] But it is unclear how effectively they act on user reports.
[146] The ASA is an example of self regulation and co-regulation and is funded by industry.

bad advertisers or a range of other anti-fraud measures. Specific steps platforms can take will be set out in future codes of practice but it is plausible at this stage to assume that the advertising duty will result in a requirement on Category 1 and 2A platforms to conduct more stringent CDD on advertisers. Based on policy intention, approximately 20 platforms are expected to be designated as Category 1 or 2A and in scope of the advertising duty.

205. To calculate direct business costs, this impact assessment takes a top-down approach. Evidence from a representative survey of SMEs conducted by the Interactive Advertising Bureau (IAB) indicates that, on average, 60% of SMEs take part in paid-for advertising online through placement of advertising.[147] Broken down by business size, this is 52% of micro businesses, 81% of small businesses, and 96% of medium-sized businesses. The IAB's findings - while representative - only included registered micro businesses. It is reasonable to assume that the proportion of unregistered micro businesses - that is, businesses too small to be registered for VAT - is likely lower than those that are registered. However, in the absence of specific evidence on this section of the economy, estimates for registered micro businesses are applied to unregistered businesses - this represents a conservative approach. The IABs survey also did not include large businesses. However it is clear that the proportion of businesses increases with firm size and, therefore, it is estimated that 99% of large businesses participate in paid-for advertising online. The proportion of each size category is then applied to BEIS' BPE[148] and UK Civil Society Almanac data[149] to determine the total number of UK businesses (or the total number of businesses likely to undergo CDD as a result of participating in paid-for advertising online).[150]

206. The proportion of businesses which advertise within each size category is expected to increase across the appraisal period. However, IAB survey data used to estimate the percentage of advertising businesses is only available for a single year. To reflect this potential growth in advertising businesses, this impact assessment uses the average growth in the proportion of UK businesses with websites between 2007 and 2019 as a proxy (or +1.5% per year).[151] A number of other real-world growth rates from different but related areas were considered but rejected as potential proxies. For example, the proportion of businesses that use social media was rejected as ONS data for this specifically excludes businesses that use social media for paid-for advertising only.[152] Digital advertising spend was also considered but evidence suggests that a small number of the largest advertisers account for the vast majority of digital advertising spending and therefore, this would significantly overestimate growth in the number of businesses (given the long tail of small and micro businesses). Within each firm size band, growth in the proportion of businesses advertising online stops when it reaches 99%. This reflects a potential saturation point at which point all potential advertisers are already placing advertisements - both medium sized and large businesses reach the saturation point within the time-period. The proportion of micro businesses advertising online grows from 52% to 59% across the period (an increase of 1.5 million businesses) and the proportion of small businesses grows from 81% to 93% (an increase of 0.1 million businesses).

207. By the first year of the appraisal period, it is estimated that approximately 3.4 million UK businesses will advertise online, this figure grows to 5.0 million by year ten. It should be noted that this approach is conservative, as some of these businesses may participate in paid-for advertising on platforms outside the scope of the advertising duty only. However, given the high levels of market

147 Powering Up: Helping UK SMEs unlock the value of digital advertising (IAB, 2020)
148 Business population estimates 2021 (BEIS, 2021)
149 UK Civil Society Almanac (NCVO, 2021)
150 In line with the rest of this IA, the number of businesses grows in line with annual business growth across the period.
151 This impact assessment conducts sensitivity analysis on a range of growth rates from 0% (no growth) to 6.4% annual growth (the largest annual increase in the proportion of businesses with websites which occurred between 2007 and 2008).
152 E-commerce and ICT activity (ONS, 2018)

concentration in this space, it is reasonable to assume that many advertise on or through a platform likely to be designated as Category 1 or 2A - Google and Facebook alone have 1.2 million UK advertisers on their platforms.[153] This approach does not account for advertisers based outside the UK that target adverts towards UK users. While any costs on those advertisers would not normally be considered in an IA, the cost on UK-based Category 1 and 2A platforms of conducting CDD would be in scope. There is no existing evidence or data on how many non-UK based businesses advertise to UK consumers using Category 1 and 2A platforms and, therefore, it has not been possible to monetise these potential costs at this stage. In future codes, Ofcom will consider the full range of impacts through comprehensive consultation with affected platforms, including the cost of anti-fraud measures as they apply to non-UK based advertisers which target UK consumers.

208. As IAB estimates are based on active advertisers (having advertised in the last 12 months), this impact assumes that 100% undergo CDD in the first year. From year two onwards, only new advertisers undergo CDD checks.[154] Across the appraisal period, there may be additional due diligence required on already authorised advertisers resulting from, for example, business changes or updates to identity documents. Given the uncertainty around specific requirements, it is not possible to reflect this possibility with any reasonable accuracy at this stage. In addition, the steps platforms will take will depend on the risk of fraud on their platform and the changing fraud landscape.

209. Of course, some businesses advertise on multiple channels and will be required to undergo CDD on more than one platform. Based on evidence from the IAB, on average, the number of channels used across SMEs overall is 1.2, 2.4, and 3.7 for micro, small and medium sized advertisers respectively. Large businesses are much more likely to advertise across a range of channels, for example by advertising on some combination of the large social media companies and Google. In the absence of specific evidence related to large businesses, this impact assessment assumes that these businesses advertise on average across 5 different in scope platforms. The number of advertising businesses within each size category is then uplifted by the average number of channels for the respective size category.

210. There are four main costs modelled using the above approach:

- **Set up costs**: the cost of updating systems and processes to account for new requirements related to CDD
- **CDD costs (platforms)**: the cost of conducting a CDD on an advertiser
- **Staff time (advertisers)**: the cost to advertisers of completing any forms associated with CDD requirements and providing appropriate information
- **Staff time (advertising agencies)**: the cost to advertising agencies of facilitating CDD between platforms and advertisers

211. Set up costs are proxied from the impact assessment supporting the Transposition of the EU Fifth Anti-Money Laundering Directive which - in the context of the cryptoasset market - estimated set up costs for each firm of between £109,000 and £438,000 (central estimate = £274,000).[155] [156] The Money Laundering Regulations required a variety of different customer due diligence activities and are a reasonable but conservative proxy for unit costs in Options 1's advertising duty. Set up costs in the context of crypto providers was based on firms without current anti-money laundering frameworks in place. Many Category 1 and 2A platforms already have anti-fraud measures in place and, therefore, proxied set up costs are expected to be an overestimate. These costs are incurred in the first year only and cover updating systems and processes. Based on baseline evidence that some large platforms

[153] Online platforms and digital advertising - Market study final report (CMA, 2020)
[154] New advertisers incorporate both the growth in the proportion of businesses that advertise online and the growth in the businesses population.
[155] All figures have been uplifted from 2017 prices to 2019 prices in the model.
[156] Transposition of the Fifth Anti-Money Laundering Directive (HMT, 2019)

already conduct similar kinds of anti-fraud due diligence and advertiser verification, this figure is likely conservative.

212. The unit costs of conducting CDD are also proxied from HMT's Money Laundering Regulations. Standard CDD is estimated to cost between £3 and £15 (central estimate = £9) and enhanced CDD is estimated to cost between £4.50 and £30 (central estimate = £19). While the vast majority of CDD resulting from the advertising duty is expected to be automated (at least to some extent), the inclusion of estimates for enhanced CDD allows for the possibility that a small number of cases require additional scrutiny, such as for advertisers operating in industries known for high levels of fraud. This impact assessment conservatively estimates that 5% of advertisers will require enhanced CDD resulting in greater costs for in scope platforms. Enhanced CDD was expected to be conducted on only 0.23% of customers in the context of anti-money laundering. However, under the OSB platforms may decide to take a more risk averse approach with more stringent checks on risky industries or types of businesses (as opposed to individual customers as is the case for anti-money laundering).[157]

213. In addition to the cost of Category 1 and 2A platforms conducting CDD, advertisers themselves will also incur costs associated with completing necessary forms and providing appropriate information to in-scope platforms. This impact assessment estimates that this will take between 10 and 30 minutes (central estimate = 20 minutes) for standard CDD and between 30 and 60 minutes (central estimate = 45 minutes) for enhanced CDD. There is limited evidence on the time taken for an advertiser to complete a process like this, but it is in line with estimates for the time taken to open a bank account in the UK (itself subject to anti-money laundering checks).[158] This impact assessment assumes that this process will be conducted by a Chief Executive in small and micro businesses and by a marketing associate in medium and large businesses. Finally, a proportion of advertising businesses will use advertising agencies who may incur costs as a result of facilitating the CDD process. To account for this, 25% of CDD checks in the model include additional staff time of between 10 and 30 minutes (central estimate = 20 minutes) for advertising agencies. This is based on evidence presented to the CMA that a quarter of advertising revenue is channelled through media agencies. Given that the majority of revenue comes from a small number of large advertisers, the actual proportion of advertisers using agencies is likely much lower and 25% is a conservative estimate.

214. Applying the methodology above, this impact assessment estimates that the fraudulent advertising duty will result in costs of **between £64.6 million and £226.7 million (central estimate = £145.8 million)** across the ten-year appraisal period.

Table 19: Fraudulent advertising duty (2019 prices, 2020 base year - 10 year PV)

	Low	Central	High
Option 1: Advertiser due diligence	£64.6 million	£145.8 million	£226.7 million

Indirect costs of fraudulent advertising duty

215. The extent to which Option 1's fraudulent advertising duty results in indirect impacts is dependent on a number of factors, all of which are at this stage unknown. While Ofcom will consider these further through consultation with industry and subsequent impacts assessments, Table 20 provides a qualitative assessment of the measures potential effect on supply, demand and price in the market:

[157] Taking HMT's estimate of 0.23% instead of 10% reduces the cost of the fraudulent advertising duty by 9.2% with <0.1% change to total policy costs.

[158] How to open a bank account online (Which?, 2021)

Table 20: Fraudulent advertising duty potential indirect impacts

Supply	The effect of Option 1 on the supply of advertising space is uncertain. Firstly, it will likely result in a short term increase in the supply of advertising space for non-fraudulent advertisers due to a reduction in fraudulent advertisers being able to advertise on platforms freeing up space for legitimate advertisers. However, the removal of fraudulent advertising would be beneficial to advertisers that don't want their products or services to appear next to harmful content or scam adverts, likely offsetting any potential increases in supply.
Demand	Demand side changes are complex and are expected to be influenced by the ability of advertisers to comply with additional checks and additional costs they may incur, as a result of completing forms and providing relevant information. If checks are too burdensome, or they can only be met by a subset of current advertisers, a fall in the demand for online advertising might be expected. However, online advertising has been performing strongly with rapid growth due to the ability to reach large audiences, the ability to engage users and drive direct sales, and the ability to target relevant audiences. These attributes mean it's unlikely further checks will result in any decrease in demand from established advertisers and agencies. Further, Category 1 and 2A platforms have a strong incentive to ensure their CDD processes are easy and user-friendly, given their reliance on advertising revenue. Finally, most of the largest companies in this space are already implementing anti-fraud measures and, therefore, it is reasonable to assume that they do not see a trade-off between checks and advertising demand. Demand for advertising could increase if advertisers - currently hesitant to advertise on social media due to harmful content and scam adverts - decide to purchase advertising space. Anti-fraud measures could also positively impact on consumer confidence which could lead to increased purchasing and increased demand for advertising space.
Price	Category 1 and 2A platforms may decide to pass costs on to advertisers who may ultimately pass costs on to consumers. There is no indication how, for example, Facebook and Google would adjust their pricing, whether a one-off joining fee, or a change in the fees charged for services for each advert purchased. This could be an increase in the cost per impression or cost per click, or a reduction in the revenue share a publisher receives. The ability of intermediaries to pass on costs to advertisers or publishers will depend on the level of market power they have. As mentioned, advertising platforms are highly concentrated in search and social display advertising. While advertisers can still go through other routes to reach audiences, they cannot access the majority of internet users that access search and social media services. Given the scale of digital advertising spend and the relatively modest estimated cost of implementing anti-fraud measures, the extent of price increases is expected to be minimal and would be considered pass-through.[159]

216. The fraudulent advertising duty is proportionate and only applies to Category 1 and 2A platforms. While this is necessary to minimise business burdens, it does create a potential risk of fraudulent advertising being displaced to smaller less well-equipped platforms. Digital advertising is highly concentrated because platforms like Facebook and Google offer large and engaged user bases. While it is possible that some fraudulent advertisers may move to smaller platforms, given the advertising market share of large social media companies and search services, Option 1 is likely to capture a large proportion of advertising activity. If a fraudulent advertiser was to move to a smaller online platform (outside of Category 1 and 2A), it could not hope to attract the same number of advert impressions and, therefore, there would be less chance of users falling victim to the scam. Ofcom will further consider potential indirect impacts and risks associated with the fraudulent advertising duty. This will include consultation with affected businesses and subsequent impact assessments.

[159] RPC case histories - direct and indirect impacts, March 2019 (RPC, 2019)

User verification and empowerment duties

217. Option 1 introduces a duty on Category 1 platforms to offer optional user verification and provide user empowerment tools. In terms of optional user verification, Category 1 services would be required to put in place a mechanism by which an adult user could verify their identity, should they wish to do so, and would have discretion on which verification measure they offer to users. The duty to provide optional user verification applies to adult users only and is separate from potential age assurance requirements under the child safety duties and pornography provision. Acceptable forms of verification will be set by the independent regulator; however, they are likely to range from verifying an email address to official ID such as a passport or driver's licence. Option 1 does not mandate that users verify their identity with the platforms just that the option is available to them.

218. In addition, Option 1 places a duty on Category 1 platforms to provide user empowerment tools so that users can have more control over their online experience. Category 1 platforms have discretion over the legal content they permit on their services. However, for the legal but harmful content identified as part of the risk assessment or designated as priority harmful content which a particular service allows on its platform, it would have to offer adult users tools to enable them not to interact with that harmful content if they did not wish to do so. The steps services can take will be set by Ofcom. It is likely that the services will use tools to filter harmful content and allow users to determine who they interact with online.

Baseline

219. Many of the largest social media platforms offer user verification already but this is largely focussed on 'notable' users. For example, on Twitter, verified users "must represent or otherwise be associated with a prominently recognized individual or brand".[160] Twitter allows users to verify their identity through provision of an official website, ID verification (examples given are driver's licence or passport) or an official email address. Instagram takes a similar approach in requiring accounts to be notable before allowing for verification and give government issued ID or business documents as evidence for verification.[161] The approach of verifying notable users is fairly consistent across the main social media platforms. However, optional user verification based on identity rather than notability doesn't seem to occur on any under the *status quo*.

220. In terms of user empowerment tools, a review of the large social media platforms found that existing tools can broadly be broken down into the following categories:

Table 21: Examples of current tools available on Category 1 platforms

Form of user empowerment tool	Description
Chat functionality controls	Platform tools that give users the ability to block or restrict who is able to chat or direct message them.
Content controls from/to specific accounts/individuals	Controls that allow a user to either block a user, preventing them from seeing and interacting with a profile; mute a user, preventing content from that user from being seen; or in some cases, choosing what types of content other specified users can see from a given account.
Inappropriate content controls	This can vary significantly from platform to platform but covers: - Options to hide comments on posts, stories, and live videos deemed by platforms to be inappropriate or offensive. - Options to filter out profanity

[160] About Verified Accounts (Twitter)

[161] How do I request a verified badge for my Instagram profile? (Instagram)

	- Options to filter out particular content containing words/phrases the user does not wish to see.
In-app purchase controls	Controls enabling users to make decisions around if and how much spending can occur on a given platform by the user/user's child.
Parental controls	Platform settings that enable parents to control the types of content or level of content their children see.
Privacy controls	Tools to provide users with control over how they can be found, the level of visible personal information, and information around location.

221. All major social media platforms reviewed as part of this IA have some level of user empowerment tools in place already, though this tends to vary quite considerably from platform to platform, and also by platform type and functionality. In terms of current coverage, the most commonplace tools available to users are tools to filter out inappropriate content, tools to enable user blocking/muting, and privacy controls. These were available across almost all large social media platforms assessed.

Cost estimates

222. In-scope platforms are likely to take a variety of approaches to optional user verification from verifying email addresses to official forms of identification. Ofcom will set out appropriate measures that Category 1 platforms can take to comply with this part of the duty. At the primary stage, this impact assessment takes a user-based approach to model potential impacts and to provide an indication of the likely scale of costs.

223. The proportion of each adult age group that uses social media sites is taken from Ofcom's most recent adult and child media use and attitudes surveys.[162] [163] The percentage of individuals within each age group ranges from 45% (for 65 year olds and older) up to 90% (for 35-44 year olds). Across the ten year appraisal period, the proportion of each age group that uses social media is assumed to increase in line with average social media use growth rates within each age group between 2015 and 2021.[164] Within each group, growth in the proportion of adults using social media stops when it reaches 95%. This reflects a potential saturation point at which point all potential social media users within each age group are using social media. Given the relatively low growth rates in age groups with over 90% already using social media, a 95% saturation point is realistic. Only one age group - namely the over 65s - does not reach the saturation point in the ten year time period, increasing from 45% to 78%.[165] Social media use estimates are then applied to ONS population data[166] to obtain the total number of people in the UK that use social media in each year of the appraisal period. In line with the rest of this IA, population is estimated to grow in line with average growth rates between 2000 and 2020 (0.65% per year).

224. As user verification is optional under Option 1, it is not clear how many social media users will decide to be verified. To estimate this, the model takes account of relevant polling data related to anonymity online and identity verification on social media sites. Based on five recent polls, it is estimated that between 50% and 78% (central estimate = 68% poll average) have either a negative view towards anonymity on social media or a positive view towards identity verification on social media. For example,

[162] Adults' media use and attitudes report 2020/21 (Ofcom)
[163] Children and parents: media use and attitudes report 2020/21 (Ofcom)
[164] Ofcom adults media use and attitudes report 2015 - 2021 (Ofcom)
[165] Varying the saturation point has minimal effects on total costs. For example, a saturation point of 99% results in a 3% increase in total verification costs compared to 95% saturation point.
[166] ONS population data - source

in one poll, 50% opposed being able to create an anonymous account[167] and in another, 78% thought that users should have to verify when signing up and/or display their real name at all times.[168] It is possible that, when faced with a specific polling question, individuals are more likely to say they support verification or oppose banning when in reality they would not opt for optional verification themselves. However, this represents a very conservative upper bound estimate of the proportion of UK users who may make use of optional verification measures. The proportion of users willing to be verified will also depend on the type of verification required and the information they are willing to give to the platform. Polling data is then applied to estimates for the number of social media users within each group. Finally, to determine the total number of potential verifications, estimates for social media users likely to opt for verification are uplifted by the average number of social media sites used (6.7).[169] The total number of potential verifications is estimated to be between 139 million and 216 million (central estimate = 188 million).

225. Verifying willing social media users is likely to be spread across a number of years as users become aware of the option and decide to be verified. This impact assessment assumes that willing users are verified equally across the first five years. From year 6 onwards, only new willing users are verified each year.[170] [171]The unit cost of verifying a user is highly dependent on the method chosen by the platform. While specific methods will be set out by Ofcom and determined by the individual platform, this impact assessment uses the lowest cost provided by third-party age verification providers in the context of verifying a users age (£0.07p per verification). While verifying identity and verifying age are related they do represent separate processes. This cost is considered a reasonable proxy to indicate the likely scale of impact. Age verification checks generally rely on official ID and therefore, this is likely an overestimate, especially for platforms that opt to verify email addresses only. Applying the methodology described above, total costs of offering optional user verification to adults are estimated to be between **£8.8 million and £13.7 million (central estimate = £11.9 million)** across the appraisal period. As platforms already verify a proportion of users under the *status quo* and social media sites in general collect a large amount of user data already, Category 1 platforms are not expected to incur significant costs associated with changing systems or extending user databases. However, estimates will be further refined by Ofcom as the regulator determines specific requirements on platforms.

Table 22: Optional user verification (2019 prices, 2020 base year - 10 year PV)

	Low	Central	High
Option 1: Optional user verification	£8.8 million	£11.9 million	£13.7 million

226. It has not been possible at this stage to monetise the potential impact associated with user empowerment tools, such as giving users the ability to filter harmful content. Firstly, thresholds for Category 1 platforms will be set out in secondary legislation. With such a high degree of variability in current coverage amongst large social media sites (potential Category 1), it is not possible to estimate with any reasonable accuracy what platforms are likely to do to comply. Further, any potential incremental costs are likely to relate to platform design. These types of changes are very difficult to monetise and evidence of historical costs is limited (or non-existent) given the sensitive nature of costs. Ofcom will work with platforms to determine specific requirements and assess the impacts of any potential platform changes.

[167] Source Left Foot Forward

[168] YouGov 2021

[169] Global Social Media Stats (Datareportal, 2021)

[170] New users reflect both population growth and any growth in social media use.

[171] In an extreme scenario where 100% of willing users are verified in the first year, total verification costs increase by 14% with less than 0.1% effect on total policy costs. It is far more likely that verifications occur over a number of years.

Assessing platform impacts on freedom of expression and privacy

Requirements

227. The largest and highest risk platforms (Category 1 services) will have additional duties to assess the impact of their safety policies and procedures on freedom of expression (FoE) and privacy and demonstrate they have taken steps to mitigate this. They will need to publish this impact assessment and keep it updated (referred to in this section as a FoE and privacy IA).

Baseline

228. The Government is not aware of any platforms currently in compliance with this requirement and considers the full cost of producing an assessment to be incremental.

Cost estimates

229. Realised costs are dependent on requirements set out in future codes of practice; however, to provide an indication of the likely scale of impact, estimates from impact assessment requirements under General Data Protection Regulations (GDPR) are used as a proxy. Under GDPR, businesses are expected to produce Data Protection Impact Assessments (DPIAs) for any processing that is likely to result in a high risk to individuals. Businesses are also encouraged as good practice to produce a DPIA for any other major project which requires the processing of personal data. A DPIA must: describe the nature, scope, context and purposes of the processing; assess necessity, proportionality and compliance measures; identify and assess risks to individuals; and identify any additional measures to mitigate those risks. The cost of producing a DPIA is considered to be a reasonable proxy for the cost of producing an FoE and privacy IA under the OSB. Costs will be considered further through Ofcom's consultations with industry and future impact assessments.

230. The unit cost of producing a DPIA comes from the Ministry of Justice's Proposal for an EU data protection regulation - Impact Assessment.[172] [173] In 2019 prices, this equates to £12,692 for a small-scale DPIA, £31,277 for a medium-scale DPIA, and £135,082 for a large-scale DPIA. The estimate for large-scale DPIAs was considered in the Ministry of Justice's IA as an extreme example of a large project involving sensitive data and was not used in its calculations. Given the uncertainties on requirements related to FoE and privacy IAs, calculations presented here use the full range of potential costs to form the low, central and high estimate. Given that Category 1 platforms will be required to ensure this assessment is updated, this IA assumes that this cost is incurred each year but reduces by 50% from year 2 onwards. Table 23 outlines the range of expected costs associated with FoE and privacy IAs:

Table 23: FoE and privacy IA (2019 prices, 2020 base year, 10-year PV)

	Low	Central	High
Option 1: FoE and privacy IA	£1.1 million	£2.7 million	£11.5 million

Reporting online CSA to designated body

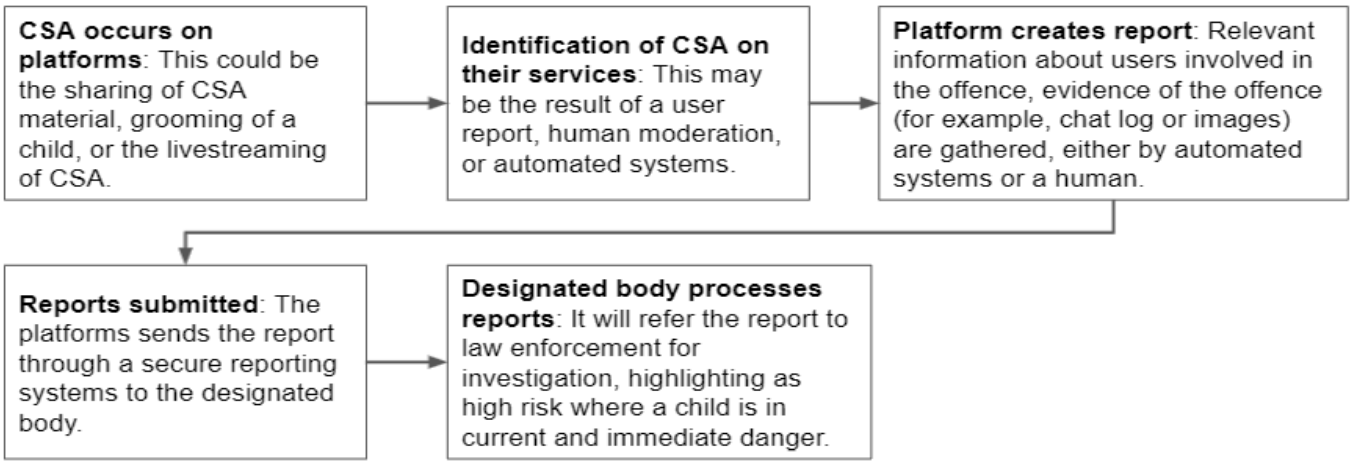
231. Option 1 introduces a legal requirement on technology businesses to report online CSA. This requirement will apply differently to platforms depending on where they are based, which is different from the approach being taken to the Online Safety regime generally, where duties will apply to all in-scope services that have UK users. UK platforms (those that provide services from within the UK) will be required to report identified CSA content to [placeholder: expected to be the NCA but not yet confirmed]. Platforms providing services from outside of the UK will only have to report identified CSA offences that are perpetrated by a UK user, and only if they do not already report CSA. 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 platforms do not have to replicate their reporting

[172] Proposal for an EU data protection regulation - Impact Assessment (MoJ, 2012)

[173] To note, Ministry of Justice estimates are themselves taken from the EU commission's own estimates.

efforts. For UK platforms, this will replace the current voluntary reporting regime within the UK. The below figure outlines the process:

Figure X: Mandatory reporting process for CSA



232. 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 platforms 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 platforms 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 CSA. From their consumer email and cloud storage services, they anticipate identifying less than 100 instances of CSA 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 platforms 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 CSA (grooming) to UK law enforcement.

233. It is estimated that about 25,000 platforms 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 CSA. To avoid duplicate reports being made, these platforms will not be required to report again under the UK reporting regime. 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 25,000 platforms will be required to report under this new requirement, and how many reports these companies are likely to make.

234. Some countries, including the USA and Canada, have legal requirements on platforms to report online CSA. Platforms based in the USA are required by law to report to 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 NCA.

235. The overall cost to the UK's technology industry on reporting CSA offences is minimal and to some extent controllable by the organisation (e.g. whether they use automated reporting). The cost for identifying CSA is already part of the platforms' costs within their identification and moderation process. These processes will vary, with some proactively identifying CSA using automation while others relying on user reports and human moderation.

236. 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 this provides an estimated cost per report of £2.10 - £4.20.

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

237. NCMEC's international reports from 2019 by electronic service providers[174] demonstrate that smaller technology platforms report less than the big technology platforms. Most large technology platforms, particularly social media sites where CSA 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 CSA; however, Table 24 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 CSA to share with the designated body.

Industry fees

238. Ofcom's operating costs will be paid through an annual industry fee. The annual industry fee will be tiered, and Ofcom will set a threshold (based on qualifying worldwide revenue) at or above which a platform would be required to notify and pay an annual fee. Platforms below the threshold will not pay the annual fee, though will still be subject to the regulatory regime. An appropriate threshold will ensure all small and medium enterprises are exempt from the direct costs of paying a fee. 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.

239. 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 £34.9 million per year and total £313.9 million across the appraisal period (2019 prices, 2020 present value base year). 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 EANDCB (although it is included in the illustrative net present social value (NPSV)).

Enforcement

240. Ofcom will have a suite of enforcement powers to take action against platforms that fail to meet their regulatory responsibilities. These are: issuing confirmation decisions, imposing fines, requiring companies to make improvements and - in the most serious cases - business disruption measures (including blocking via the Courts). Such enforcement powers will apply across different types of platforms, e.g. size, revenue, activity, overseas; and be used proportionately to potential or actual damage caused, and size and revenue. Ofcom will be required to consult and produce guidance setting out how it will use its enforcement powers. Table 25 sets out details of these enforcement measures:

Table 25: Regulatory enforcement powers

Confirmation decisions	**Description**: Where Ofcom identifies a breach, it can issue a confirmation decision confirming the breach. These can set out the steps required by the company to come into compliance with their duties and the financial penalty being imposed (if any).

[174] 2019 Reports by Electronic Service Providers - NCMEC

	Costs to platforms: As is standard practice in regulatory appraisal, this IA assumes full compliance with the regime. Therefore, any costs to platforms from rectifying actions undertaken as a result of receiving a confirmation decision is already captured in this IA's assessment of compliance costs. Confirmation decisions are therefore unlikely to result in material costs.
Fines	**Description**: 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 their duties. The approach to enforcement will aim to encourage compliance and drive positive cultural change. The regulator will support platforms to help them understand the expectations placed on them, and how the regulator's use of its enforcement powers will be proportional. Civil fines can be issued of up to £18 million or 10% of qualifying annual global turnover, whichever is higher.
	Costs to platforms: 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. In 2020/21, the ICO issued three fines for contraventions of GDPR totalling £39.7 million. In addition, for non-compliance related to nuisance calls, the ICO issued 35 fines totalling £2.3 million.[175] Fines issued by Ofcom under the online safety framework are expected to be rare but may be large if issued to large social media companies for example.
Business disruption measures	**Description**: In the most serious instances of non-compliance, Ofcom will have the power to initiate business disruption measures, to be used as a last resort. These include requiring third parties to withdraw key ancillary services (like payment or advertising services) to make it less commercially viable for non-compliant businesses (service restriction orders) and in some cases, restricting access to a non-compliant platform's service (access restriction orders).
	Costs to business: The frequency with which these measures are used depend on future codes of practice, the level of compliance, and the effectiveness of preceding regulator action on in-scope organisations, e.g. confirmation decisions and fines - all of which are unknown at this stage. To provide an indication of the likely scale of potential impacts, estimates from the IA for 'Age verification for pornographic material online'[176] which similarly involved notifying payment service providers and internet infrastructure providers, 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 **£0.6 million** per year.[177] In the context of the OSB, the two largest payment service providers engaged suggested that the cost of a small number of business disruption measures (as expected under Option 1) would be negligible to zero and would be absorbed into existing processes for responding to regulatory requests.
Deferred power to introduce senior management	**Description**: The Government will have the power to introduce criminal sanctions for senior managers who fail to ensure their company properly complies with Ofcom's information requests. These sanctions will be deferred for at least two

[175] Information Commissioner's Annual Report and Financial Statement (ICO, 2021)
[176] Age Verification for Pornographic Material Online, Impact Assessment - DCMS (2018)
[177] Converted from 2016 price and present value base year to 2019 prices and 2020 present value base year.

liability	years and will only be introduced following a post-legislative review into the effectiveness of Ofcom's enforcement and information powers.
	Cost to platforms: As above, this IA assumes full compliance with the regulatory framework, including full compliance with information requests. These sanctions are therefore unlikely to impact on costs, and will also not be effective until at least 2026-27 (if implemented).

241. While impacts associated with regulator enforcement action is not monetised at this stage, the above information provides an indication of the likely scale of impact. In addition, given the disincentivizing effect of, for example, large fines and damage to reputation, the Government expects enforcement action which results in platform fines or business disruption measures to be rare.

Cost to individuals

242. The new regulatory framework will apply to companies or individuals who provide services which host UGC or enable P2P interaction, as well as search engines. 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.

243. Given the low risk functionality exemption, the consultation stage IA noted that the vast majority (if not all) individuals are likely to be out of scope and that the Government did not have any evidence of individuals who could be in scope of the regulation. While the current scope (especially the low risk functionality exemption) is likely to have removed the vast majority of individuals, evidence provided in response to the consultation stage IA did highlight that at least some individuals will fall in scope. One individual noted that they would shut down their service as a result of potential compliance costs. The main concerns highlighted in evidence submitted include:

- Not being able to engage legal services to both consider whether the platform is in scope and compliant and to ensure continued compliance (especially as the platform is already run at personal expense); and
- The financial risk posed by Ofcom's power to fine platforms for failing to comply. It was also noted that the OSB could potentially result in trolls purposefully flooding the platform with illegal content to overwhelm current moderation systems.

244. Given the significant risks around creating a regulatory loophole, and the suspected rarity of individual cases, the Government does not consider it proportionate to exempt non-businesses from scope. However, Option 1 does include a number of provisions to ensure impacts on individuals are minimised and to avoid individuals shuttering online services, these include:

- Regulatory expectations on services will be reasonable and proportionate to the severity of the potential harm posed and the resources available to the service. If the risk of harm on a platform is low, and the platform in question has little capacity, then regulatory burdens should also be minimal.
- Ofcom will be under an obligation to create codes of practice which are feasible and which cater for all service providers, whatever their size and capacity. This would include non-business services.
- Ofcom will have a legal duty to assess the impact of its codes of practice and other significant proposals on businesses and wider society which would include individuals within the scope of the regime.
- Ofcom will produce easy to use and easy to understand guidance supporting its codes to avoid the need for individuals and smaller services to seek legal advice.
- Ofcom will take a proportionate and targeted approach to monitoring and enforcement. It will focus on the services where the risk of harm to users is highest. It will seek to engage collaboratively with companies and individuals to help them understand their new duties, and what improvements might be needed, before initiating enforcement action, where this is required. Only in the most egregious cases where regulator engagement has failed is it expected that an individual operating a site would be subject to any of the financial enforcement mechanisms.

245. The Government and Ofcom will continue to engage with individuals (and smaller services) through implementation to ensure any costs are minimal.

Cost to government

Justice impacts

246. The following aspects of Option 1 are expected to result in impacts on the criminal justice system (CJS):

- A deferred power to introduce a new criminal offence for named senior managers who fail to respond to Ofcom's information requests
- Court orders required to apply for business disruption measures
- A new appeals process, via the Upper Tribunal Administrative Appeals Chamber, heard using judicial review principles. Appeals will be made against enforcement decisions, designation as Category 1/2A/2B provider and the designation of companies in scope of the additional transparency reporting threshold.
- An impact on the number of incidences of online illegal activity and content reported to law enforcement and/or other authorities.
- Additional criminal offences in the draft OSB, under Ofcom's information gathering powers, including recklessly submitting false information, providing false or misleading information in interview or failing to attend, destruction of relevant data or falsifying data in response to Ofcom exercising its powers, and obstructing Ofcom accessing premises, data and equipment.

247. A justice impact test has been conducted which has been cleared as having a *de minimis* impact. 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 £7,000 start-up cost and £35,000 running cost (which equates to £3,500 per case and 10 cases expected per year).

248. 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.

249. The creation of an additional offence related to cyberflashing is expected to be implemented through the Online Safety Bill, with the potential for additional costs to law enforcement and the criminal justice system. This is not accounted for within this impact assessment but an additional impact assessment will be produced by the Ministry of Justice, the analysis of which will be incorporated into an updated Online Safety Bill impact assessment.

Requirement to report CSA

250. This section sets out estimated costs to the Government of establishing a body that will be responsible for receiving and processing CSA reports. The costs will vary significantly depending on the number of reports that are made.

251. 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 platforms. 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.

252. UK platforms 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 CSA. 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 CSA.

253. [placeholder: expected to be NCA but not yet confirmed] has been confirmed as the new designated body. 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.

Benefits

254. The calculation of benefits is challenging: it is not possible to develop a precise estimate of the reduction in online harm 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

255. This was the case at the time of writing the previous IA and therefore, the methodological approach remains the same. However, this IA includes one additional quantified harm, namely fraud facilitated through UGC, and also presents costs based on the most recent data. Estimated benefits are illustrative only and have not been included in the NPSV of the policy. The estimates below are the total amount of harm caused online under a number of categories of online harm; no attempt has been made to estimate the proportion of harm avoided by the introduction of this legislation (outside of illustrative scenario analysis). As with other similar uncertain policies, to address the problems with benefit estimations, this IA presents break-even analysis: estimating the reduction in online harm[178] required to exactly match the economic costs.[179]

Methodology

256. There are a wide range of different categories of harm, both illegal and legal but harmful. Of these, a total of eight defined categories of harm have been quantified, at least partially, based on available evidence. These include six illegal categories of harm:

- Contact CSA;
- Modern slavery;
- Hate crime;
- Illegal sales of drugs;
- Cyberstalking; and
- Fraud as facilitated by UGC

[178] Harms which we were able to quantify and are therefore included in the estimated benefits, are: cyberbullying, cyberstalking, intimidation of public figures, CSA, modern slavery, hate crime, drugs facilitated online.
[179] These costs comprise all monetised costs within this IA.

257. Two further categories of harm that are legal but harmful have also been costed:

- Cyberbullying
- Intimidation of public figures.

258. 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 harm. One of the challenges of estimating the online element of illegal harm is that the way in which harm occurs varies, with some harm 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).

259. With regard to the categories of illegal harm that have been quantified, this IA uses data on the prevalence of crime from the ONS which includes experimental data based on the online crime flag. In 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.

260. In addition, many of these categories of harm 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. Many of the harms are likely to have more than one source and while the OSB may address the online element, the harm may still occur through other sources. However, even where a harm may have multiple sources, a reduction or removal of the online element of the harm is not in and of itself inconsequential, and so we would still expect some degree of benefit to this occurring in these specific cases.

261. Another 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.

262. As well as issues relating to the use of the flag by police forces,[180] 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 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.

263. All quantified categories of illegal harm below contain the cost to the CJS, aside from Modern Slavery[181]. The CJS costs for cyberstalking are set out explicitly in the related cost table. For all other quantified categories of harm, 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. For some types of harm 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.

264. The previous IA assumed that quantified harm would grow at 5% per year, this was in line with growth in the amount of hours spent online between 2015-2019. This IA revises this assumption based

[180] This is defined as "An offence should be flagged where any element of the offence was committed online or through internet-based activities (e.g through email, social media, websites, messaging platforms, gaming platforms or smart devices)". Source: Counting Rules Crime Flags, Home Office, updated April 2020.

[181] 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.

on a range of potential proxies. There is evidence that online harm is growing and many survey based measures show increases in the percentage of internet users experiencing harm such as cyberbullying, misinformation, online abuse, and other key categories of online harm.[182] In addition to individuals' experience, there is also evidence that the volume of harmful content is also increasing.[183] While, there is no single indicator for the prevalence and growth of online harm, any estimate will by definition be only an attempt to mirror the real growth of online harm across a ten-year period. Potential growth rates include:

Table 26: Potential online harm growth rates

Measure	Average annual growth
Hours spent online (Ofcom)[184]	4.5% per year between 2015-2020
Total number of videos viewed online (EY)[185]	5% per year since 2017
Percentage of adult population that has recently used the internet (ONS)[186]	1.31% per year between 2015-2020
Adults that have had potentially harmful online experiences in the last 12 months[187]	1.6% between 2019-2020
Percentage of adult population that have recently accessed social networking sites (ONS)[188]	5.1% between 2011-2020 or 4.2% between 2015-2020

265. Based on the available proxies above, this IA estimates that online harm will grow by 3% per year (revised down from 5% previously estimated). The sensitivity section below uses the full range of proxy growth rates (1.3 - 5.1%) to illustrate the effect on the break-even point

Quantified harm

Contact CSA

Contact CSA[189] was estimated to result in costs of at least £10 billion in the year ending 31 March 2019.[190] While non-contact abuse is a feature of CSA, the harm from this form of abuse cannot currently be quantified. Estimates for contact CSA include the financial and non-financial (monetised) cost relating to all victims who continued to experience contact sexual abuse, or who began to experience contact sexual abuse, in England and Wales in the year ending 31 March 2019. As this cost is of victims whose abuse lasts multiple years it is not a true annual cost and as such it needs to be divided by the average length of a CSA case in order to produce an annual estimate.

266. Using the CSEW, that looked at the time abuse lasted for adults who had been abused as children,[191] this IA can estimate the average time. Two assumptions are needed to calculate the average, first that abuse for adults who responded to the survey is representative of the average length of abuse for current victims. Second, it is assumed that for those that selected 'abuse lasted for less than a year' the abuse lasted one day. This is a conservative minimum that has been chosen due to the

[182] Internet users' concerns about and experience of potential online harms (Ofcom & ICO, 2018-2020); Leading Bullying Research (Ditch the Label).
[183] National Center for Missing and Exploited Children, by the numbers (NCMEC)
[184] Ofcom's Adults' Media use and attitudes reports (2015-2020)
[185] Understanding how platforms with videosharing capabilities protect users from harmful content online (EY, 2021)
[186] Internet users, UK statistical bulletins (ONS, 2015-2020)
[187] Internet users' concerns about and experience of potential online harms (Ofcom, 2019-2020)
[188] Internet access - households and individuals (ONS, 2011-2020)
[189] For definition of contact CSA, see page 107 in Working Together to Safeguard Children (HMG, 2018)
[190] Tackling Child Sexual Abuse Strategy 2021 (HMG)
[191] Child sexual abuse in England and Wales: year ending March 2019 (ONS, 2020)

absence of evidence suggesting a longer period of abuse. This results in an estimated length of abuse of 2.17 years.

267. Dividing the estimate for the cost of contact CSA by 2.17 gives an annual cost of contact CSA of £4.93bn. This cost includes the lifetime impacts of contact CSA victims. This cost is for all forms of contact CSA and not just CSA that has an online element. The estimated proportion of contact CSA that includes an online element is estimated using the online crime flag. This assumes that the proportion of recorded offences with online elements is similar to the proportion of victims that experience abuse with online elements. This assumption is used as the online flag is the best available proxy for estimating how much of the total cost of contact CSA may be attributable to CSA with an online element. Police flagging data from April 2020 to March 2021 estimated that 20% of all recorded contact CSA offences had an online element. In this analysis it is assumed that this proportion is constant across the appraisal period.

268. The true level of online offences may be higher than 20%, due to issues with the flag and the proliferation of online technology. First, the flag is typically manually applied by officers, and therefore accurate use relies on officers being aware of the flag, remembering to apply it to specific cases, and recognising that an online element is present. This can be difficult in some cases, such as where online messaging services like WhatsApp or Kik are used, which officers may not recognise as 'internet enabled' or online. Second, the flag differs in usage between forces and is not evenly applied throughout England and Wales. Third, online technology has proliferated over the last decade, which gives offenders more opportunities to target children, with 88% of children having a smartphone aged 12.[192]. Trends like this will likely have increased the proportion of contact CSA with an online element to beyond the 20% estimated using the police recorded crimes online flag.

269. Table 27 summarises the data and calculations used to estimate the impact of contact CSA with an online element. This is calculated by multiplying the estimated annual cost of all contact CSA (£4.93bn) by the estimated proportion of contact CSA that includes an online element (20%). This gives the estimated annual cost of contact CSA with an online element (£0.99bn).

Table 27: Online contact CSA annual cost (2021/22 prices)

Harm	Estimated annual cost	Proportion online	Annual cost with online elements
Contact CSA	£4,928 m	20.1%	£993 m

270. It is worth further emphasising that this cost (£0.99bn) is a likely underestimate for the true scale and impact of contact CSA with an online element. The impact may be greater because the true cost of contact CSA is likely higher than £10.7bn due to the irregular use of the online flagging tool and the fact that estimating the full cost in all areas of abuse is difficult and sometimes unquantifiable. Additionally, the Government is aware that having a monetary cost for abuse may seem reductive to those that have experienced CSA and recognises the profound human costs of CSA to victims and survivors.

Modern slavery

271. This section considers the economic and social cost of physical modern slavery offences with an online element.

Table 28: Modern slavery annual cost (2021/22 prices)

Harm	Prevalence	Unit cost	Annual cost
Modern slavery with an online element	29	£0.4 m	£10.7 m

272. The unit cost of modern slavery is £328,720.[193] It covers the costs of physical and emotional harm, the cost of lost output and time, costs to health services, costs to victim services and law enforcement costs. This unit cost is given in 2016/17 prices. Inflating the estimate to 2021/22 prices,

[192] How many children have their own tech? (YouGov, 2020)
[193] The economic and social costs of modern slavery (Home Office, 2018)

provides an estimate of £366,065. 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.[194] 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.

273. It is important to note that the cost of modern slavery is calculated on a victim basis. It is a cost of new cases of modern slavery identified between April 2016 and April 2017 that may continue into succeeding years and does not capture those ongoing identified prior to the identification year. It is difficult to ascertain whether this will mean this an over or underestimate for the cost that modern slavery has to society as the average length of modern slavery cases vary between exploitation types. The median durations of 'labour exploitation' and 'sexual exploitation' are both 274 days, whereas 'domestic servitude' lasts 730 days on average.[195]

274. 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 contact CSA, 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.

275. The approach used above is to use the police recorded crime data,[196] where there were 8,730 recorded modern slavery offences between and April 2020 and March 2021. This is then multiplied by the proportion of cases flagged by the police as having an online element (0.33%) to give 29 cases with an online element. This figure is likely to underestimate the true prevalence of modern slavery given the limitations of the online crime flag.

276. As with other categories attempted to be quantified, it is also important to note that these figures were reported during the COVID-19 pandemic and may be unreflective of traditional crime trends with pandemic related restrictions causing disruptions to criminals and victims alike. This may have affected the ability of victims to report offences to the police, as well as potentially leading to a greater proportion of offences taking place online.

Hate Crime

277. Hate crime is not a crime category in its own right, but instead refers to a subset of other crime categories which have been motivated by hate. These offence categories include violence against the person (VATP, public order offences, criminal damage and arson offences, and other notifiable offenses. To obtain a proxy measure for the number of online hate crimes, this IA looks at offences measured by police as being racially or religiously aggravated and also flagged as having an online component. The quantification of harm is focussed on VATP, as this represents the majority of racially or religiously aggravated offences that are flagged as online and cost data is not available for the other offences. Additionally, hate crime could be motivated by other factors such as sexual orientation, disability, and transgender identity. Therefore, this cost estimate is likely to underestimate the total cost of online hate crime. The table below summarises the data and calculations used to estimate the impact of hate crime with an online component.

Table 29: Hate crime online annual cost (2021/22 prices)

Harm	Prevalence	Unit cost	Annual cost
Hate crime: racially or religiously aggravated offences with injury	30	£16,033	£0.5 m
Hate crime: racially or	688	£6,767	£4.7 m

[194] 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.

[195] The economic and social costs of modern slavery (Home Office, 2018)

[196] Police recorded crime and outcomes open data tables (Home Office, updated 2021)

religiously aggravated offences without injury			
Total			£5.1 m

278. The prevalence of online racially or religiously aggravated offences was calculated using statistics from police recorded crime from April 2020 to the end of March 2021. Within this time period, there were 65,572 racially or religiously aggravated offences recorded by police forces within England. Of these racially or religiously aggravated offences, 2,762 offences were 'with injury' offences, and 62,810 were 'without injury' offences.

279. To calculate the proportion of online racially or religiously aggravated offences within the overall category of racially or religiously aggravated offences, the police online crime flagging tool data is used. The estimate of the proportion of offences that were committed online or enabled by online devices is 1%. This gives a prevalence of 30 racially or religiously aggravated with injury offences, and 686 racially or religiously aggravated without injury offences which have the online harm flag applied.

280. This figure is likely to underestimate the true prevalence of online racially or religiously aggravated offences in the UK given that these offences often go unreported and previously noted issues with the online crime flag. Additionally, the proportion of online racially or religiously aggravated offences is low (1%) due to a high proportion of racially or religiously aggravated offences being only able to be committed online, with 73% of all racially or religiously aggravated offences being racially or religiously aggravated public fear, alarm or distress offences which are highly unlikely to be committed within the online sphere. This means the proportion of online racially or religiously aggravated offences recorded by police may be significantly lower than the amount of online hate crime committed.

Illegal sale of drugs online

281. Combating the sale of drugs online is a key element of Option 1, with drugs increasingly being sold using a variety of online methods including via social media. Using police online crime flagging data, it was estimated that 1% of all drugs supplied (using offence code 92A) are supplied using online methods or online enabled methods. This gives an estimate of 5,204 drugs supply offences that contain an online element. These recorded figures are very low compared to the total number of drugs supply offences recorded (669,114). This could be due to the way the online flag is applied to drug cases. It also could be because it is difficult to prove there is an online component, and that international drug trafficking (via import) may not be recorded as online activity and this offence type drives/fuels domestic drug supply.

282. 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.[197] Inflating this figure from 2016/17 prices to 2021/22 provides a total estimate of £22.27 billion. The proportion of recorded drugs offences flagged as online was around 0.78% in 2021/22. Combining these proportion and cost estimates provides an indicative estimate of the cost of drugs offences flagged as having an online component of around **£173.2 million**.

Table 30: Illegal sale of drugs online annual cost (2021/22 prices)

Harm	Prevalence	Unit cost	Annual cost
Illegal sale of drugs online	5,204	n/a	£173.2 m

Cyberstalking

[197] Review of Drugs - evidence relating to drug use, supply and effects, including current trends and future risks (Dame Carol Black, 2020)

283. There is no single definition of cyberstalking, however it is widely used to refer to the repeated use of online communications tools to stalk, harass or frighten a victim. Currently there is no formal governmental definition of cyberstalking and this makes collecting data specifically on this area difficult. Within this analysis, alternative proxies such as proportion of stalking offences with the online flag applied have been used but these are only proxies and may not reflect the true prevalence of this crime. Table 31 sets out the data and calculations used to estimate the impact of cyberstalking.

Table 31: Cyberstalking annual cost (2021/22 prices)

Harm	Prevalence	Unit cost	Annual cost
Cyberstalking	305,556	£33,052	£10,099 m

284. The unit cost of a cyberstalking incident is based on the cost to a victim of a stalking incident from a 2019 Home Office report.[198] Using the work of Paladin, the national stalking advocacy service, cyberstalking inflicts the same amount of psychological damage as offline stalking and therefore it was deemed appropriate to use the costings relating to all stalking in this analysis.[199] The unit cost comprises three elements (prices in 2016/17 prices): 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 2021/22 prices.

285. Due to stalking and cyberstalking being both highly personal and private crimes, they are often underreported to police and other forms of authority. For this reason, the estimate of the prevalence has been taken from the CSEW.[200] This estimated the number of victims of stalking to be 1,504,000 in the year from April 2019 to March 2020. Due to COVID-19, the CSEW for 2020-21 did not run and therefore for the purposes of this analysis, it has been assumed that the prevalence of stalking has not changed from 2019-20 to 2020-21. Therefore, it is assumed there were also 1,504,000 victims in the year from April 2020 to March 2021.

286. The CSEW does not include information relating to whether the stalking had an online element. Therefore, to calculate the proportion of cyberstalking incidents within the overall category of stalking, the police online crime flagging tool is used. Using a flagging period from April 2020 to March 2021, it is estimated that 20% of stalking offences had an online element and can be defined as cyberstalking. This leads to there being an estimated 305,556 cyberstalking victims in the year from April 2020 to March 2021. This assumes that the proportion of recorded offences with online elements is similar to the proportion of victims that experience stalking with online elements.

287. This figure is likely to underestimate the true prevalence of cyberstalking in England and Wales as cyberstalking is often unreported and the noted issues with the online crime flag. It is also important to note that due to these figures being recorded during the time of the coronavirus pandemic, an increase of online crime in the stalking and harassment area is likely to be expected. The Suzy Lamplugh Trust reports that 83% of those surveyed reported an escalation of their abuse over the period of the pandemic.[201] Their report on the changing landscape of stalking during the pandemic also highlighted that 100% of cases presented to the National Stalking Helpline now involve a cyber element.[202]

Fraud

288. There were 3.7 million instances of fraud in England and Wales in the year ending March 2020[203] and over half of these had some online element.[204] However, the legislation does not cover every form of fraud which is cyber-enabled or cyber-dependant and therefore, further consideration has been taken to provide a better estimate of the proportion of fraud which Option 1 could potentially address.

198 The economic and social costs of domestic abuse (Home Office, 2019)
199 Stalking and Harassment - a Shorthand Guide about Digital and Cyberstalking (Paladin, 2014)
200 Stalking: findings from the Crime Survey for England and Wales (ONS, 2020)
201 National Personal Safety Day 2020 - Cyber Safety at Work (Suzy Lamplugh Trust, 2020)
202 Unmasking Stalking: A changing landscape (Suzy Lamplugh Trust, 2021)
203 Crime Survey for England and Wales - year ending March 2020
204 Nature of crime: fraud and computer misuse - year ending March 2020

289. The Economic and Social Cost of Crime estimates that the average cost per fraud incident is £1,472.[205] This impact assessment draws on three main data sources to assess the impact of fraud, namely the National Fraud Intelligence Bureau (NFIB), Action Fraud (AF) and the Crime Survey for England and Wales (CSEW). The CSEW estimates 54% of fraud has 'some online element', whereas NFIB analysis suggests over 80% of fraud is cyber-enabled/dependent.[206] Note the CSEW presents a prevalence estimate, whereas NFIB analysis uses fraud reports. Due to the limitations of the data available we had to engage with expert colleagues to generate an estimate for the proportion of fraud in scope of the legislation.

290. Fraud is highly underreported with only 14% of the estimated CSEW offences reported to Action Fraud in the year ending March 2020. As such, the approach taken has been to downscale the overall prevalence estimate the CSEW provides to remove frauds which are likely to have been out of scope of the Bill. Based on the CSEW and NFIB estimates for frauds having an online element, as well as engagement with stakeholders and colleagues, the scale of frauds in scope has been estimated to be 45% as a mid-estimate, allowing for the exclusion of email enabled fraud. Low and high estimates are also given to highlight the uncertainty around the mid-estimate. It should be noted that whilst the Bill covers 45% of fraud presently and although we expect the Bill to have a significant impact on introduction, it is reasonably likely that fraudsters will displace to alternative means of defrauding victims. This approach has the following limitations:
    - The unit cost of fraud is taken from 2015/16 data and could be outdated. Additionally, this method applies the same unit cost to all fraud types.
    - This method uses historical fraud prevalence estimates and it is possible that fraudsters may divert to other methods to avoid detection because of the Bill, such as email scams which are out of scope.
    - This method may underestimate benefits if companies work beyond the scope and have a greater impact.
    - The estimate of scale is based on stakeholder and expert engagement, rather than a reliable data source due to limitations in what is available.

Table 32: Fraud annual cost (2020/21 prices)

	% potential in scope of Option 1	Number of offences	Annual cost
Low	30%	1.4 million	£2,049 million
Mid	45%	2.1 million	£3,074 million
High	60%	2.8 million	£4,098 million

291. For the main calculations the mid estimate is taken; however, both the low and high estimates are tested in the sensitivity analysis.

Cyberbullying

292. 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.[207] 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.

293. 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 day and night 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

[205] 2021/22 prices.
[206] Annual Assessment (NFIB, 2021)
[207] What is Cyberbullying? - (StopBullying, 2018)

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.[208]

294. 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. Table 33 below outlines the core unit costs for the central estimate of the economic costs of cyberbullying.

Table 33: Annual impact of cyberbullying (2019 prices)

	Category	**Prevalence**	**Unit cost**	**Annual cost**
Cyberbullying	Direct impact on victim	911,587 (19% of 10-15 year olds)[209]	£640	£583.0m
	Cost to health services of treating related depression	84,996 children accessing specialist mental health treatment	£354*	£30.1m
	Cost of treating cyberbullying related self-harm	*1,943 children*	*£838**	*£1.9m*
	Total	**911,587**	**£673**	**£613.2m[210]**

*Assumption of one incident in a given year

295. There were an estimated 4.8 million children in the UK aged 10-15 in 2020.[211] Estimates for the proportion of children who have experienced cyberbullying can vary depending on the study used. The most up-to-date studies from 2020 that address the question of prevalence are from the ONS and Ofcom. The ONS estimate that 19% of 10 to 15 year olds were cyberbullied in the year ending March 2020,[212] whilst the Ofcom figure is 26% and covers 12 to 15 year olds.[213] The ONS prevalence figure is used in this IA as it covers a broader age range of children. This central prevalence estimate of 19%, equates to 911,587 child victims of cyberbullying in the UK in a given year.

296. Based on the range of prevalence estimates in the studies observed,[214] sensitivity analysis is conducted using 7% and 26% as upper and lower bounds which results in lower and upper bound estimates of 336,000 and 1,250,000 cyber bullied children. 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.

297. 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);

[208] Bringing an end to online bullying: Whose job is it anyway? - (Anti-Bullying Alliance, 2019)
[209] Online bullying in England and Wales Online bullying in England and Wales: year ending March 2020 (ONS, 2020)
[210] This figure does not include the cost of treating cyberbullying related self harm - this is shown illustratively only given the minimal evidence base.
[211] ONS Population Estimates (ONS, 2020)
[212] Online bullying in England and Wales Online bullying in England and Wales: year ending March 2020 (ONS, 2020)
[213] Internet users' experience of potential online harms: summary of survey research (Ofcom and ICO, 2020)
[214] A number of studies were reviewed as part of work to understand the likely prevalence of cyberbullying. As well as the ONS and Ofcom studies already mentioned the following studies were also reviewed: Annual Bullying Survey (Ditch the Label, 2017); Mental Health of Children and Young People in England (NHS Digital, 2017); The Suffolk Cybersurvey (2017) Bullying in England, April 2013 to March 2018 Analysis on 10 to 15 year olds from the Crime Survey for England & Wales (DfE, 2018). 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%.

- 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).[215]

298. These are multiplied together to give an estimate of the average cost associated with the crime. 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%.[216] 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.[217] 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.[218] 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.

299. 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. This gives a central estimate of 337,000 children per year suffering from depression as a result of cyberbullying. 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.[219] 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:[220]

- 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% of these are high-intensity at a cost of £1,125.

300. 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. 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.

301. 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 **£30.1 million.**

302. The 2017 Annual Bullying Survey found that 25% of cyberbullying victims surveyed went on to self-harm. This implies that around 228,000 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 how many of these children formally seek help or attend hospital as a result is uncertain. Based on a study in the Lancet, the average cost to UK hospitals of treatment of self-harm is £809 per incident.[221] 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 cyberbullying require hospital treatment (1% or 1,943 cases), this would result in an annual cost to the NHS of £1,910,000. Given the difficulty in ascertaining exactly how many of those who self-harm due to cyberbullying would go on to require NHS

[215] 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.
[216] This represents the estimated impact of a mild episode of a depressive disorder - see below footnote for further information.
[217] The Economic and Social Costs of Crime (Home Office, 2018). The estimate for 'REDUCEQL' originally comes from Disability weights for the Global Burden of Disease 2013 (Saloman et al., 2015). The duration (DUR) (0.167 years or two months) is an average originally derived from Impact of crime on victims (Wasserman and Ellis, 2007).
[218] Annual Bullying Survey (Ditch the Label, 2017)
[219] Mental Health of Children and Young People in England, 2017 (NHS Digital, 2018)
[220] Resource impact statement: Depression and anxiety disorder (NICE, 2015)
[221] General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis (Tsiachristas et al., 2017)

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.

303. In the previous IA, an estimate for the lifelong impact of cyberbullying was included which explored the long-term economic impact associated with childhood bullying. As this estimate was based on a single academic study,[222] and was not included in the central cost estimate for cyberbullying, it has been decided not to include this estimate again until a more comprehensive and established evidence base around the long-term effects of cyberbullying becomes available.

Intimidation of public figures

304. Estimates for the baseline cost of intimidation of public figures remain unchanged since the previous IA. Work by the Alan Turing Institute is underway to better understand the prevalence and impact of this harm; however, estimates are not yet available. 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.

305. 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 34: Annual impact of online intimidation of public figures (2019 prices)

	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[5]	£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
Total (quantifiable) £4.6m				

[222] Long Term Economic Impact Associated with Childhood Bullying Victimisation (Brimblecombe et al., 2018)

306. 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.[223] 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.

307. According to research from Demos[224] (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). It is assumed that investigating and responding to a threatening tweet takes on average, one hour of police time (£37).

308. Surveys of MPs have found that many MPs feel fearful.[225] 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-respondents to the survey experienced no harm 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.[226]

309. 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[227] found that women – particularly black, Asian and minority ethnic women – experience more targeted abuse (such as gendered insults and greater incidence of threats).

Qualitative benefits

310. The approach taken in this IA is to attempt to quantify a subset of online harm which occurs under the baseline and conduct break-even and scenario analysis. Data on harm is limited and this IA is only able to monetise a small subset. This section presents evidence on additional harm for which evidence is insufficient to include within the break-even analysis, and a number of non-monetised benefits expected to accrue as a result of Option 1.

Mis- and disinformation

311. Disinformation refers to the deliberate creation and dissemination of false and/or manipulated information intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain. Misinformation refers to inadvertently spreading false information.

312. Both misinformation and disinformation are widespread online. Between 2018 and 2020, Facebook and Twitter announced that they had taken down 147 influence operations.[228] In late March 2020, 46% of adults who said they were accessing news or information about the pandemic said that they had come across news or information that they deemed false or misleading.[229] In the first quarter of 2021, six in ten UK adults who said they had seen misinformation about COVID-19 said they had seen it at least once a day.[230]

---

223 Parliamentary Monitor Snapshot - IFG (2019).

224 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.

225 For example, The Personal Security of Individuals in British Public Life - Demos (2018)

226 The Economic and Social Costs of Domestic Abuse - Home Office (2019)

227 Black and Asian Women MPs Abuse More Online - Amnesty International UK (2017)

228 How disinformation evolved in 2020 - Influence operations are referred to by Facebook as 'coordinated inauthentic behaviour' and by Twitter as 'state-backed information operations'

229 Online Nation 2021 report - Ofcom 2021

230 Online Nation 2021 report - Ofcom 2021

313. Engagement with mis/disinformation, particularly around the anti-vaxx rhetoric, has also increased. Between July and August 2020 interactions on posts criticising COVID-19 vaccines on six UK Facebook pages increased by 350%.[231] The following of anti-vaxx social media accounts has also increased to 58 million followers, a 19% increase since 2019.[232] The prevalence and impact of misinformation is also significant during election periods. During the 2016 US election the top 20 fake news stories that were circulating had more engagement than the top 20 factual news stories on social media.[233]

314. Online mis/disinformation results in real world impacts. The public's willingness to accept the COVID-19 vaccine is highly responsive to the information available about the vaccine. Based on a nationally representative survey, one study found that there was a net decrease in intent to accept a COVID-19 vaccine in respondents exposed to related misinformation. Similar results were found in a randomised control trial looking at the effects of anti-vaccine conspiracy theories.[234] Another found that a one-point shift up the disinformation scale[235] was associated with a 15% increase in negative tweets about vaccines and a two percentage point decrease in the average vaccination coverage year over year.

315. Evidence on the realised cost of mis/disinformation is limited. It is often difficult to establish a causal link between exposure online to real world impacts and in addition, many of the impacts manifest in difficult to monetise areas such as democracy and trust. Online mis/disinformation is associated with the following non-monetised impacts:

- **Health impacts**: by testing the impact of online misinformation on mask wearing through surveys, the effectiveness of masks on the spread of COVID-19, and the resulting increase in cases, hospitalisation and deaths - London Economics[236] estimated that the direct impact of online mask misinformation on the NHS could have been £22.1 million (up to Q4 2020) with a potential indirect impact of £3.6 billion over the same time-period through prolonging non-pharmaceutical interventions.
- **Impacts on democratic institutions**: Attempts to interfere in our democratic processes of the kind we saw in advance of the 2019 General Election have the potential not only to impact the outcome of our democratic events, but also to lead to questions about their validity, resulting in decreasing trust in government and democratic institutions. The British Social Attitudes Survey[237] found that only 15% of respondents trust the Government either 'most of the time' or 'just about always' - the lowest level in more than 40 years.
- **Impacts on trust in media**: due to the difficulty in assessing the validity of online information and the speed with which it can be created, mis/disinformation can lead to individuals becoming increasingly sceptical of the mainstream news media.
- **Market impacts**: through large scale falls in stock markets such as a fall of 38 points in the S&P 500 based on false information relating to the US-Russia election inquiry.[238]
- **Wasted advertising revenue**: well-known brands are finding that a large part of their advertising spend is going to disinformation sites, including extremist websites. This spend can damage advertisers' brands and make spreading disinformation a profitable activity.
- **Costs to business and government of counter-disinformation efforts**: businesses like Facebook are spending large amounts of money annually on their safety systems, which includes countering disinformation - one estimate puts this spend at over £2.3bn.[239] Governments around the world are also incurring costs to counter disinformation, for example, Australia's Electoral Integrity Task Force created to counter cyber-attacks including disinformation campaigns.

---

[231] Online Nation 2021 report - Ofcom 2021

[232] The Anti-Vaxx Industry - Centre for Countering Digital Hate 2020

[233] How Fake News Affects US Elections - 2020

[234] The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions - Jolley and Douglas 2014

[235] This indicator tracks the opinion of over 3,000 scholars in 180 countries and asks them "How routinely do foreign governments and their agents use social media to disseminate misleading viewpoints or false information to influence domestic politics in this country?" on a 5-point Likert scale.

[236] The Cost of Lies: Assessing the human and financial impact of COVID-19 related online misinformation on the UK - London Economics - 2020

[237] Fairness and justice in Britain

[238] The Economic Cost of Bad Actors on the Internet - Fake News - 2019

[239] The Economic Cost of Bad Actors on the Internet - Fake News - 2019

316. Option 1 is expected to reduce the prevalence and impacts of online mis/disinformation compared to the *do-nothing* option through user reporting, additional content moderation, transparency reporting, and ensuring platforms properly assess the risk of this harm.

Terrorism/extremism

317. Terrorism and its associated fear and damage to society are a major harm that the OSB seeks to mitigate and prevent. This is through tackling the way that terrorists and those that want to harm our society use the internet to recruit, radicalise and propagandise.

318. The evidence of terrorist content online and its harm is often hard to show due to the way the content is hidden and the lack of a direct route to show its impact. However, there is clear evidence that the internet is a route for those being radicalised to become extremists and from then terrorists. Evidence from a Ministry of Justice sample of extremist prisoners found that for cases prior to 2005, 83% were radicalised face to face with only 17% radicalised using a mixture of online and face to face. For cases from 2015 to 2017 this had increased dramatically to 56% radicalised using a mixture of online and offline, 27% purely online and 17% just offline, showing a reversal of the pathways to extremism.[240] This shows that the internet is increasingly being used by terrorists to radicalise those that go on to commit or attempt to commit harm in society.

319. Additionally, the internet is used by terrorists to propagandise and promote terrorism. This was most obvious on 15 March 2019 where there was a terrorist attack in Christchurch, New Zealand, with 51 fatalities and 50 injured. The perpetrator live streamed the attack while committing it, and over the 17 minutes that attack lasted it was viewed around 4,000 times before being removed.[241] Since then it will have been viewed countless times as it was recorded by those supportive of extreme right-wing terrorism and shared to other websites and forums. While this is one horrific incident, the scale of the issue is also large, with Twitter suspending 371,309 accounts from December 2017 to December 2018.[242] This is more than a thousand accounts each day, showing the way that terrorists flood the internet with their insidious message in order to convince others to join or support them.

320. Although it is hard to quantify the benefit of the removal of terrorist content and activity from the online sphere, it's removal will almost certainly have an effect on the level of terrorism in society. This decrease is challenging to estimate but any decrease would be significant in terms of the scale of the problem. In 2018 RAND Europe estimated that between 2004 and 2016, the UK suffered around £41 billion of economic losses (in 2021/22 prices) due to terrorist attacks and related activity.[243] This shows just how damaging terrorism is to society and the economy and any activity to hinder the ability of terrorist groups to radicalise, recruit and propagandise carries a significant benefit.

Underage exposure to pornographic content

321. A significant proportion of children access pornography online both inadvertently and intentionally. 51% of children as young as 11-13 years old have seen pornography, with this rising to 66% and 79% for 14-15 year olds and 16-17 year olds respectively. Many children - some as young as 7 years old - stumble upon pornography online. 61% of the 11-13 year olds in who have seen pornography, describe their viewing as mostly unintentional. Underage exposure to pornography can result in children feeling 'grossed out', 'confused', 'disturbed' or upset. Many of the children who had seen pornography at such a young age felt that it was unhealthy to have seen such content at that age.

322. Current evidence does not allow robust quantification of the baseline impacts of children's access to pornography. However, there are a range of short and long term impacts that clearly demonstrate the scale of the problem. For example, evidence suggests that pornography can create damaging insecurities in children and young people. 35% of children said they worry about what other people think of their body because they do not look like the actors they see in pornography. 19% of girls, and 17% of the boys said that they had "learnt if I look normal naked" from watching porn and 29% said that pornography makes them feel bad about their body. The content of pornography can also skew young

[240] Exploring the role of the Internet in radicalisation and offending of convicted extremists (MoJ, 2021)
[241] Christchurch Call (Ministry of Foreign Affairs and Trade NZ)
[242] Twitter suspended 166,153 accounts for terrorism content in second half 2018 (Reuters, 2019)
[243] The cost of terrorism in Europe (RAND, 2018)

people's view of sex, 30% of boys and girls agree that "real sex hasn't lived up to my expectations from watching porn".

323. Pornography can also influence young people's sexual behaviours and expectations towards more "rough" and "forceful" sexual encounters. Meta-analysis from 2017 shows how those who consume porn frequently are more likely to hold sexually aggressive attitudes and be engaged in sexual behaviour that is conducive to sexual agression. Some young girls are worried that boys who watch porn will think it is normal to proceed being rough and forceful when a women's body language indicates that they do not want sex.

324. Some children also feel that porn has affected their or their partner's view of consent, since it is often only implied in porn and not explicitly given. Children who intentionally sought out pornography had the most worrying ideas around consent (by a factor of between three and six in comparison to those who had mostly seen it by accident). 29% of these children did not think consent was needed if "you knew the person really fancies you", in comparison to only 5% of those that had mostly seen pornography by accident.

325. Research also finds that boys who consumed pornography when they were 12-14 years of age, are more likely to have engaged in aggressive, sexual behaviour. In a longitudinal study, 10-15 year olds that consumed violent pornography were six times more likely to be sexually agressive than those who did not consume it, or than those who consumed less aggressive pornography.

326. While it is not possible to monetise the impact of underage exposure to pornography, it has clear and significant effect on children's attitudes and behaviours. Both the core child safety duties and pornography provision will ensure platforms protect children from this content. On this basis, Option 1 is expected to result in material reductions in the short and long term impacts of children's access to pornography.

Other non-monetised benefits

327. In addition to a reduction in online harm over the appraisal period, Option 1 is also expected to result in the following non-monetised benefits:

- **Benefit to law enforcement**: these benefits are expected to accrue both in terms of a general reduction in online crime and through creating a transparent regulatory system, making it easier for law enforcement to tackle crime online. Requirements to report CSA content to law enforcement, transparency reporting, and Ofcom's information gathering powers are expected to contribute towards the accrual of these benefits, which are likely to take the form of cost savings or efficiency gains. The level of online crime reduction and the way in which both platforms and law enforcement operate within the online safety framework is unknown at this stage and this IA is therefore unable to monetise this potential benefit.
- **Increase in media literacy levels**: many of the steps businesses take to comply with the OSB are likely to result in improvements in media literacy levels. For example, these benefits may accrue from steps taken by platforms to keep users safe, such as warnings and flags, giving users the ability to control the content they see, and other tools related to literacy by design. In addition, the Government's related non-regulatory media literacy interventions are aimed at improving core media literacy skills and giving users the ability to keep themselves safe online. Given that platform actions are unknown at this stage and measurement of media literacy is still evolving, this potential benefit remains non-monetised.
- **SafetyTechnology**: Option 1 is expected to result in an increase in demand for safety technology and the Government is supporting the sector through a series of non-regulatory interventions, such as research, investment, and challenge funds. Modelling conducted by Perspective Economics[244] estimated that a combination of the incoming online safety regulations and non-regulatory initiatives could create an additional £900 million in revenue and 3,500 FTE jobs in the lead up to the regime. This benefit remains outside of the scope of this IA for two reasons: first, benefits are expected to accrue before the appraisal period for this IA and long-term modelling has not been conducted; and second, results do not distinguish between the impact of the legislation and the impact of non-regulatory initiatives. Any benefit to the Safety Technology

[244] Internal modelling conducted for DCMS

sector resulting from the legislation would be considered 'resources used to comply with regulation' as set out in RPC guidance.

- **Evidence**: Option 1 is expected to result in an increase in the evidence base underpinning online harm through greater transparency and data availability.

Break even and scenario analysis

328. As outlined in the above sections, this IA quantifies the annual social cost - under baseline - of a subset of online harm, including both illegal and legal but harmful. Online harm is assumed to grow at 3% per year and Table 35 outlines the estimated cost in the first year of the appraisal period.

Table 35: Annual social cost of online harm in the first year of the appraisal period (2024)

Harm	Annual cost to society
Contact CSA	£916.4 million
Modern slavery	£9.8 million
Hate crime	£4.7 million
Illegal sale of drugs	£159.8 million
Cyberstalking	£9,315.7 million
Fraud (central)	£2,835.2 million
Cyberbullying	£601.4 million
Intimidation of public figures	£4.5 million
Total annual	**£13.8 billion**
Total across the appraisal period (10-year PV)	**£135.5 billion**

Note these cost figures do not align exactly with the analysis above of the individual categories of harm. These have been uplifted in line with the expected growth in online harm to 2024 - the first year of the appraisal period and converted to 2019 prices and 2020 present value base year.

329. Given the difficulty in providing an evidenced estimate for a percentage reduction in online harm resulting from Option 1, the benefits remain purely illustrative and are not considered in the calculation of the NPSV.

330. The illustrative benefit is the value of a reduction in online harm. Given the data limitations described above, this IA has only been able to quantify estimated benefits for a reduction in the subset of online harm outlined above. It is assumed that, once enacted, a policy will start to reduce online harm in the second year of the appraisal period.[245]

331. As outlined in the previous IA, evidence on the likelihood of benefits occurring remains 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 under Option 1. This is due to the complex nature of online harm, 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 harm resulting from platforms' responses to Option 1. However, the OSB is expected to lead to a

[245] The reduction is relative to the estimates of harm under BAU and is not applied cumulatively. The year in which reductions would start will depend on the year in which regulation is enacted.

reduction in online harm compared to a *do-nothing* baseline through the following mechanisms (this is not an exhaustive list):

Table 36: Qualitative assessment of why the OSB is expected to result in reduced harm

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[246]. This highlights how systems and processes to moderate content can mitigate the impact of online harm. The OSB is expected to lead to some platforms conducting additional content moderation to address online harm. 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[247][248]. 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 OSB, platforms will be expected to accommodate user reporting of harm 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.
Age assurance	Both the core child safety duties and pornography provider provision will result in increased age assurance processes online. This will ensure children are protected from age inappropriate material and mitigate the short and long-term impact of harms such as children's exposure to pornographic content. Under the *status quo*, the vast majority of pornography sites (and sites where pornography can be accessed) do not have age assurance systems. Where they do exist, they are light-touch and ineffective, such as a user confirming they are over 18 by ticking a box. Option 1 will ensure that platforms hosting pornography have effective and accurate age verification processes in place minimising children's access to pornography online.
Anti-fraud measures	Fraud facilitated both by UGC and advertisements online, lead to significant victim losses. Reporting and content moderation measures alongside increased customer due diligence on advertisers is likely to result in a material reduction in online scams.
Transparency and user behaviour	Category 1, 2A and 2B services will be expected to publish transparency reports under the OSB and Ofcom will have a range of information gathering powers as well as a responsibility to conduct research into online harm. In addition, alongside the OSB the Government is undertaking a number of projects and initiatives aimed at improving media literacy. Giving users more information about the risks and prevalence

[246] Community Standards Enforcement Report - Facebook (2021)

[247] User reporting was the first source of detection

[248] YouTube Community Guidelines Enforcement - YouTube (2020)

	of online harm on platforms and the Government's initiatives related to media literacy are both expected to increase user safety online and mitigate some of the impacts associated with online harm.
Risk assessments	The OSB 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.

332. Given the uncertainty around the reduction in online harm that could be achieved under Option 1 (as described above), this IA estimates the reduction in the subset of quantified online harm required to exactly match the costs, that is, the scale of the reduction of harm required to deliver a benefit-cost ratio of precisely 1. The results are shown in Table 29.

Table 37: Break-even point

	Low	Central	High
Option 1	1.5%	2.1%	2.7%

333. 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 Option 1 in reducing harm:

- Low reduction scenario = 1% per year compared to a *do nothing* counterfactual
- Mid reduction scenario = 3% per year compared to a *do nothing* counterfactual
- High reduction scenario = 5% per year compared to a *do nothing* counterfactual

334. Based on these scenarios, Table 30 compares the costs and benefits.

Table 38: Benefit cost ratios (BCR) under illustrative scenario (central estimate only)

Low reduction scenario					
Low estimate		**Central estimate**		**High estimate**	
Implied BCR	0.68	Implied BCR	0.49	Implied BCR	0.37
Mid reduction scenario					
Low estimate		**Central estimate**		**High estimate**	
Implied BCR	2.04	Implied BCR	1.46	Implied BCR	1.11
High reduction scenario					
Low estimate		**Central estimate**		**High estimate**	
Implied BCR	3.40	Implied BCR	2.43	Implied BCR	1.85

Indirect costs and benefits

Freedom of expression

335. The Government has engaged extensively with rights groups, industry, trade associations, Parliament, and civil society on freedom of expression implications. This engagement has helped shape the legislation to ensure that it is designed to strengthen freedom of expression and has built in safeguards to avoid any potential negative impacts. This section sets out why the Government expects these proposals to enhance freedom of expression online rather than limit it.

336. Under the *status quo*, major technology companies already exercise significant power over what lawful speech is considered acceptable online. Many users complain about the opaque, arbitrary removal of their legitimate content and the lack of clear routes to appeal the takedown. Decisions on how to moderate content involve trade-offs with freedom of expression and absent regulation, these decisions are being made by companies without democratic oversight. The requirements on Category 1 platforms to ensure they have clear and accessible terms of service and user redress mechanisms are expected to minimise freedom of expression impacts currently inherent under the *status quo.* These platforms will not be able to arbitrarily remove harmful content. They will need to be clear what content is acceptable on their services and enforce the rules consistently and users will have access to effective mechanisms to appeal content that is removed without good reason. They will also be required to have regard for freedom of expression when fulfilling their safety duties.

337. In addition, it is clear that some individuals and groups do not engage online through fear of being the targets of online abuse. For example, an international survey of female journalists found 64% had experienced online abuse – death or rape threats, sexist comments, cyberstalking, account impersonation, and obscene messages.[249] 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. Additionally, in the 2017 Annual Bullying Survey,[250] of those that had been the victims of cyberbullying, 26% deleted their social media profiles and 24% stopped using social media altogether.[251] The framework takes an approach which benefits and protects all users. It will empower adults, including vulnerable users, to keep themselves safe online, and to enjoy their right to freedom of expression, reducing the risk of bullying or being attacked on the basis of their identity.

338. Given that the OSB addresses legal but harmful content and is likely to result in increased (or more effective) content moderation, a number of stakeholders have raised concerns relating to potential negative impacts on freedom of expression. Table 31 sets out some of the main concerns:

Table 39: Main stakeholder concerns around potential impacts on freedom of expression

The Bill forces platforms to delete legal but harmful content which will have a negative impact on freedom of expression	The OSB does not require platforms to remove any content that is legal but harmful to adults. Category 1 services must only set out their policies with regard to content that is legal but harmful, and must enforce these policies consistently. The OSB also contains protections for freedom of expression that require platforms to balance their duties to protect users from harm with consideration of the importance of free expression. Similar protections apply to content of democratic importance and journalistic content on Category 1 services.
The Bill's definitions of harmful content are too vague and could result in the over removal of content.	The OSB requires platforms to take action against illegal content on their service where it is an existing UK offence that gives rise to harm to an individual. The OSB requires platforms that are likely to be accessed by children to

[249] IFJ global survey shows massive impact of online abuse on women journalists - IFJ (2018)
[250] The 2017 survey is used here as it included a deep dive on cyberbullying specifically. It is not possible to disaggregate the impacts of traditional bullying and cyberbullying in more recent editions.
[251] Annual Bullying Survey 2017 (Ditch the Label, 2017)

	protect children on their service. Companies will need to take action to protect children against content that poses a material risk of it having - or indirectly having - a significant adverse physical or psychological impact on a child of ordinary sensibilities. Action may include restricting children's access to that content (rather than removing such content entirely). The OSB also requires Category 1 services to set out their policies in relation to content or activity that is considered harmful to adults, although as set out above, does not require them to remove such content. Companies must set clear terms of service for content that poses a material risk of it having - or indirectly having - a significant adverse physical or psychological impact on an adult of ordinary sensibilities.
Large fines will cause platforms to overreact and remove content that is legal.	Platforms will have no legal duty to remove legal content and will have duties to protect freedom of expression when carrying out the safety duties. Ofcom enforcement will apply equally to all duties in the OSB, including those regarding freedom of expression, such that the OSB ensures against platforms 'overreacting'. Ofcom has the option of imposing substantial fines to encourage compliance (and to reflect instances of serious user harm). However, the cap is a ceiling. Ofcom will only impose fines proportionate and appropriate to the breach that has occurred. Escalating enforcement sanctions will avoid incentivising content takedown, with judicial oversight required for the most severe sanctions.
The journalistic content and content of democratic importance protections provide some people with a higher level of protection, creating a two-tier system online.	The protections for journalistic content and content of democratic importance focus on the content, not the actor. Anyone who posts this content will benefit from the protections. The protections themselves are important to ensure democratic debate is protected online and users have access to quality journalism.
The Bill provides Ofcom with too much power and allows it to regulate free speech.	Ofcom is accountable to Parliament in how it exercises its functions. It is required to present its annual report and accounts before both houses and to appear before Select Committees to answer questions about its regulatory operations. Parliament will have a role in approving a number of aspects of the regulatory framework through its scrutiny of both the primary and secondary legislation. The Government has ensured that, in addition to judicial review through the High Court, there is an accessible and affordable alternative means of appealing the regulator's decisions. The OSB will establish the Upper Tribunal as the alternative route to appeal Ofcom's decisions. As a public body, Ofcom is required to protect freedom of expression when carrying out its duties. This means that Ofcom will not be able to put in place any measures that restrict users' freedom of expression unless it is lawful, necessary and proportionate to do so.
The Bill sets a worrying international precedent for oppressive regimes.	There is a vast difference between UK internet safety regulations and the actions taken by some governments that use censorship and suppression to stifle political dissent. The UK supports freedom of expression as both a fundamental right in itself and as an essential element of a full range of human rights. That is why freedom of expression is an inherent principle in the design of the OSB. The UK approach does not seek to change the open and free principles upon which the Internet is based and the Government will continue to be an active voice in multilateral discussions

	to maintain a free, open and secure Internet.
Removing the right of an individual to remain anonymous online will limit freedom of expression	Option 1 does not remove the right of an individual to remain anonymous online. The government agrees that placing restrictions on anonymity online could disproportionately impact users without official ID (such as refugees, migrants and those from lower socio-economic backgrounds), or those who are reliant on ID from family members, and would experience a serious restriction of their online experience, freedom of expression and rights. The OSB requires platforms to provide optional user verification and allow users to determine the content and kinds of users they interact with online. Under the user verification duty, users are still able to be completely or pseudo anonymous online, verifying their identity only if they wish to do so. There is a trade-off between preventing online abuse and maintaining freedom of expression and optional user verification is the right balance.

339. While Option 1 is expected to enhance certain aspects of freedom of expression online it also includes a number of protections - both in the design and specific safeguards - to ensure any negative impacts are mitigated. In its comparative analysis of online harm regulations in eight jurisdictions,[252] Linklaters identified that regimes can broadly be divided into those that focus on individual pieces of content and those that instead focus on the 'systems and processes' that platforms must have in place. The online safety framework is a 'system and processes' approach which means it does not set specific obligations on platforms to remove content within a certain time limit. For example, Germany's NetzDG requires platforms to remove illegal content within 24 hours. This approach was copied in France's "Avia Law" (see international context section) but was deemed by the French Constitutional Court to be incompatible with the right to freedom of expression, given the risk that platforms would "over-block" to avoid enforcement action. By focussing on systems and processes, there will be less of an incentive for platforms to take too cautious an approach and restrict freedom of expression online.

340. Finally, Option 1 includes a number of built-in safeguards to protect freedom of expression, these include:

- All in-scope companies must have regard to the importance of protecting freedom of expression when implementing safety policies and procedures. This mitigates the risk that companies adopt highly restrictive measures to fulfil their statutory duties.
- Codes of practice will set out steps relating to companies' processes for considering the balance between user safety and freedom of expression when introducing content moderation or other online safety measures. Companies will be assessed as having fulfilled their duty to have regard to the importance of protecting freedom of expression if they follow these steps.
- Companies must have systems and processes in place to enable users to complain and seek redress if their content has been unfairly removed or restricted, or if they have been suspended or banned from a service.
- 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.
- Super-complaints will allow organisations to lodge concerns on behalf of users, which can include concerns about limits on freedom of expression.
- [placeholder: to note policy still under discussion] Companies must make clear in their terms of service that users have a right to bring a claim in court for breach of contract where their content is removed in breach of that company's terms of service.
- Category 1 services will need to assess the impact on freedom of expression and privacy both when deciding on safety policies and after they have adopted those policies. They will also need to demonstrate they have taken positive steps to mitigate this impact.

[252] Online harms a comparative analysis (Linklaters)

- Category 1 services are required to put in place clear policies about how they will protect users' access to content of democratic importance[253] when making content moderation decisions. Providers must take into account the importance of users' free expression in relation to content of this kind.
- Content of democratic importance will apply to content, not people. Therefore, content that supports or opposes government policy will be captured whether the creator of that content is a government minister or an individual political campaigner. This definition of democratic content does not, therefore, privilege politicians and/or specific political parties. For example, a service cannot provide a higher level of protection for left-wing views compared to right-wing ones.
- Users will be able to appeal to the platforms if they consider that the platform is not complying with its duties to protect content of democratic importance.
- [placeholder: to note policy still under discussion] Category 1 services are required to put in place clear policies about how they will protect the content of electoral and referendum candidates and campaigners when making content moderation decisions during the pre-election and pre-referendum periods. Protections must include an expedited complaints procedure for those candidates and campaigners to appeal against decisions companies have taken with regard to their content of democratic importance they have generated, shared or created.
- Category 1 services will be required to put in place clear policies to protect journalistic content[254] when making content moderation decisions. Protections must include an expedited complaints procedure for users who are the creators of such content (including recognised news publishers) to appeal against decisions companies have taken with regard to journalistic content they have generated, shared or created**.**
- Ofcom must fulfil its new functions in a way that protects users' rights to freedom of expression. There will be a robust appeals process against regulator decisions for anyone materially affected by a decision by the regulator.

341. The online safety framework limits platforms ability to arbitrarily remove lawful content, protects vulnerable users' right to express views online, and is designed to protect freedom of expression online. Based on the above qualitative assessment of freedom of expression implications, Option 1 is expected to enhance freedom of expression online rather than limit it.

Privacy impacts

342. There are a number of areas within the OSB that have the potential to result in privacy implications. For this reason, it includes strong privacy protections and Ofcom and the ICO will work together to ensure consideration of how personal data is processed as part of the duties.

343. 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. 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.

344. End-to-end encrypted services are in scope of the OSB and Ofcom will take steps to ensure that these services are meeting their obligations under the duties. The Government is supportive of strong encryption to protect user privacy, however, there are concerns that a move to end-to-end encrypted systems, when public safety issues are not taken into account, is eroding a number of existing online safety methodologies. This could have significant consequences for tech companies' ability to tackle grooming, sharing of CSA material, and other harmful or illegal behaviours on their platforms. Companies will need to regularly assess the risk of harm on their services, including the risks around end-to-end encryption. They would also need to assess the risks ahead of any significant design changes such as a move to end-to-end encryption. Service providers will then need to take reasonably practicable steps to mitigate the risks they identify.

[253] 'Content of democratic importance' is defined as content, including news publisher content, which is, or appears to be, intended to contribute to democratic political debate in the UK at a national or local level. This includes content promoting or opposing government policy and content promoting or opposing a political party.

[254] 'Journalistic content' will apply to content, including news publisher content, which is generated for the purpose of journalism and which is UK-linked

345. In addition, 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 CSA on public and, where proportionate, private channels.

346. Age assurance requirements in the Bill require the use of users' personal data. Under Option 1, platforms that host pornographic content will likely be required to verify the age of their users to prevent children from accessing this content. Concerns related to user privacy were raised under Part 3 of the DEA; however, the Online Safety Bill, combined with existing data protection law, will provide strong legal safeguards for user privacy. The Data Protection Act 2018 already provides a high standard of data protection legislation in the UK, which age verification providers will need to comply with and which has strong sanctions for malpractice. The Information Commissioner's Office recently published an opinion about the use of age assurance technologies and compliance with data protection law, which makes clear that providers using age verification must comply with data protection principles of transparency, fairness, lawfulness, accuracy, data minimisation and purpose limitation. The ICO also suggests companies use appropriately certified solutions. The Online Safety Bill will also place an explicit duty on providers to carry out privacy impact assessments. In addition, the ICO recently approved a new certification scheme for age assurance, through which companies can demonstrate their commitment to following the DPA 2018 when using age assurance technologies. DCMS is also leading further work to incentivise the development of further privacy-protecting solutions, this includes an international standard which DCMS is working on with industry, the British Standards Institute and the International Organization for Standardization. Furthermore, there is a growing range of solutions available on the market that minimise the amount of personal data users are required to share and provide platforms with an anonymised 'yes/no' answer to whether the user is over 18. Many pornography users are comfortable sharing data with pornographic websites, for example their credit card details, which is why there is a market for premium content and 'live cam sites'.

347. More broadly, all in-scope companies must have regard to the importance of protecting users from unwarranted infringements of privacy when implementing safety policies and procedures. Codes of practice will set out steps relating to companies' processes for considering the balance between user safety and privacy when introducing content moderation or other online safety measures. Companies will be assessed as having fulfilled their duty to have regard to the importance of protecting users from unwarranted infringements of privacy if they follow these steps.

348. The Government has consulted a range of stakeholders on end-to-end encryption and privacy implications more generally. This included businesses, Parliament, charities and privacy-focussed organisations. Proposals have included banning end-to-end encryption or greater consequences for companies when illegal material such as CSA is found on their systems. However, there are also a number of privacy-focussed organisations who are concerned about how the regulatory framework will impact on user privacy.

349. Recognising the potential risk of an impact to users' privacy, the preferred option includes a number of protections for privacy and mitigations against potential privacy implications.

Table 40: Overview of mitigations against privacy impacts

Mitigation	Description
Platforms must take steps to protect against unwarranted infringements of privacy when carrying out their safety duties.	The OSB includes specific provisions that require service providers to protect against unwarranted infringements of privacy in the fulfilment of their safety duties and reporting and redress duties. This is to ensure service providers do not, for example: ○ actively monitor more content than is necessary for safety features to function ○ track the activity of children more than it is needed to ensure they are only served appropriate content
Platforms must take steps to enable users and other affected persons to report	This includes a platform protecting against unwarranted infringements of privacy. If a complaint is upheld, platforms are expected to seek to rectify the issue by making changes to their policies and procedures to

concerns about a platform's non-compliance with their duties.	bring themselves into compliance.
Ofcom must put together codes of practice that explain how platforms can comply with their duties. Ofcom must consult on these codes. Platforms must comply with these codes or take alternative steps that achieve the same ends.	Companies will be expected to be clear about how they can protect against unwarranted infringements of privacy when fulfilling their duties. Throughout the codes, Ofcom would set out how platforms can fulfil each of their safety and redress duties in such a way that protects users from unwarranted infringements of privacy. For example, Ofcom may refer to service providers' existing duties under data protection law and include specific steps that service providers can take to guard against privacy infringements when implementing safety systems and processes.
Stringent safeguards relating to Ofcom requiring the use of technology	This will only be used as a last resort where alternative measures are not working and will be subject to stringent safeguards to protect users' rights. 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.
Ofcom can enforce the privacy duties on platforms.	Ofcom will be able to enforce the privacy duties to hold platforms to account.
Collaboration between Ofcom and the ICO	Ofcom will work closely with the ICO when developing codes so that platforms are clear about what they have to do to comply with both regimes and inefficiencies are reduced. Each regulator will provide guidance for platforms and users about how the regimes interact. Operationally, both regulators will work closely together to resolve issues as they arise, for example, flagging complaints that are relevant to the other regulator and passing on complaints that are for the other regulator to investigate.
As part of the Government's related non-regulatory interventions, the Safety Tech Challenge Fund awarded five organisations funding to prototype and evaluate innovative ways in which sexually explicit images or videos of children can be detected and addressed within end-to-end encrypted environments, while ensuring user privacy is respected.	In addition to the measures within the OSB, the Government is supporting the development of technological solutions to mitigate against the public safety challenges arising from the use of end-to-end encryption.
The Secretary of State must review how effective the regulatory framework is at protecting users from unwarranted infringements of privacy.	Given the novelty and complexity of the regime, monitoring work and the post-implementation review will consider freedom of expression and privacy implications

350. There are inevitably trade-offs between user safety and technologies such as end-to-end encryption which seek to increase user privacy. Option 1 recognises this and includes strong protections for user privacy online. Alongside Ofcom, the Government will continue to consult with stakeholders

through implementation of the regime and beyond to ensure any potential privacy implications are minimised.

Calculations

351. 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. Under Option 1, the NPSV is estimated to be -£2,507m (central) with an EANDCB of £250.6 million (central). This NPSV and EANDCB are illustrative only and are based on a 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.

352. Estimated costs have increased since the consultation stage IA, with the NPSV increasing from -£2,118 million to -£2,507 million and the illustrative EANDCB increasing from £205.8 million to £250.6 million. The main factors include:

- **Content moderation costs**: while there has been no change to the methodology or analytical assumptions here, increases in the number of businesses in scope (to reflect an implementation date of 2024) and inclusion of the latest revenue data has led to increases in this cost.
- **Additional costs**: the inclusion of the pornography provision, the fraudulent advertising duty, and duties related to user verification and empowerment have added to the NPSV by approximately £190 million.
- **Consultation evidence**: familiarisation costs and transition costs have also been increased to reflect input from stakeholders in areas such as the potential need for legal advice and representing SMB staff time with Chief Executive wage estimates instead of estimates for regulatory professionals.

353. Given that specific business requirements are unknown at this stage, the EANDCB calculated here remains 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).[255]

Key assumption sensitivity analysis

354. This IA presents low, central and high estimates throughout to reflect the range of potential impact scenarios on business. Additionally, this section brings attention to the key assumptions used in the production of the estimates and varies them in isolation to outline how sensitive the central estimate is to each.

Table 41: Risks and sensitivity analysis

Assumption	Lower bound sensitivity	Central	Upper bound sensitivity
Number of businesses in scope of the regulation	19,438	25,051	139,387
Illustrative EANDCB	£174.2m	£250.6m	£322.7m
Illustrative NPSV	-£1,849m	-£2,507m	-£3,128m
Evidence: The central estimate is based on a random stratified sample of 500 businesses from the IDBR. This is varied to reflect the number of organisations identified by RR prior to supplementing with additional known types of organisations likely to be in scope and the upper bound of RR estimates.			
Assumption	**Lower bound sensitivity**	**Central**	**Upper bound sensitivity**

[255] RPC Case Histories: assessment and scoring of primary legislation measures (2019)

Cost to Category 1 organisations of additional content moderation	1% of turnover	7.5% of turnover	15% of turnover
Illustrative EANDCB	£187.1m	£250.6m	£323.8m
Illustrative NPSV	-£1,961m	-£2,507m	-£3,138m
Evidence: The central estimate is based on the midpoint of estimates provided by in-scope businesses during the interview phase of RR research project. This is varied to reflect the range of responses.			
Assumption	**Lower bound sensitivity**	**Central**	**Upper bound sensitivity**
Cost to Category 2 organisations of additional content moderation	0.3% of turnover	1.9% of turnover	3.8% of turnover
Illustrative EANDCB	£126.4m	£250.6m	£402.3m
Illustrative NPSV	-£1,438m	-£2,507m	-£3,813m
Evidence: 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 is varied to reflect the range of responses.			
Assumption	**Low**	**Central**	**High**
Growth rate of online harms under the baseline	1.3%	3%	5.1%
Break-even point	2.4%	2.1%	1.8%
Evidence: The central estimate is in line with growth in the amount of hours spent online. This is varied to reflect a realistic range in terms of potential growth rates.			
Assumption	**Low**	**Central**	**High**
Percentage of fraud within scope of the OSB	30%	45%	60%
Break-even point	2.2%	2.1%	1.9%
Evidence: This reflects an illustrative range of between 30%-60% (central: 45%) of the relevant fraud offences likely to be in scope of the OSB.			

Small and micro business assessment

Justification for non-exemption

355. As explained in guidance from the RPC, the default position is to exempt SMBs fully from the requirements of new regulatory measures.[256] However, the evidence suggests that the objectives of the regulations would be compromised by exempting SMBs.

[256] Small and Micro Business Assessments: guidance for departments, with case history examples - RPC (2019).

356. First, there is evidence of harm occurring on smaller platforms. In particular, law enforcement and NGOs regularly see CSA offenders active on small chat forums, live streaming apps and file sharing/hosting services. The IWF notes that online harm exists 'in vast quantities' on smaller platforms.[257] 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.

357. 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.[258] 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. 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.

Impacts on SMBs

358. This IA estimates that there are around 21,500 SMBs within scope of the OSB. The in-scope SMBs are estimated to fall within the following risk categories:

Table 42: Estimated number of SMBs in each risk tier (rounded to the nearest ten)

	Low risk	Mid risk	High risk	Category 1
Micro	10,090	10,090	60	0
Small	580	580	60	0

359. Tables 43 to 44 outline the costs that SMBs are expected to incur as a result of the regulations (with medium and large businesses included for comparison):

Table 43: SMB transition costs excluding additional user reporting costs (2019 prices)

	Low risk	Mid risk	High risk
Micro	£610	£635	£786
Small	£827	£878	£1,080
Medium	£1,044	£1,094	£1,346
Large	£1,604	£1,654	£1,957

Table 43 represents the per business transition costs. It does not reflect costs to the 10% of businesses in the central estimate that are expected to incur higher costs as a result of not currently enabling user reporting.

Table 44: SMB transition costs including additional user reporting costs (2019 prices)

	Low risk	Mid risk	High risk
Micro	£1,545	£1,570	£1,721
Small	£1,763	£1,813	£2,015
Medium	£1,979	£2,029	£2,282
Large	£2,539	£2,589	£2,892

Table 44 represents the per business transition costs. It reflects the cost to 10% of businesses in the central estimate that are expected to incur higher costs as a result of not currently enabling user reporting.

257 IWF Online Harms White Paper Response (2021)

258 UK launch of tech against terrorism at Chatham House - Tech Against Terrorism (2017).

360. As Tables 43 to 44 illustrate, the largest per business transition costs are expected to fall on medium and large businesses who are better placed to absorb them. While costs are expected to be higher for medium and large businesses in absolute terms, small and micro businesses that do not currently allow users to report harm are expected to incur comparable costs when considered in relative terms. Allowing users to report harm is fundamental to the success of the OSB and to keeping users safe online and therefore, the Government considers these costs to be proportionate.

361. Both the Government and Ofcom will work with small and micro businesses through implementation to ensure transition costs are minimised through for example, clear and accessible codes and guidance and proportionate expectations based on the size of business and risk of harm.

Table 45: SMB compliance costs excluding additional content moderation and risk assessment costs (2019 prices)

	Low risk	Mid risk	High risk	Category 1
Micro	£114	£114	£114	£0
Small	£114	£114	£114	£0
Medium	£389	£389	£389	£0
Large	£779	£779	£779	£0.4 million

Table 45 represents the per business compliance costs. It does not reflect the cost to the 2.5% of businesses in the central estimate that are expected to incur higher costs as a result of not currently assessing risks. It also does not reflect the 10% of larger mid-risk firms and the 25% of high-risk firms expected to conduct additional content moderation.

Table 46: SMB compliance costs including additional content moderation and risk assessment costs (2019 prices)

	Low risk	Mid risk	High risk	Category 1
Micro	£5,959	£5,959	£9,048	£0
Small	£5,959	£5,959	£57,807	£0
Medium	£6,234	£0.3m	£0.3m	£0
Large	£6,623	£4.0m	£4.0m	£16.4m

Table 46 represents the per business compliance costs. It reflects both the cost to the 2.5% of businesses in the central estimate that are expected to incur higher costs as a result of not currently assessing risks and the 10% of larger mid-risk firms and the 25% of high-risk firms expected to conduct additional content moderation.

362. While per business costs are expected to be higher for medium and large businesses, it is important to consider the possibility that some in-scope SMBs will have limited resources for compliance. To minimise burdens on SMBs, it will be vital for Ofcom to work with businesses and to ensure both requirements and enforcement are proportionate to the risk of harm and resources available to businesses. Proportionality in the context of effective safety measures must be balanced against the risk of harmful content being displaced to smaller and less well-equipped platforms. The government and Ofcom will work with SMBs to ensure that steps taken are effective in both reducing harms and minimising compliance costs. The government's Safety by Design framework and guidance is targeted at SMBs to help them design in user-safety to their online services and products from the start thereby minimising compliance costs.

363. The pornography provision is estimated to bring into scope an additional 11 SMBs (10 micro businesses and one small business),[259] made up of high risk UK-based pornography providers. These businesses will only incur costs associated with preventing children from accessing pornography. This impact assessment was only able to estimate illustrative site costs and a total economic cost of the pornography provision and it is not possible to determine the per business cost on these 11 SMBs.

[259] Based on business demographics within creative industries - DCMS Sectors Economic Estimates 2019: Business Demographics (DCMS)

Ofcom - through its assessment of codes and regulator guidance - will further consider potential impacts on these businesses.

Unregulated SMBs

364. While the fraud advertising duty only applies to Category 1 and 2A platforms (costs reflected in Table 19 above) it is also expected to result in costs to a significant number of out-of-scope SMBs that advertise on in-scope platforms. Costs here are expected to occur as a result of providing information to support anti-fraud checks. This impact assessment estimates that approximately 3.1 million micro businesses and 0.2 million small businesses will incur some costs in the first year. The two tables below outline the per business cost to SMBs expected to undergo standard and enhanced CDD which is made up of staff time to provide necessary information:

Table 47: Fraudulent advertising duty per business costs (standard CDD)

	Low risk	Mid risk	High risk
Micro	£16	£16	£16
Small	£7	£7	£7

Table 48: Fraudulent advertising duty per business costs (standard CDD)

	Low risk	Mid risk	High risk
Micro	£36	£36	£36
Small	£16	£16	£16

365. It should be noted that while a significant number of SMBs are expected to undergo CDD, these costs are one-off and once a business is verified to advertise on Category 1 and 2A platforms, they are not expected to incur any additional costs in the appraisal period. 95% of these businesses are expected to undergo standard CDD with 5% incurring costs associated with enhanced CDD.

366. 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 49: Total costs for each size band

	Total costs (10 year PV)	Number of businesses (to nearest ten)	Percentage of total costs	Percentage of in-scope businesses
Micro	£66.0 million	20,250	3.0%	80.8%
Small	£10.2 million	1,220	0.5%	4.9%
Medium	£608.7million	2,910	27.2%	11.6%
Large	£1,552.9million	680	69.4%	2.7%

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.

Findings from SMB engagement

367. The Government has engaged extensively with industry including with SMBs since the OHWP and more recently during pre-legislative scrutiny. SMBs (either themselves or through trade and industry associations) noted the following key concerns relating to ensuring that the OSB does not disproportionately affect smaller platforms:

Table 50: SMB concerns and mitigations

Potential impacts on competition: the need to ensure that innovative and smaller companies are not disproportionately negatively impacted. Large in scope companies are more likely to design products already in line with regulatory requirements.	Ofcom has a proven track record of balancing robust consumer protection with the need to ensure the regulatory environment is conducive to growth and innovation. Under Option 1, Ofcom will have a legal duty to assess the impact on SMBs and have regard to innovation in production of its codes.
SMB awareness: the need to reach out to SMBs to ensure they understand their obligations and to reduce the cost of familiarisation.	Ofcom and the Government will work together to engage SMBs throughout implementation and ensure obligations are clear and aimed at SMBs.
Technology requirements: the need to ensure a balance between mandating technology and ensuring SMBs are not required to employ technology which they cannot afford.	The Government will only mandate specific technologies in very limited circumstances such as to identify and remove illegal terrorist content or CSA content and only where this is the only effective, proportionate and necessary action available, and the regulator is confident that the tools available are highly accurate
Clear codes and guidance: the need for clear and easy to understand codes and guidance. The majority of SMBs do not have teams of regulatory compliance staff and prefer things such as checklists.	Guidance and codes produced by Ofcom will be clear, accessible and easy to understand. It will also ensure guidance is aimed specifically at SMBs.
Transparency reporting: the need to ensure thresholds are set at such a point to avoid the unnecessary inclusion of SMBs within this requirement.	Thresholds for designation as Category 1, 2A and 2B will be set out in secondary legislation. However, given the thresholds likely to focus on reach and risk of harm, it is expected that transparency reporting will only be required of the largest and highest risk services.
Non-prescriptive guidelines for risk assessments and transparency reports: prescriptive requirements related to the way in which platforms assess risk and report on harm is likely to disproportionately impact SMBs.	While certain information will be required in both a platform's assessment of risk and reporting of harm, on the whole the information requested and the systems and processes used by platforms will vary greatly. Ofcom will consult SMBs to ensure guidelines are not overly prescriptive.
Alignment with existing global regulations: the need to avoid creating unnecessary burdens on SMBs and ensure requirements align with other countries' regulations.	The Government and Ofcom is continuing to assess potential areas of alignment in terms of compliance activities, and are 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 harm.
Proportionality: the need to ensure the principle of proportionality through implementation of the legislation.	Proportionality is at the heart of Option 1 and Ofcom will work closely with affected SMBs to ensure requirements are feasible and proportionate.
Continued engagement with SMBs: the need for the Government and Ofcom to continue to engage with SMBs.	Engagement with SMBs is ongoing and will continue throughout implementation of the regime.

368. SMB concerns raised during engagement have been instrumental in the design of Option 1 and the government's commitment to proportionality. Ofcom will continue to engage SMBs on future codes in an attempt to ensure impacts are proportionate to both the risk of harm and a platform's resources.

SMB mitigations

369. This section sets out how the potential mitigations for SMBs identified by the RPC have been considered.[260]

Table 51: SMB mitigations

Potential mitigations (as suggested by the RPC)	How they have been considered in the OSB
Differentiated regulatory approach and requirements, which will likely apply to the majority of small businesses	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 *all* users. Additionally, only Category 1, 2A, and 2B businesses will be required to publish transparency reports. We expect a small number of only large businesses to be designated..
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)	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 UGC. Many of these will be SMBs. Enforcement measures will begin with confirmation decisions 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.
More discretion for smaller businesses to meet regulatory requirements* (e.g. extended transition period or temporary exemption)	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. Laying of codes will undergo consultation and IAs and will be staggered allowing time for SMBs to comply with individual codes, as opposed to a specific date in which the whole regime comes into force at once.
Simpler and clearer guidance on how to comply. More compliance support for small businesses from the Government and regulators	As well as the requirement to be consistent with the principle of risk-based and proportionate action, Ofcom will also be required to have regard to the need to: ○ ensure all businesses are able to understand and fulfil their responsibilities and

[260] Checklist tool for a high-quality SaMBA - RPC (August 2019)

	○ cater for all businesses whatever their risk level and capacity (for example by providing support to start-ups and SMBs, drawing on best practice in other sectors). 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. 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. 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.
Stronger culture of transparency and learning*	Ofcom is 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. It will have a dedicated digital, data and innovation function to lead these efforts. Ofcom has 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 collaborative methods for policy and implementation, and focus on inclusion of a broad range of stakeholders. 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.
Different requirements for different sizes of businesses	As mentioned above, not all businesses will be expected to respond to all categories of harm: many, and most SMBs, will only be required to respond to illegal harm 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).

	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.
Financial aid (e.g. reimbursement of compliance costs)	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. 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. The regulator should not be in a position to reimburse businesses or not be able to cover any regulatory costs.
Opt-in and voluntary solutions	Voluntary approaches have been tested in the sector but have not been successful (see rationale for intervention).

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?

370. The OSB will apply to any in-scope service provided to UK users regardless of where the service is based. The scope of the framework's core duties is functionality based, i.e. it is both good/service and sector agnostic. It is difficult in the context of online platforms and online harm 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 be a minor part of the online presence of a business which attains revenue from an unrelated good or service.

Where UGC/P2P interaction is the main offering

371. The UK is an important market for many of the most affected types of organisations, for example:

- **Social media and search engines**: social media businesses' main offering to users and advertisers is the UGC and P2P interaction[261]. As a result, Facebook accounts for over 50% of the display advertising market and with regard to search engines, Google controls 90% of the search advertising market.[262] The UK is the 13th largest market in terms of user base for Facebook[263], 8th for Instagram[264], 5th for Twitter[265], 4th for Snapchat[266], and 3rd for Pinterest. In

[261] 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.
[262] Online platforms and digital advertising - Market study final report (CMA, 2020)
[263] Leading Countries Based on Facebook Audience Size as of January 2021 - Statista
[264] Instagram Demographic Statistics: How many people use Instagram in 2021? - Brian Dean
[265] Leading Countries Based on Number of Twitter Users as of January 2021 - Statista
[266] Leading Countries Based on Snapchat Audience Size as of January 2021 - Statista

terms of traffic, the UK is responsible for 4.1% of traffic to Google (the third largest share behind the US and Brazil).[267]

- **Online marketplaces**: the UK is the third largest market for Amazon and second largest market for eBay representing 8% and 18% of total global traffic to the sites respectively. The UK market is of equal importance to smaller online marketplaces placing second for its share of global traffic for platforms such as Etsy and Wayfair[268].

372. Given the value of the UK market to these businesses, it is unlikely that the OSB would lead to a reduction in services offered to UK users (or UK advertisers). Platforms offering UGC and P2P services to UK users will not be at a significant disadvantage from those that operate elsewhere as the regulatory landscape for online platforms is evolving internationally. Similar regulations to the OSB are being developed or have already been implemented internationally - Germany's NetzDG was implemented in 2018 and the EU is developing their Digital Services Act. Other countries are also expected to follow suit.

373. Compliance costs associated with Option 1's fraudulent advertising duties, will make the process of advertising to UK consumers more expensive (if compliance costs are passed on to advertisers). However, many platforms are already deciding to implement anti-fraud measures and the cost of both conducting (for in scope platforms) and undergoing (for advertisers) a customer due diligence check is expected to be relatively modest for each individual business. This impact assessment does not expect anti-fraud measures to negatively impact the provision of advertising space or the decision of non-UK-based companies advertising to UK consumers.

374. Unlike a business providing an online service, if the cost of regulatory compliance becomes excessive in one country for a business manufacturing goods, 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 same for businesses in the digital markets whose main offering is UGC and P2P interaction. Due to the nature of digital markets, there are limited constraints on the provision of an online service, e.g. on the number of users/consumers. In a digital market the decision to provide a service is solely based on whether the benefits from providing the service in that country, for example, ad revenue or similar, exceed the cost of compliance. This IA estimates a relatively modest per business cost of compliance which is proportional to business risk, the likelihood of online platforms withdrawing their services from the UK in favour of providing their services elsewhere as a result of the proposed regulation is minimal.

375. For services currently offered to UK users only, who may in the future, look to enter other markets, this IA does 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.[269] Moreover in Ofcom's Online Nation 2021 report, 61% of respondents agreed with the statement: "Internet users must be protected from seeing inappropriate or offensive content".[270] Over the past year there has been increasing public pressure on platforms to take further steps in addressing online harm, particularly for categories of harm such as disinformation and online abuse. Given the general public's concerns about internet safety, compliance with the OSB could be considered to be a competitive advantage for UK providers[271] on the international stage.

Where UGC/P2P interactions are secondary

376. Some businesses - that may 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 OSB 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

[267] Regional distribution of desktop traffic to Google.com as of June 2021, by country (Statista, 2021)
[268] https://www.webretailer.com/b/online-marketplaces-uk/
[269] Internet Users' Experience of Potential Online Harms: Summary of Survey Research - Ofcom (2020)
[270] Online Nation 2021 report - Ofcom (2021)
[271] UK providers here refers to platforms providing services to UK users only.

risk' functionality exemption has removed a large proportion of these types of businesses from scope, for example, small hospitality, beauty and health businesses, where there is simply a comment function for reviews on their products. 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. This could result in a reduction in the quality of the customer experience when engaging with such businesses. 73% of customers find live chat the most satisfactory form of communication with a company[272].

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

377. 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 paving the way in this regulatory landscape with countries worldwide following suit. 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?

378. 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.

379. 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 sticky investments in the UK market. They also have large investments in value-add employment (that is, 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 harm, the Government would therefore expect there to be a minimal impact upon their investment and business activity within the UK.

WTO Notification

380. 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.

Competition assessment

Competition in digital markets

381. In July 2020, the Competition Markets Authority (CMA) published the final report of its market study into online platforms and digital advertising.[273] The findings highlight a number of characteristics of digital markets that inhibit entry and expansion by rivals, undermining effective competition. These include network effects and economies of scale, the power of default placement (for example being assigned the default search engine on an internet browser),[274] unequal access to user data,[275] lack of

[272] https://99firms.com/blog/live-chat-statistics/#gref

[273] Online Platforms and Digital Advertising Market Study - final report - CMA (2020)

[274] In 2019, Google paid around £1.2 billion in return for default position in the UK, a majority of which was to Apple for being the default on the Safari browser. Such payments are one of the most significant factors inhibiting competition in the search engine market.

[275] Analysis of a trial run by Google in 2019, comparing the revenue publishers received from personalised advertising with revenue from non-personalised ads, suggests that UK publishers earned 70% less revenue when

transparency (in terms of decisions made by platforms), ecosystems of complementary products and services, and vertical integration as large digital platforms are present at multiple stages of supply chains. On this basis, there are significant barriers to competition present under the baseline.

382. In terms of online search engines, Google has persistently had a high and stable share of the general search market, with a share of supply between 89% and 93% over the last three years.[276] In June 2021 Bing and Yahoo Search had the next two highest shares at 5.6% and 2.7% respectively.[277] Similarly to Google, Bing holds extensive default positions through Microsoft's agreements with Windows PC manufacturers. These extensive default positions limit the expansion of rival search engines through limiting their accessibility to consumers, preventing new entrants from developing into strong competitors. Existing smaller platforms in the market are often syndication partners of Google or Bing, relying on the larger search engines for their search results and adverts[278]. These businesses seek to attract customers through other means, for example DuckDuckGo's unique selling point is its focus on privacy. In search advertising, Google is by far the largest player with the CMA stating that potential rivals can no longer compete on equal terms.[279]

383. In the social media market, the extent of competition between platforms is dependent on the degree to which users consider them as substitutes. Social media platforms offer similar types of functionality although they are differentiated based on particular consumer needs based on the type of communication and content consumption provided. Despite this, evidence indicates that Facebook has a significant and enduring market power in social media. Between July 2015 and February 2020, Facebook had a share of 54% of user time spent in social media[280].

384. When looking at the market for VSPs, based on analysis of the number of users watching videos on platforms and the number of video views on such platforms, the Herfindahl–Hirschman Index (HHI)[281] is greater than 2,500 indicating a highly concentrated sector, this has been the case since 2017. Only a limited number of platforms have entered the sector and achieved scale in recent years. Consumers use a limited number of platforms to view videos online, 38% of people said that the main reason they use a platform to watch videos is because it was the first platform they used, suggesting a degree of consumer inertia.[282]

385. While it is important to acknowledge potential competition impacts of Option 1, many of the main online markets are highly uncompetitive currently. Many of the requirements under Option 1 such as transparency reporting, user redress, and privacy protections may go some way to mitigating some of the current problems.

Potential impacts on competition of the OSB

386. While the rapid evidence assessment of Germany's NetzDG did not find any evidence that the policy had any impact on market competition, the proposals under Option 1 are not limited to large social media companies. Option 1 could potentially impact competition in the market if:

- compliance costs create – or are viewed by potential new entrants as - a barrier to entry; or
- costs fall disproportionality on SMBs, i.e. they are not able to absorb the costs (in unit terms) as easily as larger businesses; or
- compliance costs dissuade foreign investment and/or encourage UK based businesses to disinvest in the UK; or
- compliance with the legislation creates friction for users' consumption of online platforms.

---

they were unable to sell personalised ads. The inability of smaller platforms and publishers to access user data therefore creates a significant barrier to entry.

[276] Online Platforms and Digital Advertising Market Study - final report - CMA (2020)

[277] Worldwide market share of search engines - Statista (2010-2021)

[278] Online Platforms and Digital Advertising Market Study - final report - CMA (2020)

[279] Online Platforms and Digital Advertising Market Study - final report - CMA (2020)

[280] Online Platforms and Digital Advertising Market Study - final report - CMA (2020)

[281] The Herfindahl–Hirschman Index (HHI) is used to assess the level of concentration in a sector.

[282] Understanding how platforms with video sharing capabilities protect users from harmful content online - EY, DCMS (2021)

Will the measure indirectly or directly limit the range or number of suppliers?

387. 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. The growth of the UK digital economy outpaces that of most other sectors.[283] The fast-paced nature of this evolving market can result in platforms scaling rapidly, however, the financial benefits of the achieved scale can be delayed. Therefore, it is possible that a firm be deemed high-risk and not yet have the financial resources available to comply with the legislation. This could potentially result in realised costs of compliance forcing platforms out of the market. The proportionate enforcement expected of the regulator will be essential in minimising this impact.

388. For a low risk in scope micro-businesses, beyond familiarising themselves with the regulations, they 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'). The impact on such businesses is expected to be limited. 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.

389. Option 1 is expected to result in impacts on some out-of-scope SMBs through requirements under the fraudulent advertising duty. These SMBs - that participate in paid-for advertising on Category 1 and 2A platforms - will incur costs associated with providing necessary information to ensure they are legitimate businesses. Platforms like Google and Facebook do offer low friction advertising opportunities especially to small businesses with an estimated 63% of SMEs advertising this way.[284] These costs on out-of-scope SMBs are expected to be minimal, involving between 10-30 minutes of staff time only. It is still the case that Ofcom must engage with SMB advertisers as it develops codes of practice and ensure any compliance burdens are minimal.

Will the measure limit the ability of suppliers to compete?

390. 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.

391. 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 harm (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 depending on the requirements set out in future codes. Additionally, based on the intention of the policy, small or micro businesses are not expected to be designated as Category 1. It should be noted that the pornography provision ensures that all businesses regardless of size will be required to prevent children from accessing pornography. While this has the potential to result in burdens on SMB platforms that host pornography, the government considers the protection of children a core objective of the Bill. Further, given the nature of costs for age verification solutions (largely based on the number of checks), sites with a larger user base will pay more.[285]

Will the measure limit the suppliers' incentives to compete vigorously?

392. Regulation of online platforms 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,

[283] DCMS Sectors Economic Estimates 2019 (provisional) Gross Value Added - DCMS (2020)
[284] Powering Up: Helping UK SMEs unlock the value of digital advertising (IAB, 2020)
[285] The number of users is not a perfect proxy for platform size but they are related.

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. However, a thriving digital economy is at the heart of the government's vision for long-term economic growth. As such, the growth of digital markets will be supported by initiatives including the pro-competition regime for digital markets which will encourage competition in this sector.[286]

Will the measure limit the choices or information available to consumers?

393. 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 and purchase of online goods and services, driving greater competition between businesses to implement measures meeting regulatory and consumer demands for increased safety on online platforms.

Innovation test

Innovation in digital markets

394. Investment in primary technologies, including artificial intelligence and machine learning, provide an indication of the level of innovation within digital markets. In 2020, the UK had the second highest proportion of venture capital investment into these foundational technologies, accounting for 54% of total venture capital investment.[287] UK investment in the technology sector has significantly increased over recent years. Impact tech investment[288] in the UK has more than doubled since 2018, a 106% increase, in the same period the US saw only a 15% increase.[289] The UK is the third in the world for impact tech investment. These large-scale investments into the technology sector indicate high levels of innovation, providing the resources for innovation in digital markets.

395. The success of online marketplaces illustrates the value of eCommerce innovation. In the UK the largest marketplaces such as Amazon and eBay accommodate millions of customers with 407million visits and 298 visits in April 2021.[290] Marketplaces are able to provide a streamlined process of servicing and selling with access to an extensive global consumer base. Innovation in this market over the years has enhanced consumers' experiences. This includes the use of smart eCommerce which enables the supply of a customised list of recommendations based on consumer behaviour and history to provide a tailored online experience. AI has enabled marketplaces to provide 24/7 customer service through the use of chatbots, providing instant answers to simple questions.

396. There has also been considerable innovation in the gaming industry. In the past year Fortnite have hosted events including in-game concerts and movie trailer premiers.[291] It also anticipated that these innovations could develop into the creation of a digital metaverse, a virtual experience going beyond gaming to provide an array of media experiences.[292] The development of virtual reality has also augmented the gaming experience through providing an immersive gaming environment.

397. There is currently large-scale investment in research and development among the largest online platforms, indicating significant levels of innovation. In 2020 Amazon's R&D expenditure amounted to $42.7billion[293] (£33.3billion), similar levels of investment in the same period were seen among other

[286] A new pro-competition regime for digital markets - DCMS (2021)
[287] The Future UK Tech Built, Tech Nation Report 2021 - Tech Nation
[288] Impact tech investments are investments in technology made to generate positive social and environmental impacts alongside a financial return#.
[289] The Future UK Tech Built, Tech Nation Report 2021 - Tech Nation
[290] Leading online marketplaces in the United Kingdom as of April 2021, based on number of monthly visits - Statista
[291] The 10 most innovative social media companies of 2021 - Fast Company
[292] The 10 most innovative social media companies of 2021 - Fast Company
[293] Amazon Research and Development Expenses 2006-2021 - Macrotrends

platforms including Google $27.5billion (£21.4billion) and Facebook $18.4billion (£14.3billion).[294] In the past GAFAM companies have delivered breakthrough and disruptive innovations improving consumers' lives and creating jobs. Digital firms have also disrupted existing markets including the taxi and hotel industries.

398. However, there are concerns that the dominant firms that have emerged from the growth of digital markets are constraining further innovation.[295] As explored in the 'Competition Assessment', certain characteristics of digital markets inhibit the entry and expansion by rivals. In digital markets innovation requires access to data, users and fair returns.[296] The biggest digital platforms control some, if not all, of these elements. Established platforms have access and control over data,[297] a loyal consumer base,[298] and can exploit their market power to extract an unfair share of returns from successful innovation.[299] Therefore, the aim of Option 1 is to minimise any indirect impacts of regulatory compliance on wider innovation.

Potential impacts on innovation of the OSB

399. While sector agnostic in its design, Option 1 is risk-based and therefore, the majority of requirements will fall on businesses with websites offering high levels of UGC and P2P interaction functionalities, for example, social media and other digital technology businesses. These types of businesses are high-growth and highly profitable businesses, as such these companies invest considerably into research and development. The compliance requirements of this framework will therefore disproportionately fall on highly innovative sectors. However, these platforms are already investing substantially into user safety and it is therefore assumed that they do not necessarily see a trade-off between user safety and innovation.

400. 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.

401. 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
- Ofcom will have a legal duty to pay due regard to innovation in the exercise of all of its functions
- 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 categories of online harm (even if that changes).

[294] The average exchange rate (1 USD = 0.7798 GBP) in 2020 was used to present figures in GBP.
[295] Competition and Innovation in Digital Markets - BEIS (2020)
[296] Big Tech: how can we promote competition in digital platform markets? - Amelia Fletcher, Economics Observatory (2021)
[297] Big Tech: how can we promote competition in digital platform markets? - Amelia Fletcher, Economics Observatory (2021). Amazon, Apple, Facebook and Google are estimated to hold around 1.2 billion gigabytes of data between them.
[298] Understanding how platforms with video sharing capabilities protect users from harmful content online - EY, DCMS (2021) There is evidence to suggest that a degree of inertia exists among consumers in the VSP industry.
[299] Big Tech: how can we promote competition in digital platform markets? - Amelia Fletcher, Economics Observatory (2021)

- The approach taken will be technology neutral and therefore encompass future changes to how the architecture of the internet functions.
- Development of the online safety implementation measures which will focus on researching emerging harm 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 UGC. Many of these will be SMBs.

402. 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. 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.

Equalities impact

Statutory Equalities Duties	Completed
Proposals set out in the OSB 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 OSB, 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 theirs 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. **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.** **The Senior Responsible Officer has agreed with these findings.**	**Yes**

403. The Government has a legal obligation to consider the effects of policies on those with protected characteristics[300] under the Public Sector Equality Duty 2011 and the Equality Act 2010.

404. Overall, these proposals are expected to have a positive impact on users with protected characteristics. This is incorporated in the overarching aim of the policy; to make the internet a safe place for all users. Reducing online harm 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:

- A 2019 report by the Alan Turing Institute found that Black people and those of 'Other' ethnicities are far more likely to be targeted by, and exposed to, online abuse than White and Asian people, with 39% of Black people having observed hateful/cruel content online compared to 27% of White People.
- Between January and June 2021, Community Security Trust recorded 1,308 anti-Jewish hate incidents nationwide in the first half of this year. This is a 49% increase from the 875 incidents recorded in the first six months of 2020, and is the highest total CST has ever recorded in the first half of any year.[301]

[300] Age, disability, sex, gender reassignment, pregnancy and maternity, race, religion or belief and sexual orientation

[301] Antisemetic incidents January - June 2021 (CST, 2021)

- Users with disabilities have been forced to leave social media as a result of the abuse they had experienced online.[302]
- Women tend to be disproportionately affected by online offences like harassment, stalking, revenge pornography.

405. Vulnerable groups, particularly those with mental health problems, are at a much higher risk of falling victim to online scams. A 2021 report found that people who have experienced mental health problems are nearly three times more likely to have been a victim of an online scam than the rest of the population (23% of those with mental health problems were victims of online scams vs 8% of the wider population)[303].

406. It should also be acknowledged that there are potential distributional impacts as a result of the possibility of introducing age assurance processes. For example, those from disadvantaged backgrounds, including those with disabilities, are often less likely to hold form identification[304]. How much of an impact will be heavily dependent upon the level of verification and identification required.

407. The assessment of prospective equality impacts that Option 1 may have on those with protected characteristics is considered in regards to both direct and indirect discriminaton:

408. 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.

409. 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 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.
- **Improving media literacy**: 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 harm. The media literacy efforts incorporated in this policy may therefore be particularly important to enable these users to be able to keep themselves safe 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).
- **Requesting that redress mechanisms are easily accessible**: 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.

410. 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. This will be monitored post-commencement. However, the focus of the framework on systems and processes, as opposed to content, is intended to avoid this.

411. 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.

Devolution test

412. 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.

[302] House of Commons Petitions Committee report (2018)
[303] Caught in the Web - Online Scams and Mental Health (Money and Mental Health Policy Institute, 2020)
[304] Public Opinion Tracker 2021, Electoral Commission

413. 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 categories of harm in scope and media literacy.

414. While some of the categories of harm 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, Option 1 clarifies the responsibility of businesses to tackle this activity on their services.

415. 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 implementation.

Monitoring and evaluation

416. As part of developing a coherent and comprehensive evaluation framework, DCMS will be commissioning an independent evaluator or team of evaluators to assess the current evidence base and advise on appropriate governance structures, evaluation methodologies, and comment on future online harm research plans. On this basis, this section lays out the current proposed plans for monitoring and evaluation (M&E); however, these are subject to change as M&E work commences and the programme of research underpinning it progresses. The approach will be iterative and will draw on expertise from across government and external experts.

417. Any review will take a holistic approach and will evaluate the entirety of the online safety framework, including the OSB, the regulator, future codes of practice and secondary legislation and the impact on the digital sector more broadly. There are three main areas of evaluation:

- A review of the wider online safety framework;
- Evidence from the implementation of individual codes of practice; and
- An assessment of the government's overall online safety strategy, including the online safety implementation measures, such as media literacy initiatives, child and adult online safety initiatives, investments in the safety tech sector, and safety by design interventions.

Review clause

418. The OSB contains 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, 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.

Review governance

419. The review will be led by the DCMS Secretary of State and they will be responsible for delivering the PIR. However, given that the OSB is a joint policy, both DCMS and the Home Office will share responsibility and work closely with Ofcom to ensure appropriate monitoring and develop the underlying evidence base for online harm. In addition, the review will require input from:

- other government departments, such as Ministry of Justice (justice impacts), the Department for Business, Energy and Industrial Strategy (SMBs and business impacts more broadly), the Foreign Commonwealth and Development Office and the Government Communication Headquarters (evidence related to mis and disinformation), the Department for Education (media literacy and safety by design education initiatives), and others;
- regulated online platforms;
- civil society groups; and
- wider society.

420. DCMS, the Home Office and Ofcom are expected to set up an analytical evaluation working group to coordinate on baselining activities, the development of online harm metrics, and research pipelines. This work will be overseen by a senior evaluation steering group, again with representation from the three main stakeholders. Advice and potential involvement will also be sought from established Whitehall expert groups such as the cross-government evaluation group, the Cabinet Office's evaluation task force, and the independent RPC.

Review plans

421. At a high level, 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)

422. The majority of initial M&E work is focussed on baselining, developing key metrics, and ensuring that there is a coordinated programme of research to fill evidence gaps. A key strand in the evaluation work will be an assessment of the policy's stated objectives:

- **Objective 1 - to increase user safety online**: Work to understand and baseline the current prevalence of a number of key types of online harm is underway and this work is expected to result in clear and measurable indicators for both illegal and legal but harmful types of harm.
- **Objective 2 - to preserve and enhance freedom of speech online:** This will be monitored through the collection and reporting of transparency data, such as the amount of content removed/restored; and user satisfaction, such as measuring the effectiveness of redress mechanisms**.** Ofcom already conducts regular high-quality user attitude surveys which will be key indicators for this objective.
- **Objective 3 - to improve law enforcement's ability to tackle illegal content online**: This is expected to materialise as efficiency gains or cost savings for law enforcement. This can be measured using crime data and the level of understanding of the drivers of crime, including the specific role of activities in scope in facilitating crime. Addressing online crime will help drive economic growth and enable a stronger online business environment. Assessing the policy against this objective will require consultation with law enforcement and relevant enforcement authorities.
- **Objective 4 - to improve users' ability to keep themselves safe online:** This will draw on Ofcom's comprehensive programme of media literacy and internet use-related research and evaluation of media literacy initiatives.
- **Objective 5 - to improve societies understanding of the harm landscape**: This links closely to Objective 1 and the need to have a clear understanding of how harm manifests and how it can be measured. While important, success against this objective is more subjective than the others. However, the Government will draw on Ofcom's programme of user experience research to assess wider understanding of online harm and the joint programme of harm research planned.

Key measures and sources of data

423. The table below outlines some of the potential key measures for the OSB evaluation. As noted, these will largely depend on both government and Ofcom research programmes between now and implementation.

Table 52: Potential metrics for evaluation

Link to objective	Metric/measures	Sources (non exhaustive)
1 & 2 &3	Reductions in prevalence of priority and non-priority online harms on in-scope platforms	● Ofcom's adult and child media literacy trackers ● Online Harms Observatory (Alan Turing Institute)

		● Annual bullying survey ● Police recorded crime data (online flag) ● Counter disinformation monitoring (HMG) ● NFIB fraud reports ● Home Office and Ministry of Justice data
1 & 3	Reductions in the spread and flow of illegal and harmful content within and across platforms, such as mis- disinformation	● Platform transparency reports ● Counter disinformation monitoring (HMG) ● Ofcom's information gathering powers
1 & 3 & 4	Reductions in children's exposure to illegal content and age innapropriate content such as pornography	● Ofcom's child media literacy tracker ● Independent research on children's pornography use
3	Improvements in platform performance in areas such as responding to user reports, content moderation, and minimising the algorithmic spread of harmful content	● Ofcom's information gathering powers ● Platform transparency reporting ● Ofcom's compliance reporting
4 & 5	Increases in media literacy indicators, such as awareness of safety features, critical thinking skills, and interacting with other users safely online	● Ofcom's child media literacy tracker ● Independent media literacy research
2	Improvements in platforms' handling of content takedown challenges	● Ofcom's information gathering powers ● Platform transparency reporting
1 & 2 & 4	Improvements in users' experience of the online environment, with particular focus on children's experiences	● Ofcom's programme of user experience research ● DCMS VSP user experience research ● Ofcom's child media literacy tracker

424. The government will also consider a number of measures against areas such as competition, innovation and further measures related to privacy and freedom of expression (beyond content removal). Some example key evaluation questions include:

- How effective is Ofcom's regulation of the online harms framework?
- Was the online harms framework implemented successfully?
- Are there gaps in the regulatory framework and if so, where?
- Have regulatory requirements been communicated to businesses effectively?
- Did the OSB reduce the prevalence of priority and non-priority illegal content and activity?
- Did the OSB reduce the prevalence of priority and non-priority legal but harmful content and activity?
- Did the OSB reduce children's exposure to harmful content?
- To what extent can changes in the prevalence of harms be attributed to the OSB?
- Did the OSB affect freedom of expression online?
- Did the OSB produce or contribute to any unintended consequences?

- Are costs experienced by in-scope platforms in line with those estimated in previous impact assessments?
- What is the ratio of costs to benefits?
- Has the OSB affected competition and innovation in digital markets and if so, how?

425. Finally, planned M&E work is likely to be structured as three separate phases of activity

- **Phase one**: this phase involved both the development of a series of theories of change for each individual aspect of the policy. Theories of change are tools to enable causal link monitoring and allow the Government to identify assumptions made and fill evidence gaps. Additionally, this phase involved a piece of feasibility research[305] to develop an approach to the measurement of harm. This has also been complemented by Ofcom's ongoing surveys.
- **Phase two**: will take forward research to measure and baseline harm identified as priority illegal and priority 'legal but harmful' harm. This will complement existing qualitative research by Ofcom. Stage Two will also involve seconding in independent evaluation expertise to develop and finalise the M&E framework. The work conducted in Stage 2 will allow for the development of specific and measurable evaluation questions and metrics, such as 'the amount of children who have experienced cyberbullying in the last 12 months' tracked using results from DfE and Ofcom surveys alongside platform transparency reports.
- **Phase three**: will involve the ongoing collection of data and commissioning of research on an annual basis.

The online harm landscape is a fast-moving policy area and the OSB is ground-breaking and novel in its approach. The Government recognises the need for a comprehensive and adaptable M&E framework to ensure the policy achieves its objectives and minimises the potential for unintended consequences.

[305] Forthcoming – to be published 2022