HM Government

Online Harms White Paper

April 2019

CP 57

Online Harms White Paper

Presented to Parliament
by the Secretary of State for Digital, Culture, Media & Sport
and the Secretary of State for the Home Department
by Command of Her Majesty

April 2019

CP 57

© Crown copyright 2019

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at www.gov.uk/government/publications

Any enquiries regarding this publication should be sent to us at public.enquiries@homeoffice.gsi.gov.uk

ISBN 978-1-5286-1080-3
CCS0219683420 03/19

Printed on paper containing 75% recycled fibre content minimum

Printed in the UK by the APS Group on behalf of the Controller of Her Majesty's Stationery Office

Table of contents

Joint Ministerial foreword **3**

Executive summary **5**

PART 1: Introduction **11**
1: The challenge 11
2: The harms in scope 30

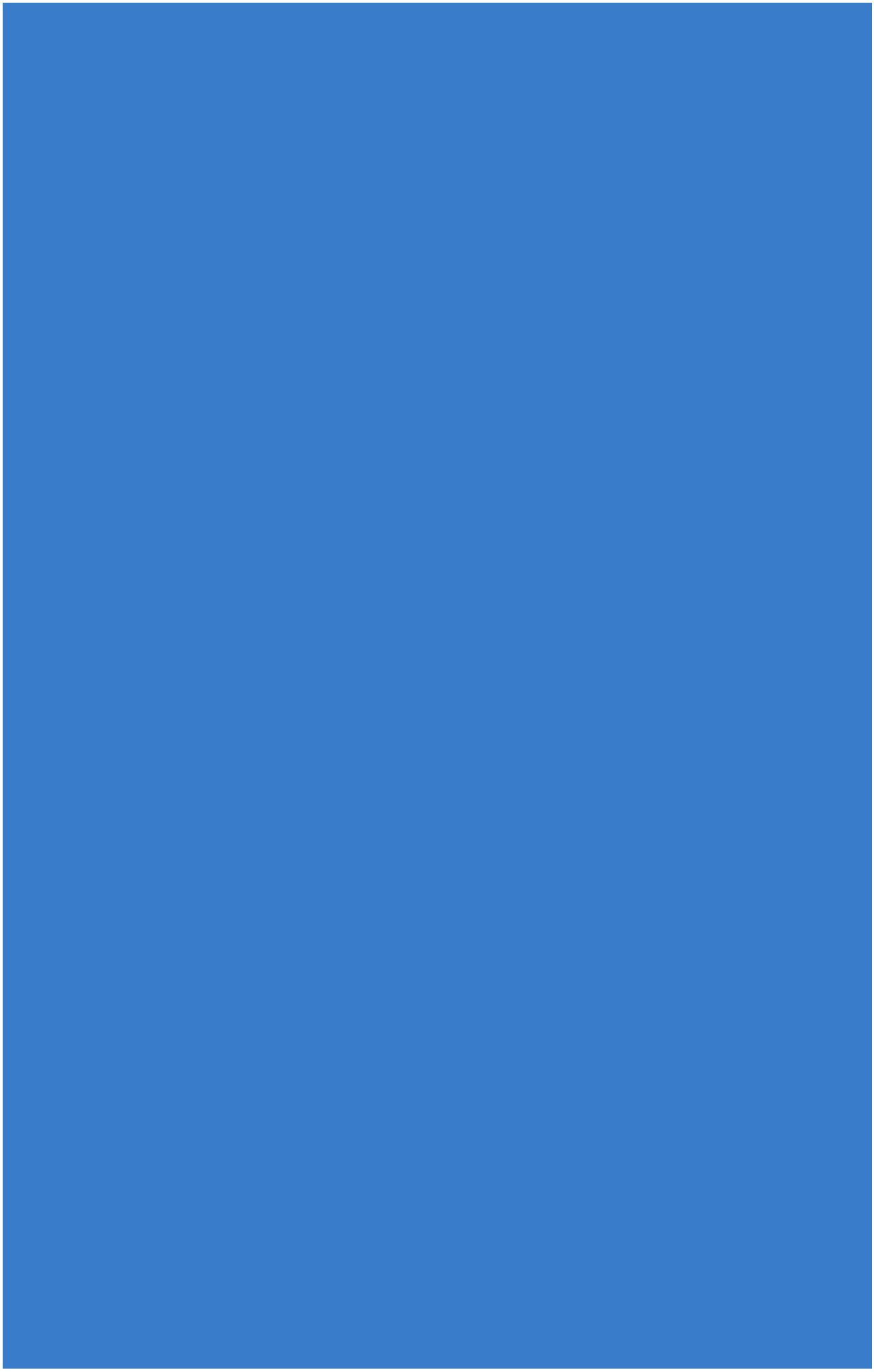
PART 2: Regulatory model **41**
3. A new regulatory framework 41
4: Companies in scope of the regulatory framework 49

PART 3: Regulation in practice **53**
5: A regulator for online safety 53
6: Enforcement 59
7. Fulfilling the duty of care 64

PART 4: Technology, education and awareness **77**
8: Technology as part of the solution 77
9. Empowering users 85

Part 5: Conclusion and next steps **95**
10: Conclusion and next steps 95

Annex A: How to respond to the consultation **97**



Joint Ministerial foreword

The internet is an integral part of everyday life for so many people. Nearly nine in ten UK adults and 99% of 12 to 15 year olds are online. As the internet continues to grow and transform our lives, often for the better, we should not ignore the very real harms which people face online every day.

In the wrong hands the internet can be used to spread terrorist and other illegal or harmful content, undermine civil discourse, and abuse or bully other people. Online harms are widespread and can have serious consequences.

Two thirds of adults in the UK are concerned about content online, and close to half say they have seen hateful content in the past year. The tragic recent events in New Zealand show just how quickly horrific terrorist and extremist content can spread online.

We cannot allow these harmful behaviours and content to undermine the significant benefits that the digital revolution can offer. While some companies have taken steps to improve safety on their platforms, progress has been too slow and inconsistent overall. If we surrender our online spaces to those who spread hate, abuse, fear and vitriolic content, then we will all lose.

So our challenge as a society is to help shape an internet that is open and vibrant but also protects its users from harm. The UK is committed to a free, open and secure internet, and will continue to protect freedom of expression online. We must also take decisive action to make people safer online.

This White Paper therefore puts forward ambitious plans for a new system of accountability and oversight for tech companies, moving far beyond self-regulation. A new regulatory framework for online safety will make clear companies' responsibilities to keep UK users, particularly children, safer online with the most robust action to counter illegal content and activity.

This will be overseen by an independent regulator which will set clear safety standards, backed up by reporting requirements and effective enforcement powers.

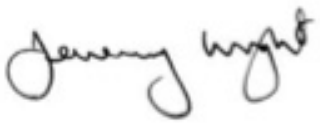
Although other countries have introduced regulation to address specific types of harm, this is the first attempt globally to address a comprehensive spectrum of online harms in a single and coherent way.

The UK's future prosperity will depend heavily on having a vibrant technology sector. Innovation and safety online are not mutually exclusive; through building trust in the digital economy and in new technologies, this White Paper will build a firmer foundation for this vital sector.

As a world-leader in emerging technologies and innovative regulation, the UK is well placed to seize these opportunities. We want technology itself to be part of the solution, and this White Paper proposes measures to boost the tech-safety sector in the UK, as well as measures to help users manage their safety online.

We believe the approach in this White Paper can lead towards new, global approaches for online safety that support our democratic values, and promote a free, open and secure internet; and we will work with other countries to build an international consensus behind it.

Online safety is a shared responsibility between companies, the government and users. We would encourage everyone to take part in the consultation that accompanies this White Paper, and work with us to make Britain the safest place in the world to be online.



Rt Hon Jeremy Wright MP
Secretary of State for Digital, Culture, Media and Sport

Rt Hon Sajid Javid MP
Home Secretary

Executive summary

1. The government wants the UK to be the safest place in the world to go online, and the best place to start and grow a digital business. Given the prevalence of illegal and harmful content online, and the level of public concern about online harms, not just in the UK but worldwide, we believe that the digital economy urgently needs a new regulatory framework to improve our citizens' safety online. This will rebuild public confidence and set clear expectations of companies, allowing our citizens to enjoy more safely the benefits that online services offer.

The problem

2. Illegal and unacceptable content and activity is widespread online, and UK users are concerned about what they see and experience on the internet. The prevalence of the most serious illegal content and activity, which threatens our national security or the physical safety of children, is unacceptable. Online platforms can be a tool for abuse and bullying, and they can be used to undermine our democratic values and debate. The impact of harmful content and activity can be particularly damaging for children, and there are growing concerns about the potential impact on their mental health and wellbeing.
3. Terrorist groups use the internet to spread propaganda designed to radicalise vulnerable people, and distribute material designed to aid or abet terrorist attacks. There are also examples of terrorists broadcasting attacks live on social media. Child sex offenders use the internet to view and share child sexual abuse material, groom children online, and even live stream the sexual abuse of children.
4. There is also a real danger that hostile actors use online disinformation to undermine our democratic values and principles. Social media platforms use algorithms which can lead to 'echo chambers' or 'filter bubbles', where a user is presented with only one type of content instead of seeing a range of voices and opinions. This can promote disinformation by ensuring that users do not see rebuttals or other sources that may disagree and can also mean that users perceive a story to be far more widely believed than it really is.
5. Rival criminal gangs use social media to promote gang culture and incite violence. This, alongside the illegal sale of weapons to young people online, is a contributing factor to senseless violence, such as knife crime, on British streets.
6. Other online behaviours or content, even if they may not be illegal in all circumstances, can also cause serious harm. The internet can be used to harass, bully or intimidate, especially people in vulnerable groups or in public life. Young adults or children may be exposed to harmful content that relates, for example, to self-harm or suicide. These

experiences can have serious psychological and emotional impact. There are also emerging challenges about designed addiction to some digital services and excessive screen time.

Our response

7. This White Paper sets out a programme of action to tackle content or activity that harms individual users, particularly children, or threatens our way of life in the UK, either by undermining national security, or by undermining our shared rights, responsibilities and opportunities to foster integration.
8. There is currently a range of regulatory and voluntary initiatives aimed at addressing these problems, but these have not gone far or fast enough, or been consistent enough between different companies, to keep UK users safe online.
9. Many of our international partners are also developing new regulatory approaches to tackle online harms, but none has yet established a regulatory framework that tackles this range of online harms. The UK will be the first to do this, leading international efforts by setting a coherent, proportionate and effective approach that reflects our commitment to a free, open and secure internet.
10. As a world-leader in emerging technologies and innovative regulation, the UK is well placed to seize these opportunities. We want technology itself to be part of the solution, and we propose measures to boost the tech-safety sector in the UK, as well as measures to help users manage their safety online.
11. The UK has established a reputation for global leadership in advancing shared efforts to improve online safety. Tackling harmful content and activity online is one part of the UK's wider ambition to develop rules and norms for the internet, including protecting personal data, supporting competition in digital markets and promoting responsible digital design.
12. Our vision is for:
    - A free, open and secure internet.
    - Freedom of expression online.
    - An online environment where companies take effective steps to keep their users safe, and where criminal, terrorist and hostile foreign state activity is not left to contaminate the online space.
    - Rules and norms for the internet that discourage harmful behaviour.
    - The UK as a thriving digital economy, with a prosperous ecosystem of companies developing innovation in online safety.
    - Citizens who understand the risks of online activity, challenge unacceptable behaviours and know how to access help if they experience harm online, with children receiving extra protection.
    - A global coalition of countries all taking coordinated steps to keep their citizens safe online.
    - Renewed public confidence and trust in online companies and services.

Clarity for companies

13. Increasing public concern about online harms has prompted calls for further action from governments and tech companies. In particular, as the power and influence of large companies has grown, and privately-run platforms have become akin to public spaces,

some of these companies now acknowledge their responsibility to be guided by norms and rules developed by democratic societies.

14. The new regulatory framework this White Paper describes will set clear standards to help companies ensure safety of users while protecting freedom of expression, especially in the context of harmful content or activity that may not cross the criminal threshold but can be particularly damaging to children or other vulnerable users. It will promote a culture of continuous improvement among companies, and encourage them to develop and share new technological solutions rather than complying with minimum requirements.
15. It will also provide clarity for the wide range of businesses of all sizes that are in scope of the new regulatory framework but whose services present much lower risks of harm, helping them to understand and fulfil their obligations in a proportionate manner.

A new regulatory framework for online safety

16. The government will establish a new statutory duty of care to make companies take more responsibility for the safety of their users and tackle harm caused by content or activity on their services.
17. Compliance with this duty of care will be overseen and enforced by an independent regulator.
18. All companies in scope of the regulatory framework will need to be able to show that they are fulfilling their duty of care. Relevant terms and conditions will be required to be sufficiently clear and accessible, including to children and other vulnerable users. The regulator will assess how effectively these terms are enforced as part of any regulatory action.
19. The regulator will have a suite of powers to take effective enforcement action against companies that have breached their statutory duty of care. This may include the powers to issue substantial fines and to impose liability on individual members of senior management.
20. Companies must fulfil the new legal duty. The regulator will set out how to do this in codes of practice. If companies want to fulfil this duty in a manner not set out in the codes, they will have to explain and justify to the regulator how their alternative approach will effectively deliver the same or greater level of impact.
21. Reflecting the threat to national security or the physical safety of children, the government will have the power to direct the regulator in relation to codes of practice on terrorist activity or child sexual exploitation and abuse (CSEA) online, and these codes must be signed off by the Home Secretary.
22. For codes of practice relating to illegal harms, including incitement of violence and the sale of illegal goods and services such as weapons, there will be a clear expectation that the regulator will work with law enforcement to ensure the codes adequately keep pace with the threat.
23. Developing a culture of transparency, trust and accountability will be a critical element of the new regulatory framework. The regulator will have the power to require annual transparency reports from companies in scope, outlining the prevalence of harmful content on their platforms and what countermeasures they are taking to address these. These reports will be published online by the regulator, so that users and parents can make informed decisions about internet use. The regulator will also have powers to require additional information, including about the impact of algorithms in selecting

content for users and to ensure that companies proactively report on both emerging and known harms.

24. The regulator will encourage and oversee the fulfilment of companies' existing commitments to improve the ability of independent researchers to access their data, subject to appropriate safeguards.

25. As part of the new duty of care, we will expect companies, where appropriate, to have effective and easy-to-access user complaints functions, which will be overseen by the regulator. Companies will need to respond to users' complaints within an appropriate timeframe and to take action consistent with the expectations set out in the regulatory framework.

26. We also recognise the importance of an independent review mechanism to ensure that users have confidence that their concerns are being treated fairly. We are consulting on options, including allowing designated bodies to make 'super complaints' to the regulator in order to defend the needs of users.

27. Ahead of the implementation of the new regulatory framework, we will continue to encourage companies to take early action to address online harms. To assist this process, this White Paper sets out high-level expectations of companies, including some specific expectations in relation to certain harms. We expect the regulator to reflect these in future codes of practice.

28. For the most serious online offending such as CSEA and terrorism, we will expect companies to go much further and demonstrate the steps taken to combat the dissemination of associated content and illegal behaviours. We will publish interim codes of practice, providing guidance about tackling terrorist activity and online CSEA later this year.

The companies in scope of the regulatory framework

29. We propose that the regulatory framework should apply to companies that allow users to share or discover user-generated content or interact with each other online.

30. These services are offered by a very wide range of companies of all sizes, including social media platforms, file hosting sites, public discussion forums, messaging services and search engines.

31. The regulator will take a risk-based and proportionate approach across this broad range of business types. This will mean that the regulator's initial focus will be on those companies that pose the biggest and clearest risk of harm to users, either because of the scale of the platforms or because of known issues with serious harms.

32. Every company within scope will need to fulfil their duty of care, particularly to counter illegal content and activity, comply with information requests from the regulator, and, where appropriate, establish and maintain a complaints and appeals function which meets the requirements to be set out by the regulator.

33. Reflecting the importance of privacy, any requirements to scan or monitor content for tightly defined categories of illegal content will not apply to private channels. We are consulting on definitions of private communications, and what measures should apply to these services.

An independent regulator for online safety

34. An independent regulator will implement, oversee and enforce the new regulatory framework. It will have sufficient resources and the right expertise and capability to perform its role effectively.

35. The regulator will take a risk-based approach, prioritising action to tackle activity or content where there is the greatest evidence or threat of harm, or where children or other vulnerable users are at risk. To support this, the regulator will work closely with UK Research and Innovation (UKRI) and other partners to improve the evidence base. The regulator will set out expectations for companies to do what is reasonably practicable to counter harmful activity or content, depending on the nature of the harm, the risk of the harm occurring on their services, and the resources and technology available to them.

36. The regulator will have a legal duty to pay due regard to innovation, and to protect users' rights online, taking particular care not to infringe privacy or freedom of expression. We are clear that the regulator will not be responsible for policing truth and accuracy online.

37. The government is consulting on whether the regulator should be a new or existing body. The regulator will be funded by industry in the medium term, and the government is exploring options such as fees, charges or a levy to put it on a sustainable footing. This could fund the full range of the regulator's activity, including producing codes of practice, enforcing the duty of care, preparing transparency reports, and any education and awareness activities undertaken by the regulator.

Enforcement of the regulatory framework

38. The regulator will have a range of enforcement powers, including the power to levy substantial fines, that will ensure that all companies in scope of the regulatory framework fulfil their duty of care.

39. We are consulting on which enforcement powers the regulator should have at its disposal, particularly to ensure a level playing field between companies that have a legal presence in the UK, and those which operate entirely from overseas.

40. In particular, we are consulting on powers that would enable the regulator to disrupt the business activities of a non-compliant company, measures to impose liability on individual members of senior management, and measures to block non-compliant services.

41. The new regulatory framework will increase the responsibility of online services in a way that is compatible with the EU's e-Commerce Directive, which limits their liability for illegal content until they have knowledge of its existence, and have failed to remove it from their services in good time.

Technology as part of the solution

42. Companies should invest in the development of safety technologies to reduce the burden on users to stay safe online.

43. In November 2018, the Home Secretary co-hosted a hackathon with five major tech companies to develop a new tool to tackle online grooming, which will be licensed for free to other companies, but more of these innovative and collaborative efforts are needed.

44. The government and the regulator will work with leading industry bodies and other regulators to support innovation and growth in this area and encourage the adoption of safety technologies.

45. The government will also work with industry and civil society to develop a safety by design framework, linking up with existing legal obligations around data protection by design and secure by design principles, to make it easier for start-ups and small businesses to embed safety during the development or update of products and services.

Empowering users

46. Users want to be empowered to keep themselves and their children safe online, but currently there is insufficient support in place and many feel vulnerable online.

47. While companies are supporting a range of positive initiatives, there is insufficient transparency about the level of investment and the effectiveness of different interventions. The regulator will have oversight of this investment.

48. The government will develop a new online media literacy strategy. This will be developed in broad consultation with stakeholders, including major digital, broadcast and news media organisations, the education sector, researchers and civil society. This strategy will ensure a coordinated and strategic approach to online media literacy education and awareness for children, young people and adults.

Next steps

49. This is a complex and novel area for public policy. To this end, as well as setting out the government's proposed approach, this White Paper poses a series of questions about the design of the new regulatory framework and non-legislative package. A full list of these questions is included at the end of this White Paper.

PART 1: Introduction

1: The challenge

Summary

- Illegal and unacceptable content and activity is widespread online, and UK users are frequently concerned about what they have seen or experienced.
- The prevalence of the most serious illegal content and activity on the internet, which threatens our national security or the physical safety of children, is unacceptable. The ease and extremity of the most serious online offending such as child sexual exploitation and abuse (CSEA) continues to increase.
- The impact of harmful content and activity can be particularly damaging for children and young people, and there are growing concerns about the potential impact on their mental health and wellbeing.
- Tackling illegal and harmful content and activity online is one part of the UK's wider mission to develop rules and norms for the internet, including protecting personal data, supporting competition in digital markets and promoting responsible digital design.

1.1 The internet is an integral part of everyday life for so many people. Nearly nine in ten UK adults are online and adult users spend around one day a week on the internet.[1] This is also true for children and young people, with 99% of 12-15 year olds going online, spending an average of twenty and a half hours a week on the internet.[2]

1.2 The internet can be a powerful force for good. It serves humanity, spreads ideas and enhances freedom and opportunity across the world. Online services facilitate the exchange of information, goods and services. They match supply and demand with great efficiency, increase consumer choice and lower distance between participants.

1 Ofcom (2018). Adults' Media Use and Attitudes Report. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0011/113222/Adults-Media-Use-and-Attitudes-Report-2018.pdf

2 Ofcom (2018). Children and parents: media use and attitudes report 2018. Available at: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2018

1.3 However, there is growing evidence of the scale of harmful content and activity that people experience online. Online services can be used to spread terrorist propaganda and child abuse content, they can be a tool for abuse and bullying, and they can be used to undermine civil discourse. Despite the many benefits of the internet, more than one in four adult users in the UK have experienced some form of harm related either to content or interactions online.[3]

1.4 Social media platforms and other technology companies increasingly acknowledge that they have a greater responsibility to protect their users from harm. British citizens want to feel empowered to keep themselves and their children safe and secure online. Both the government and industry have a responsibility to ensure this is the case.

Online harms suffered by individuals

1.5 The most appalling and horrifying illegal content and activity remains prevalent on an unacceptable scale. Existing efforts to tackle this activity have not delivered the necessary improvements, creating an urgent need for government to intervene to drive online services to step up their response.

1.6 There is a growing threat presented by online CSEA. In 2018 there were over 18.4 million referrals of child sexual abuse material by US tech companies to the National Center for Missing and Exploited Children (NCMEC).[4] Of those, there were 113, 948 UK-related referrals in 2018, up from 82,109 in 2017. In the third quarter of 2018, Facebook reported removing 8.7 million pieces of content globally for breaching policies on child nudity and sexual exploitation.[5]

1.7 Not only is the scale of this offending increasing, so is its severity. The internet Watch Foundation (IWF) estimates that 55% of the child sexual abuse material they find online contains children aged ten or under, and 33% of this imagery is in the most serious category of abuse.[6]

1.8 Terrorists also continue to use online services to spread their vile propaganda and mobilise support (see Box 2). Terrorist content online threatens the UK's national security and the safety of the public.

1.9 All five terrorist attacks in the UK during 2017 had an online element, and online terrorist content remains a feature of contemporary radicalisation.[7] It is seen across terrorist investigations, including cases where suspects have become very quickly radicalised to the point of planning attacks. This is partly as a result of the continued availability and deliberately attractive format of the terrorist material they are accessing online.

1.10 Terrorist groups work to find new ways to spread their propaganda and evade government and law enforcement efforts to prevent this. These threats are not only restricted to the largest, best-known services, but are prevalent across the internet. Terrorist groups and their supporters constantly diversify their reliance on the online services they use to host their

3 Ofcom and ICO (2018). Internet users' experience of harm online 2018. Available at: https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm-online

4 NCMEC. Available at: http://www.missingkids.com/footer/media/vnr/vnr2

5 Facebook (2018). Transparency Report. Available at: https://transparency.facebook.com/community-standards-enforcement#child-nudity-and-sexual-exploitation

6 Internet Watch Foundation (2017). Annual Report 2017. Available at: https://annualreport.iwf.org.uk/

7 Speech at Digital Forum, San Francisco by the Rt Hon Amber Rudd, 13 February 2018.

material online. While Facebook reported removing over 14 million pieces of content related to terrorism or violent extremism in 2018,[8] the terrorist group Daesh used over 100 platforms in 2018, making use of a wider range of more permissive and smaller platforms.

1.11 We have also seen terrorists and their supporters adopting new techniques, with material being shared using hacked social media accounts, and propaganda videos being edited in an effort to avoid detection.

1.12 Terrorist groups place a huge premium on quickly reaching their audiences. A third of all links to Daesh propaganda, for example, are disseminated within an hour of upload, while in the immediate aftermath of the terrorist attack in Christchurch, there was a co-ordinated cross-platform effort to generate maximum reach of footage of the attack. It is therefore vital to ensure that there is the technology in place to automatically detect and remove terrorist content within an hour of upload, secure the prevention of re-upload and prevent, where possible, new content being made available to users at all.

1.13 The threat continues to evolve with terrorists' relentless desire to seek out new ways to share their propaganda in an effort to radicalise and recruit. The most effective way to combat this adaptive threat is to have a consistent cross-platform response to ensure there are no safe spaces for terrorists to operate online.

1.14 Rival gangs use social media to glamourise weapons and gang life, as well as to directly depict or incite acts of violence. Alongside the illegal sale of weapons to young people online, this is a contributing factor to incidents of serious violence, including knife crime, in the UK. The latest police recorded crime figures, for the year ending September 2018, show an 8% increase in knife crime (to 39,818 offences) compared with the previous year. Homicide figures have risen by 14% (excluding terrorist attacks) over the same period.[9]

Harm: Child sexual exploitation and abuse online — Box 1

Threat:

Child sex offenders use the internet to view and share Child Sexual Abuse Material (CSAM), groom children online, and live stream the sexual abuse of children. The sheer scale of CSEA online is horrifying.

- In 2017, the IWF assessed 80,319 confirmed reports of websites hosting or linking to images of child sexual abuse. A total of 43% of the children in the images were aged 11-15 years old, and 57% were ten years old or younger. Two per cent were aged two or younger.[10]
- Sexual exploitation can happen to any young person – whatever their background, age, gender, race or sexuality or wherever they live.
- In the most horrific cases, child sex offenders in developing countries are abusing children at the instigation of offenders in the UK who commission the abuse online and watch it over live stream for a fee.

8 Facebook (2018). Transparency Report. Available at: https://transparency.facebook.com/community-standards-enforcement#child-nudity-and-sexual-exploitation

9 ONS (2019). Crime in England and Wales, Year Ending September 2018. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2018

10 Internet Watch Foundation (2017). Annual Report 2017. Available at: https://annualreport.iwf.org.uk/

Impact:

- Victims of abuse report ongoing trauma caused by the knowledge that images of their abuse are still being circulated and viewed by child sex offenders online. Victims also fear being recognised as a result of their images being available online.
- Victims of online grooming suffer lasting harm after being blackmailed and coerced into sharing indecent images of themselves or live-streaming themselves to offenders, and live in fear that those images could be used against them.

Harm: Terrorist content online

Box 2

Threat:

Terrorists, including Islamist groups such as Daesh and Al-Qaeda as well as far right terrorists, use the internet to spread propaganda designed to radicalise vulnerable people, and distribute material designed to aid and abet terrorist attacks. There are also examples of terrorists broadcasting attacks live on social media. Terrorist use of the internet poses a threat to national security and the safety of the public.

- As larger platforms take more action against terrorist propaganda, terrorist groups have spread out to a wider range of more permissive and smaller platforms.
- Terrorist groups are adopting new techniques to avoid detection, including sharing material via hacked social media accounts, and subtly altering propaganda videos.
- Terrorist groups place a huge premium on quickly reaching their audiences.

Impact:

- The availability and spread of terrorist content online has been shown to contribute to terrorist attacks on UK soil.
- All five of the terrorist attacks undertaken in the UK during 2017 had an online element to them.[11]
- Online terrorist content is seen across terrorist investigations, including cases where suspects have become very quickly radicalised to the point of planning attacks.

11 Speech at Digital Forum (2018). San Francisco by the Rt Hon Amber Rudd, 13 February 2018

Harm: Content illegally uploaded from prisons

Box 3

Threat:

There are an increasing number of cases where online content originating from prisons is illegally uploaded by prisoners to social media.

- Some prisoners transmit videos, images and messages from prisons using prohibited devices such as mobile phones.
- They can use social media accounts to harass and intimidate their victims.

Impact:

- This can lead to victims of crime feeling that they have no escape from their tormentors, even when they have been imprisoned.
- Prisoners openly uploading content from prisons can also undermine public confidence in the prison service.

Tackling serious violence online

Box 4

Rival gangs use social media to promote gang culture, taunt each other and incite violence. Content can also either directly depict or incite real world violence or glamourise gang life and the use of weapons. Government and law enforcement are taking action to tackle this threat:

- We have provided £1.4 million to support a new national police capability to tackle gang related activity on social media.
- This will bring together a dedicated team to take action against online material, focusing on investigative, disruption and enforcement work against specific gang targets, as well as making referrals to social media companies so illegal and harmful content can be taken down.
- Prior to this, a new action group was established to bring together government, social media companies, police and community groups to tackle violent material available via social media.

Harm: The sale of opioids online — Box 5

Threat:

Powerful and dangerous opioids are marketed and sold online. Fentanyl and its analogues (substances with similar but slightly altered chemical structures) are a group of powerful synthetic opioids. They have similar effects to other opioids such as morphine and heroin, but are significantly more potent.

- Since December 2016, there have been at least 143 recorded deaths in the UK attributed to Fentanyl and its analogues.[12] Fentanyl has been sold on several well-known social media sites. The products are marketed as top-quality substances with fast and secure delivery, with mobile phone numbers provided for follow-up contact.[13]
- Of particular concern is that some social media groups and threads, including those used by vulnerable people, are being targeted. This includes people suffering from chronic pain, where there is a risk of accidental overdose and people dealing with depression, where there is a risk that the fentanyl may be used to assist suicide.

Impact:

- Whilst these products continue to be made available there is a risk that fatalities will increase.
- There is also a risk that health professionals and other first responders will continue to be exposed to potentially harmful environments.

1.15 Beyond illegal activity, other behaviour online also causes harm. In 2017, one in five children aged 11 to 19 reported having experienced cyberbullying in the past year[14]; 21% of women have received misogynistic abuse online,[15] and half of girls aware of sexist abuse on social media say this has restricted what they do or aspire to in some way.[16] The House of Commons Petitions Committee has highlighted the extreme abuse experienced online by disabled people, which has forced some of them to leave social media.[17]

1.16 Victims have also described a qualitative difference between online and offline harms, particularly in reference to online abuse. The Law Commission noted the perceived anonymity of offenders as one of the characteristics of online abuse that may result in a different experience for victims – see Box 6.[18] Many users also feel that the market currently offers

12 NCA analysis.
13 Ibid.
14 NHS Digital (2018). Mental Health of Children and Young People in England, 2017. Available at: https://files.digital.nhs.uk/C9/999365/MHCYP%202017%20Behaviours%20Lifestyles%20Identities.pdf
15 Amnesty International (2017). The impact of online abuse against women. Available at: https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/
16 Girl Guiding (2016). Girls Attitudes Survey 2016. Available at: https://www.girlguiding.org.uk/globalassets/docs-and-resources/research-and-campaigns/girls-attitudes-survey-2016.pdf
17 House of Commons Petitions Committee (2019). Online abuse and the experience of disabled people. The Petitions Committee, 2019. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmpetitions/759/759.pdf
18 Law Commission (2018). Abusive and Offensive Communications. Available at: https://www.lawcom.gov.uk/abusive-and-offensive-online-communications/

them very few alternative, safer online services. For example, the 2018 Doteveryone Digital Attitudes[19] report found that almost half of respondents felt they had no choice but to sign up to online services, even where they had concerns.

Tackling online anonymous abuse — Box 6

The internet can be used to harass, bully or intimidate. In many cases of harassment and other forms of abusive communications online, the offender will be unknown to the victim. In some instances, they will have taken technical steps to conceal their identity. Government and law enforcement are taking action to tackle this threat.

- The police have a range of legal powers to identify individuals who attempt to use anonymity to escape sanctions for online abuse, where the activity is illegal. The government will work with law enforcement to review whether the current powers are sufficient to tackle anonymous abuse online.
- We are enhancing law enforcement's ability to tackle anonymous online abuse by investing in training that is designed to improve digital capability across policing. For example, as part of the £4.6 million Police Transformation Fund allocated by the Home Office, the Digital Investigation and Intelligence programme will build police capability to respond to the full range of digital crime types, through investment in technology and training.
- We are also making it easier for the public to report online crimes. Through the Digital Public Contact programme, we will provide the public with a digitally accessible police force with a consistent set of online capabilities to use in engaging and transacting with police services through a single online channel.
- We also expect companies to do substantially more to keep their users safe and counter online abuse, particularly where this is illegal. Companies need to take responsibility for tackling abusive behaviour on their services. More detail is set out in Chapter 3.

Online harms suffered by children and young people

1.17 Being online can be a hugely positive experience for children and young people – see Box 7. Recent research by internet Matters found that seven in ten parents think screen time is essential for their children's learning development and two thirds of parents feel that devices give their children another outlet for creativity, particularly so for children aged 6-10.[20]

1.18 However, the impact of harmful content and activity can be particularly damaging for children, as set out in Box 1 above and Boxes 8-10 below. There is also growing concern about the relationship between social media and the mental health of children and young people. The Children's Commissioner's report published in November 2018 *Who knows what about me* sets out the huge size and growth of children's digital footprint and the associated

19 Doteveryone (2018). People, Power and Technology: The 2018 Digital Attitudes Report. Available at: https://attitudes.doteveryone.org.uk/files/People%20Power%20and%20Technology%20Doteveryone%20Digital%20Attitudes%20Report%202018.pdf

20 Internet Matters (2018). Look Both Ways: Practical Parenting in the Age of Screens. Available at: https://www.Internetmatters.org/about-us/screen-time-report-2018/

risks and benefits.[21] Internet Matters reported in February 2019 that vulnerable young people are more likely to suffer online harms and less likely to receive online safety advice and education.[22]

The positive impact of being online for children and young people — Box 7

Most children have a positive experience online, using the internet for social networking and connecting with peers, as well as to access educational resources, information, and entertainment. The internet opens up new opportunities for learning, performance, creativity and expression.

- A literature review by the UK Council for Child Internet Safety (2017) highlights evidence that young people recognise the positive role of the internet in relation to self-expression, developing understanding, bringing people together and respecting and celebrating differences. [23] Research by UNICEF (2017) shows that use of technology is beneficial for children's social relationships, enabling them to enhance existing relationships and build positive friendships online.[24]
- A report by The Royal Society for Public Health in 2017 found that young people reading blogs or watching vlogs on personal health issues helped improve their knowledge and understanding, prompted individuals to access health services, and enabled them to better explain their own health issues or make better choices.[25] They also found that young people are increasingly turning to social media as a means of emotional support to prevent and address mental health issues.
- More recently, research by Ofcom showed that nine in ten social media users aged 12-15 state that this use has made them feel happy or helped them feel closer to their friends. Two thirds of 12-15 year olds who use social media or messaging sites say they send support messages, comments or posts to friends if they are having a difficult time. One in eight support causes or organisations by sharing or commenting on posts.[26]
- In the 2019 UK Safer Internet Centre survey,[27] 70% of young people surveyed said that being online helps them understand what's happening in the world, with 60% noting they have only seen or heard about certain issues or news because they heard about them from the internet. 43% said they have been inspired to take action because of something they saw online, with 48% stating being online makes them feel that their voice or actions matter.

21 Children's Commissioner (2018). Who knows what about me? Available at: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2018/11/who-knows-what-about-me.pdf

22 Internet Matters (2019). Vulnerable Children in a Digital World. Available at: https://pwxp5srs168nsac2n3fnjyaa-wpengine.netdna-ssl.com/wp-content/uploads/2019/02/Vulnerable-Children-in-a-Digital-World-FINAL.pdf

23 UKCCIS Evidence Group (2017). Children's online activities, risks and safety. Available at: https://www.gov.uk/government/publications/childrens-online-activities-risks-and-safety-a-literature-review-by-the-ukccis-evidence-group

24 UNICEF (2017). How does the time children spend using digital technology impact their mental well-being, social relationships and physical activity? Available at: https://www.unicef-irc.org/publications/pdf/Children-digital-technology-wellbeing.pdf

25 RSPH (2017). Status of mind: Social media and young people's mental health and wellbeing. Available at: https://www.rsph.org.uk/uploads/assets/uploaded/62be270a-a55f-4719-ad668c2ec7a74c2a.pdf

26 Ofcom (2018). Children and parents: media use and attitudes report 2018. Available at: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2018

27 UK Safer Internet Centre (2019). Our internet, Our Choice Report. Available at: https://www.saferInternet.org.uk/safer-Internet-day/safer-Internet-day-2019/our-Internet-our-choice-report

Harm: Cyberbullying Box 8

Threat:

In 2017, one in five children surveyed aged 11-19 reported having experienced cyberbullying in the past year.[28]

- The prevalence of cyberbullying is higher for some groups, such as women, religious minorities, LGBT+, BME and disabled individuals.[29]

Impact:

- Cyberbullying has been shown to have psychological and emotional impact. In a large survey of young people who had been cyberbullied, 41% had developed social anxiety, 37% had developed depression, 26% had suicidal thoughts and 25% had self-harmed.[30]
- These figures are all higher than corresponding statistics for 'offline' bullying, and indicated the increased potential for harm of cyberbullying.

Harm: Self-harm and suicide Box 9

Threat:

In a survey of young adults, 22.5% reported self-harm and suicide-related internet use, including 8.2% and 7.5% who had actively searched for information about self-harm and suicide respectively.[31]

- Amongst those who had harmed with suicidal intent, 70% reported self-harm and suicide-related internet use.[32]
- The prevalence of using the internet to view related content has also been found to be higher in children than adults. One study of those presenting to hospital following self-harm found that 26% of children had viewed self-harm and suicide content, compared to 8.4% of adults.[33]

28 NHS Digital (2018). Mental Health of Children and Young People in England, 2017. Available at: https://files.digital.nhs.uk/C9/999365/MHCYP%202017%20Behaviours%20Lifestyles%20Identities.pdf

29 Ditch the Label (2017). 'The Annual Bullying Survey 2017'. Available at: https://www.ditchthelabel.org/wp-content/uploads/2017/07/The-Annual-Bullying-Survey-2017-2.pdf

30 Ibid.

31 Mars, B et al. (2015). Exposure to, and searching for, information about suicide and self-harm on the internet: Prevalence and predictors in a population based cohort of young adults' Journal of affective disorders,185, 239-45. Available at: https://doi.org/10.1016/j.jad.2015.06.001

32 Ibid.

33 Padmanathan, P. et al. (2018). Suicide and Self-Harm Related internet Use. Crisis. Available at: https://doi.org/10.1027/0227-5910/a000522

Impact:

- The National Confidential Inquiry into Suicide and Safety in Mental Health (NCISH) analysed the characteristics of 595 children and young people (aged under 20) who had died by suicide in the UK between 2014 and 2016.
- The NCISH found that suicide-related internet use (i.e. searching the internet for information on suicide methods) was reported for almost a quarter (23%) of these children and young people.[34]

Harm: Underage sharing of sexual imagery Box 10

Many children and young people take and share sexual images. Creating, possessing, copying or distributing sexual or indecent images of children and young people under the age of 18 is illegal, including those taken and shared by the subject of the image.

- Surveys provide tentative evidence that between 26%[35] and 38%[36] of 14-17 year olds have sent sexual images to a partner, and between 12% and 49% have received a sexual image.[37]
- The proportion of young people sending images varies with age, with one study indicating that 26% of 14 year olds had sent and received sexual images, rising to 48% of 16 year olds.[38]

Impact:

- Sharing sexual images can expose children and young people to bullying, humiliation, objectification and guilt. These images can be shared widely and appear on offender forums or adult pornography sites, or be used to extort further imagery. This puts children and young people in a vulnerable position and at risk of harm. It is a criminal offence to produce, possess or share sexual images of under 18 year olds.
- The National Society for the Prevention of Cruelty to Children (NSPCC) reported that sexting was discussed in 1,392 counselling sessions with children and young people on their helplines that year, representing a 15% increase on the year before.[39]

1.19 The UK Chief Medical Officers (UK CMOs) commissioned independent researchers to carry out a systematic evidence review on the impact of social media use on children and young people's mental health. The review covered important and diverse issues including cyberbullying, online gaming, sleep problems and problematic internet use, which is also known as 'internet addiction'.

34 National Confidential Inquiry into Suicide and Safety in Mental Health (2018). Annual Report: England, Northern Ireland, Scotland, Wales. University of Manchester. Available at: http://documents.manchester.ac.uk/display.aspx?DocID=38469
35 Brook (2017). Digital Romance. Available at: https://www.brook.org.uk/press-releases/digital-romance
36 UKCCIS Evidence Group (2017). Children's online activities, risks and safety. Available at: https://www.gov.uk/government/publications/childrens-online-activities-risks-and-safety-a-literature-review-by-the-ukccis-evidence-group
37 Ibid.
38 Ibid.
39 Ibid.

1.20 Overall the research did not present evidence of a causal relationship between screen-based activities and mental health problems, but it did find some associations between screen-based activities and negative effects, such as increased risk of anxiety or depression.[40] It is important that parents and carers support their children to have positive experiences online.

1.21 While there is not yet sufficient evidence about the impact of screen time to support detailed guidelines for parents or requirements on companies, we will continue to support research in this area and ensure high quality advice is available to families. We also welcome efforts from the industry to develop tools to help individuals and families understand and manage how much time they spend online – more information on these is in Box 33.

Emerging challenge: Screen time — Box 11

Screen time and its impact on children is an issue of growing concern. Research by Internet Matters found that nearly half of parents (47%) are concerned about the amount of time their child spends online and 88% take measures to limit their child's use of devices.[41]

- The UK CMOs recently conducted a systematic evidence review on children and young people's screen and social media use. The CMO subsequently produced advice for parents and carers to encourage them to discuss boundaries with children around online behaviours and time spent using screens, and to lead by example.
- For example, the UK CMOs advised that:
  - Sleep matters. Getting enough good quality sleep is very important. Leave phones outside the bedroom when it is bedtime.
  - Sharing sensibly. Talk about sharing photos and information online and how photos and words are sometimes manipulated. Parents and carers should never assume that children are happy for their photos to be shared. For everyone – when in doubt, don't upload!
  - Education matters. Make sure you and your children are aware of, and abide by, their school's policy on screen time.
  - Keep moving! Everyone should take a break after a couple of hours sitting or lying down using a screen. It's good to get up and move about a bit. #sitlessmovemore
  - Safety when out and about. Advise children to put their screens away while crossing the road or doing an activity that needs their full attention.
  - Talking helps. Talk with children about using screens and what they are watching. A change in behaviour can be a sign they are distressed – make sure they know they can always speak to you or another responsible adult if they feel uncomfortable with screen or social media use.

40 Department of Health and Social Care (2019). United Kingdom Chief Medical Officers' commentary on Screen-based activities and children and young people's mental health and psychosocial wellbeing: a systematic map of reviews. Available at: https://www.gov.uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-of-reviews

41 Internet Matters (2018). Look Both Ways: Practical Parenting in the Age of Screens. Available at: https://www.Internetmatters.org/about-us/screen-time-report-2018/

- Family time together. Screen-free meal times are a good idea – you can enjoy face-to-face conversation, with adults giving their full attention to children.
- Use helpful phone features. Some devices and platforms have special features – try using these features to keep track of how much time you (and with their permission, your children) spend looking at screens or on social media.

Future action – building our understanding:

Given the amount of time many children spend online, and the level of parental concern on this issue, we urgently need to build a better understanding.

- While we do not expect the regulator to set requirements around screen time, both government and the regulator will continue to support research in this area to inform future action in this space.
- We need to develop a better understanding of not just of the impact of screen time as a whole, but also between different types of screen time and children's development and wellbeing.
- As part of this, we also expect companies to support the developing evidence base around screen time, for example by providing access to anonymised data to researchers as recommended by the CMOs
- If the emerging evidence base demonstrates a strong link between different elements of screen time and damage to children's wellbeing or development, companies will be expected to take appropriate action to fulfil their duty of care.

Threats to our way of life

1.22 The UK's reputation and influence across the globe is founded upon our values and principles. Our society is built on confidence in public institutions, trust in electoral processes, a robust, lively and plural media, and hard-won democratic freedoms that allow different voices, views and opinions to freely and peacefully contribute to public discourse.

1.23 Inaccurate information, regardless of intent, can be harmful – for example the spread of inaccurate anti-vaccination messaging online poses a risk to public health. The government is particularly worried about disinformation (information which is created or disseminated with the deliberate intent to mislead; this could be to cause harm, or for personal, political or financial gain).

1.24 Disinformation threatens these values and principles, and can threaten public safety, undermine national security, fracture community cohesion and reduce trust.

1.25 These concerns have been well set out in the wide-ranging inquiry led by the Digital, Culture, Media and Sport (DCMS) Select Committee report on fake news and disinformation, published on 18 February 2019. This White Paper has benefited greatly from this analysis and takes forward a number of the recommendations. The government will be responding to the DCMS Select Committee report in full in due course. We also note the recent papers from the Electoral Commission and Information Commissioner's Office on this and wider issues, and are considering these closely.

Harm: Online disinformation — Box 12

Threat:

Online disinformation – spreading false information to deceive deliberately – is becoming more and more prevalent. Misinformation refers to the inadvertent sharing of false information.

- A recent study from the University of Oxford's Computational Propaganda Project has found evidence of organised social media manipulation campaigns in 48 countries in 2018.[42] According to the Reuters Institute, 61% of people want the government to do more to separate what is real and fake on the internet (2018).[43]
- One of the major technological challenges in disinformation is the continued development of AI systems. AI techniques can be used to target and manipulate individual voters, with highly sophisticated micro-targeting based on individual psychology.
- AI can be beneficial in the automatic detection of content, or automatically fact-checking articles. But developments in AI also make it possible to generate fake content (text, audio and video) which is difficult to detect by humans and algorithms – known as 'deepfakes'. As a result, it is becoming even easier to create and disseminate false content and narratives.
- The Russian State is a major source of disinformation. The Kremlin has used disinformation to obfuscate and confuse audiences around their illegal annexation of Crimea, intervention in eastern Ukraine and the shooting down of Malaysian Airlines flight MH17, which led to the deaths of 298 people including ten UK citizens. After the attempted murder of Sergei and Yulia Skripal in Salisbury in March 2018, the Russian State led a concerted disinformation campaign to distract from their culpability. This included the use of state media and covert social media accounts to sow over 40 different narratives as to what happened.

Impact:

- Most users are not always aware that much of the content they see is determined by sophisticated algorithms that draw on data about their online activity, such as their browsing history, their social media networks and what they post.
- Research by Doteveryone suggests that 62% of people do not realise that their social networks can affect the news they see,[44] while only three in ten adult online users questioned by Ofcom were aware of the ways in which companies can collect data about them online.[45]

42 Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation.Available at: https://comprop.oii.ox.ac.uk/research/cybertroops2018/

43 Newman, N. et al. (2018). Reuters Institute Digital News Report 2018. Available at: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475

44 Miller, C., Coldicutt, R., and Kitcher, H. (2018). People, Power and Technology: The 2018 Digital Understanding Report Doteveryone. Available at: http://understanding.doteveryone.org.uk/files/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf

45 Ofcom (2018). Adults' Media Use and Attitudes Report. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0011/113222/Adults-Media-Use-and-Attitudes-Report-2018.pdf

Harm: Online manipulation — Box 13

Threat:

Propaganda and false information have long been used to persuade and mislead, but the internet, social media and AI provide ever more effective ways to manipulate opinion.

- The tolerance of conflicting views and ideas are core facets of our democracy. However, these are inherently vulnerable to the efforts of a few to manipulate and confuse the information environment for nefarious purposes, including undermining trust. A combination of personal data collection, AI based algorithms and false or misleading information could be used to manipulate the public with unprecedented effectiveness.
- The distinction between legitimate influence and illegitimate manipulation is not new. The government took action to prevent subliminal broadcast advertising in the Broadcasting Act 1990. The government gave the Independent Television Commission (replaced by Ofcom) a duty to ensure that licensed services complied with requirements not to include technical devices which convey messages or influence individuals without them being aware. We believe the government should make sure there are similar boundaries between legitimate and illegitimate practices online. The techniques and practices used are still emerging. We are developing a better understanding of the nature and scale of the potential problem and effective interventions.

Harm: Online abuse of public figures — Box 14

Threat:

In recent years we have seen a worrying rise in the amount of abuse, harassment and intimidation directed at those in public life. Much of this abuse happens on social media.

Impact:

- An international survey of female journalists found two thirds (64%) had experienced online abuse – death or rape threats, sexist comments, cyberstalking, account impersonation, and obscene messages.[46] Almost half (47%) did not report the abuse they had received, and two fifths (38%) admitted to self-censorship in the face of this abuse.[47]
- The Guardian's research into the 70 million comments left on its site over a ten year period highlighted that of the ten most abused writers, eight were women and two were black men. This is in spite of the fact that the majority of the regular opinion writers for The Guardian are white men. This was then compared to the ten writers who received the least abuse – who were all men.[48]

46 IFJ (2018). IFJ global survey shows massive impact of online abuse on women journalists. Available at: https://www.ifj.org/media-centre/news/detail/article/ifj-global-survey-shows-massive-impact-of-online-abuse-on-women-journalists.html
Note no similar data is available for male journalists.

47 Ibid.

48 The Guardian (2016). The dark side of Guardian comments. Available at: https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments

There are too many stories of public figures closing their social media accounts following waves of abuse.

- In December 2017 the Committee for Standards in Public Life, which was commissioned by the Prime Minister, published its report on intimidation in public life.[49] The consultation sought views on a range of ideas including establishing a new offence of intimidation, and requiring imprints on electronic campaigning. The report included examples of the extent of intimidation of those in public life.
- "It is hard to explain how it makes you feel. It is anonymous people that you've never met, true, but it has a genuinely detrimental effect on your mental health. You are constantly thinking about these people and the hatred and bile they are directing towards you." – Rachel Maclean MP
- "I spoke on a number of occasions in the House of Commons in different committees about the rights of women. To which I suffered daily attacks on Twitter, on my email system or endless online articles written about how people wished to see me raped." – Jess Phillips MP
- The report also makes a number of recommendations for actions that social media companies should take in relation to intimidatory content, including implementing tools to enhance the ability of users to tackle online intimidation and supporting users who become victims of this behaviour.[50] These recommendations have helped to shape the indicative list of steps the regulator may want to include in codes of practice.
- The government's response to the Committee on Standards in Public Life's Review of Intimidation in Public Life was published in March 2018 and set out a number of actions for government based on the Committee's recommendations. As part of this work the government has undertaken a public consultation entitled Protecting the Debate: Intimidation, Influence and Information which closed in October 2018. The government's response will be published in due course.

  This abuse is unacceptable – it goes beyond free speech and free debate, dissuades good people from going into public life, and corrodes the values on which our democracy rests.

  The new regulatory framework will make clear companies' responsibility to address this harm.

Other online harms

1.26 There are other harms associated with the internet and online technology. For example, Ofcom and ICO's report on Internet Regulation highlighted users' concerns around privacy and hacking.[51] This White Paper is part of the government's wider programme of work to establish the right norms and rules for the internet.

49 Committee on Standards in Public Life (2017). Intimidation in Public Life. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1__2_.pdf

50 Government action to date is available here: https://www.parliament.uk/business/publications/written-questions-answers-statements/written-statement/Commons/2019-03-07/HCWS1389/

51 Ofcom and ICO (2018). Internet users' experience of harm online. Available at: https://www.ofcom.org.uk/research-and-data/Internet-and-on-demand-research/Internet-use-and-attitudes/Internet-users-experience-of-harm-online

Responsible and ethical technology

1.27 The government takes both the protection of personal data and the right to privacy extremely seriously. The General Data Protection Regulation (GDPR) and the Data Protection Act 2018 (DPA), alongside the increased powers for the ICO to gather evidence, inspect artificial intelligence (AI) and levy significant fines on those who break the law, update our data protection laws fit for the digital age.

1.28 The DPA also includes an important new provision requiring the Information Commissioner to produce an age-appropriate design code. This provision breaks new ground by addressing the approach to the design of online services likely to be used by children. It ensures platforms and service providers put child user interests at the centre of the design process, and protects them from risks that arise from the use of their personal data online, including the algorithms and profiling that serves them with personalised content.

1.29 However, the increased use of data and AI is giving rise to complex, fast-moving and far-reaching ethical and economic issues that cannot be addressed by data protection laws alone. Increasingly sophisticated algorithms can glean powerful insights, which can be deployed in ways that influence the decisions we make and the services we receive. It is essential that we understand, and respond to, barriers to the ethical deployment of AI.

1.30 That is why the government has set up the Centre for Data Ethics and Innovation. The Centre will provide independent, impartial and expert advice on the ethical and innovative deployment of data and AI. The Centre will publish its first strategy document in spring 2019, setting out further details on its key priorities.

1.31 The way that technology is designed, who it is designed by and the outcomes it is trying to achieve also influence how it impacts its users and wider society. There is an increasing amount of evidence that social media platforms and other digital services can impact people's habits, sleep patterns, productivity at work, attention spans and even voting preferences – see Box 15. We are looking carefully at how we can ensure that digital products and services are designed in a responsible way, with their users' well-being in mind. Chapter 8 of this paper looks specifically at how we are working with companies to include considerations around safety in the design of their products.

Emerging challenge: Designed addiction — Box 15

Some online products have been designed to encourage continuous use. They include seemingly small but influential features, which incentivise people to keep using the app or platform for longer. One common example is the 'infinite scroll', in which information is loaded continuously as the user scrolls down the page, encouraging the user to keep scrolling.

- A recent report by 5Rights highlighted other elements of 'persuasive design', such as 'typing bubbles', quantifying friends, and notifications. Even 'likes' can be powerful tools for keeping users online.[52]

52 5Rights (2018). Disrupted Childhood: The Cost of Persuasive Design, 5Rights. Available at: https://5rightsfoundation.com/static/5Rights-Disrupted-Childhood.pdf

- These techniques could exacerbate addictive behaviours. The number of people suffering from clinical addiction in this way has not been reliably quantified, but there are well-documented extreme cases of vulnerable individuals for whom addiction has got in the way of social lives, sleep, physical activity and other parts of a healthy, balanced lifestyle.
- There is also evidence that a wider range of people experience less extreme forms of compulsive or habitual behaviour online. Some experts would stress that this compulsive behaviour is not clinical addiction.

Future action:

- The government shares concerns around designed addiction and is determined to ensure that we have sufficient evidence on this risk, and the right expectations of companies to design their products in safe ways.
- In the future, we expect the regulator will continue to support research in this area to inform future action and, if necessary, set clear expectations for companies to prevent harm to their users.
- We also expect companies to be transparent about design practices which encourage extended engagement, and to engage with researchers to understand the impact of these practices on their users
- DCMS is continuing to work with the Gambling Commission and the industry on player protections in the online sector. In May 2018, we published the response to the Consultation on Proposals for Changes to Gaming Machines and Social Responsibility Measures, which set out a clear plan to strengthen player protections.
- Since then, a number of changes have been made to make gambling fairer and safer, including tightening advertising rules and launching GAMSTOP, the online self-exclusion scheme. Additionally, from May, the Gambling Commission will bring in changes that mean that age and identity must be verified before consumers can deposit money and gamble, and will require age verification before customers can access free-to-play demo games.

Thriving digital markets

1.32 The digital sector makes a huge contribution to our economy at £130.5 billion gross value added in 2017, equivalent to 7% of the UK gross value added. The sector has seen strong growth with an increase of 33% since 2010, compared to 29% for the total UK economy.[53] As the digital economy has grown, powerful new companies have emerged, often with very dominant market positions. This has raised questions about the competitiveness of digital markets and what this means for consumers.

1.33 The government's Modernising Consumer Markets Green Paper sought views on how well equipped the UK's competition regime is to manage emerging challenges, including the growth of fast-moving digital markets. We continue to consider policy options across the range of measures proposed in the green paper and are conducting a review of

53 DCMS (2019). Sectors Estimates 2017 (provisional): Gross Value Added. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/759707/DCMS_Sectors_Economic_Estimates_2017__provisional__GVA.pdf

digital markets, due in summer 2019. This will be informed by the work of the independent Digital Competition Expert Panel, led by Professor Jason Furman which published its recommendations for government on 13 March 2019.

1.34 Professor Furman and the panel found that the digital economy has brought significant benefits but does not have enough competition. They called for a new digital markets unit to set and enforce a code of conduct so the largest digital companies know what are acceptable rules for competition. This new unit would give people more control over their data by enabling people to switch between platforms more easily. They also recommended changes to merger rules, and updating existing rules to improve enforcement over anticompetitive conduct. The government will consider these proposals and respond later in the year.

1.35 Thriving digital markets also rely on the innovative, efficient and fair use of data. In June 2018, the Secretary of State for DCMS announced that we would develop a National Data Strategy to ensure the UK is a world-leading data economy – unlocking the power of data across government and the wider economy, while building public trust and confidence in its use.

Online advertising

1.36 Online advertising plays a crucial role in the digital economy, with many free digital services, such as search engines or social networks, funded by advertising revenues. The online advertising ecosystem is highly complex, with much of the advertising space online bought through automated processes, and the velocity with which adverts are created and displayed is far higher than offline. Online advertising encourages and rewards the collection of user data (the more data a service has on a user the more effectively it can target adverts at them) and the holding of people's attention (the longer they use a service the more adverts they see).

1.37 This combination of factors has given rise to a number of issues caused by or related to online advertising. Work is already underway to address some of these:

- A Home Office-led working group on CSEA and terrorist content linked to advertising met in March 2019, following an initial meeting in December 2018, comprising representatives from advertising trade bodies, agencies, brands, law enforcement and the IWF. The working group's activity to date comprises actions to help ensure advertising is not supporting this kind of illegal activity.
- As part of the government's Childhood Obesity Plan, DCMS and the Department of Health and Social Care launched a consultation on 18 March 2019 on introducing a 9pm watershed on TV advertising of products high in fat, salt or sugar, and similar protections for children viewing adverts online.
- The Competition and Markets Authority is considering further work on digital advertising, although this is dependent on the outcome of EU exit negotiations.
- In November 2018 the Advertising Standards Authority published its strategy More Impact Online, which aims to put the protection of consumers online at the heart of its work over the next five years, and makes commitments to explore, for example, the use of machine learning and AI to improve regulation.
- In 2018, the ICO conducted an investigation into data analytics and micro targeting of political advertising online. The report, 'Democracy Disrupted?', highlighted the risks of personal data being abused in digital campaigning and made a number

of recommendations to improve transparency and data protection compliance. The ICO has also commenced a broader examination of the use of personal data in adtech.[54]

1.38 As announced in the DCMS Secretary of State's immediate response to the Cairncross Review, DCMS will conduct a review of how online advertising is regulated in the UK.

54 ICO (2018). Democracy disrupted? Personal information and political influence. Available at: https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf

2: The harms in scope

Summary

- This White Paper sets out government action to tackle online content or activity that harms individual users, particularly children, or threatens our way of life in the UK, either by undermining national security, or by reducing trust and undermining our shared rights, responsibilities and opportunities to foster integration. It sets out an initial list of content and behaviour which will be in scope, as well as a list of harms which will be excluded.
- There is currently a patchwork of regulation and voluntary initiatives aimed at addressing these problems, but these have not gone far or fast enough to keep UK users safe online.
- Many of our international partners are also developing new regulatory approaches to tackle online harms, but the UK will be the first to tackle online harms in a coherent, single regulatory framework that reflects our commitment to a free, open and secure internet.

Harmful content or activity in scope of the White Paper

2.1 Table 1 below shows the initial list of online harmful content or activity in scope of the White Paper, based on an assessment of their prevalence and impact on individuals and society.

2.2 This list is, by design, neither exhaustive nor fixed. A static list could prevent swift regulatory action to address new forms of online harm, new technologies, content and new online activities.

Table 1: Online harms in scope

Harms with a clear definition	Harms with a less clear definition	Underage exposure to legal content
• Child sexual exploitation and abuse. • Terrorist content and activity. • Organised immigration crime. • Modern slavery. • Extreme pornography. • Revenge pornography. • Harassment and cyberstalking. • Hate crime. • Encouraging or assisting suicide. • Incitement of violence. • Sale of illegal goods/ services, such as drugs and weapons (on the open internet). • Content illegally uploaded from prisons. • Sexting of indecent images by under 18s (creating, possessing, copying or distributing indecent or sexual images of children and young people under the age of 18).	• Cyberbullying and trolling. • Extremist content and activity. • Coercive behaviour. • Intimidation. • Disinformation. • Violent content. • Advocacy of self-harm. • Promotion of Female Genital Mutilation (FGM).	• Children accessing pornography. • Children accessing inappropriate material (including under 13s using social media and under 18s using dating apps; excessive screen time).

2.3 There is already an effective response to some categories of harmful content or activity online. These will be excluded from the scope of the new regulatory framework to avoid duplication of existing government activity.

2.4 The following harms will be excluded from scope:

- All harms to organisations, such as companies, as opposed to harms suffered by individuals. This excludes harms relating to most aspects of competition law, most cases of intellectual property violation, and the organisational response to many cases of fraudulent activity. The government is leading separate initiatives to tackle these issues. For example, the Joint Fraud Taskforce is leading an ambitious programme of work to tackle fraud, including online fraud, through partnership between banks, law enforcement and government.
- All harms suffered by individuals that result directly from a breach of the data protection legislation, including distress arising from intrusion, harm from unfair processing, and any financial losses. Box 16 explains how the UK's legal framework provides protection against online harms linked to data breaches.

- All harms suffered by individuals resulting directly from a breach of cyber security or hacking. These harms are addressed through the government's National Cyber Security Strategy.
- All harms suffered by individuals on the dark web rather than the open internet. These harms are addressed in the government's Serious and Organised Crime Strategy. A law enforcement response to criminality on the dark web is considered the most effective response to the threat. As set out in the strategy, the government continues to invest in specialist law enforcement skills and capability.

Stronger regulation of personal data online — Box 16

The UK already enjoys high standards of data protection law, that were modernised in 2018 with the introduction of the GDPR and the Data Protection Act 2018. The government chose to go further than other countries, by providing stronger powers to apply to the investigation and enforcement of specific online threats.

Key protections for online harms involving personal data include:

- An obligation to provide clear and accessible privacy information, tailored for children when they are the users of online services.
- A legal obligation to accountability, making companies responsible for placing data protection at the centre of the design of online services in a way that mitigates the risk to users' information. This also includes a requirement to undertake data protection impact assessments, and have them approved by the ICO where high risks persist.
- A right to erasure of personal data online, with stronger provisions where data has been gathered from a child user.
- An age-appropriate design code, which gives the design standards we will expect providers of online services and apps used by children to meet when they process their data.
- A power to inspect algorithms in situ, to understand their use of personal data and whether this leads to bias or other detriment.
- A power to require information to be handed over to the ICO wherever it is held, including on cloud servers.

Shortcomings of the current regulatory landscape

2.5 Currently there is a range of UK regulations aimed at specific online harms or services in scope of the White Paper, but this creates a fragmented regulatory environment which is insufficient to meet the full breadth of the challenges we face. The current regulatory framework includes:

- GDPR and the Data Protection Act enforced by the ICO. This includes collection and use of personal data, including when online. The GDPR also has extraterritorial scope and can be enforced against companies outside the UK who offer services to UK users.[55]

55 The Information Commissioner's Office. Available at: https://ico.org.uk/

- The Electoral Commission's oversight of the activity of political parties, and other campaigners, including activity on social media.[56]
- Forthcoming age verification requirements for online pornography.[57]
- The Equality and Human Rights Commission's oversight of the Equality Act 2010 and Freedom of Expression.[58]
- Ofcom's existing oversight of video-on-demand services.[59]
- The revised EU Audiovisual Media Services Directive, which will introduce new high-level requirements for video sharing platforms such as YouTube.[60]
- The Gambling Commission's licensing and regulation of online gambling.[61] DCMS has been working with the Commission to tighten advertising rules on gambling and launched GAMSTOP, the online self-exclusion scheme. Additional age-verification requirements are expected to come take effect in from May this year[62].
- The Competition and Markets Authority's (CMA) enforcement of consumer protection law online. See Box 17 for further details.

Consumer enforcement by the Competition and Markets Authority Box 17

Businesses risk breaching consumer protection law where their online behaviour misleads consumers or treats them unfairly. The CMA has undertaken a range of recent enforcement activity examining potentially unfair or misleading online behaviour, including:

- Online gambling – the CMA worked with the Gambling Commission to sanction unfair online 'bonus' promotions by major gambling firms. The CMA was concerned that players' money could effectively be trapped under the terms of these promotions, or that they could be caught out by unclear or imbalanced promotion rules. Changes were agreed with a number of firms, including William Hill and Ladbrokes.
- Online reviews and endorsements – the CMA has an ongoing programme of work to tackle fake or misleading online reviews and endorsements. Most recently, 16 celebrities, reality stars and social media influencers committed to always be clear

56 The Political Parties, Elections and Referendums Act 2000 (PPERA) provides the Electoral Commission with the powers and functions to regulate political finance in the UK. Electoral law is also enforced by the police, who lead on the Representation of the People Act offences. The Electoral Commission has powers to investigate breaches of the rules to funding and spending for election and referendum campaigns, which includes digital campaigning.

57 The Digital Economy Act 2017 provides for the regulation of providers of online commercial pornography to ensure that pornographic material is not normally accessible by those under 18, and that content which is deemed to be extreme pornographic material is not made available to any user. The BBFC is the designated regulator. These requirements will come into force shortly.

58 The Equality and Human Rights Commission. Equality Act 2010. Available at: https://www.equalityhumanrights.com/en/equality-act/equality-act-2010

59 The EU's Audiovisual Media Services Directive 2010 provides Ofcom with the power to regulate editorial content (programming) on UK 'video-on-demand' services – overseeing compliance on content requirements that cover protecting under 18s, preventing incitement to hate, and commercial references in programmes.

60 The EU's revised Audiovisual Media Services Directive (2018) will place requirements on 'video sharing platforms' to take 'appropriate measures' to protect minors from harmful content, protect the general public from illegal content and content that incites violence and/or hatred, and will introduce basic requirements around advertising. A regulator is still being selected, and these requirements are scheduled to come into force by September 2020.

61 The Gambling Act 2005 provides the Gambling Commission with powers to license and regulate all forms of gambling, including online gambling.

62 From May 2019, the Gambling Commission will bring in changes that mean that age and identity must be verified before consumers can deposit money and gamble, and will require age verification before customers can access free-to-play demo games.

in their social media posts where they have been paid to post content online. The CMA is now examining the responsibility of social media platforms to ensure that paid-for content is always properly disclosed.

- Secondary tickets – as a result of action by the CMA, including court proceedings against Viagogo, consumers will always receive essential information before they purchase a ticket from online resale platforms, in particular if there is a risk that the consumer will not be able to get into the event or venue. The court order secured against Viagogo also requires that 'pressure selling' messages are removed from their website.
- Online hotel booking – the CMA recently agreed changes with companies in the Booking.com and Expedia corporate groups in relation to potentially misleading online practices. These include new requirements to be clear about the role that commission plays in the order of search results and that any claims about the limited availability of hotel rooms are accurate and do not risk misleading consumers.

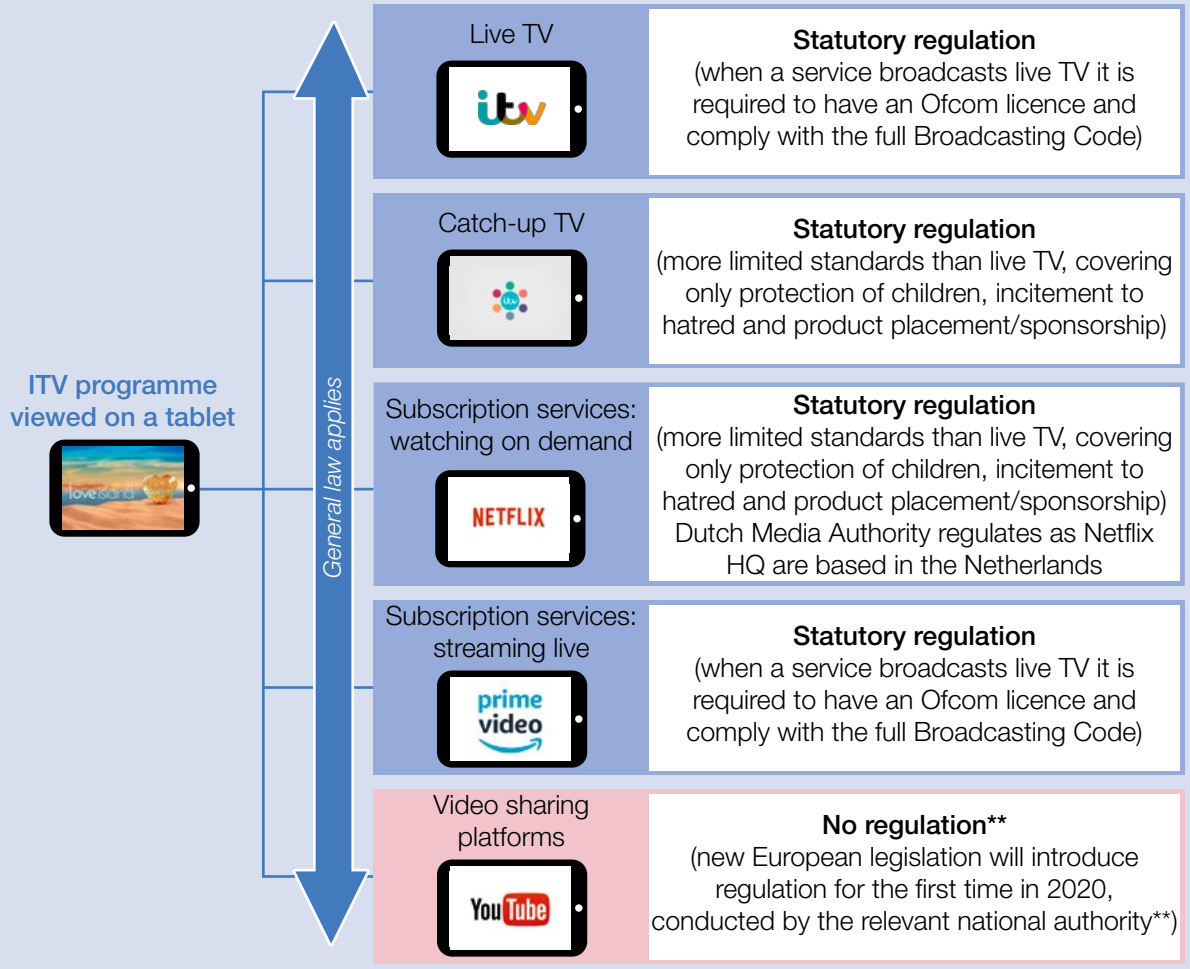
2.6 Under the current liability regime, which is derived from the EU's e-Commerce Directive, platforms are protected from legal liability for any illegal content they 'host' (rather than create) until they have either actual knowledge of it or are aware of facts or circumstances from which it would have been apparent that it was unlawful, and have failed to act 'expeditiously' to remove or disable access to it. In other words, they are not liable for a piece of user-generated illegal content until they have received a notification of its existence, or if their technology has identified such content, and have subsequently failed to remove it from their services in good time.

2.7 For illegal harms, it is also important to make sure that criminal law applies online in the same way as it applies offline. In February 2018 the Prime Minister announced a review by the Law Commission of the law in relation to abusive and offensive online communications, to highlight any gaps in the criminal law which cause problems in tackling this abuse. In its scoping report last year, the Law Commission concluded that behaviour is broadly criminalised to the same extent online as offline and recommended a clarification of existing communication offences. The government is now finalising the details of the second phase of the Law Commission work.

2.8 For legal harms, the same piece of content can be subject to different regulatory standards depending on the platform on which it appears. Ofcom's report Addressing Harmful Content Online sets out how the same programme would be regulated to differing degrees depending on whether it is broadcast on TV, viewed on-demand, or on an online video sharing platform (see Box 18). This means that there are significant gaps in consumer protection.[63]

63 Ofcom (2018). Addressing Harmful Online Content. Available at: https://www.ofcom.org.uk/phones-telecoms-and-internet/information-for-industry/online-policy-research/addressing-harmful-online-content

Regulation of the same content on different services Box 18



Source: Addressing Harmful Content Online, Ofcom 2018.

Voluntary approaches

2.9 Beyond this range of regulatory requirements, the government's Internet Safety Strategy Green Paper, published in 2017, focused on a voluntary approach to countering harmful behaviour and content online. The green paper recognised that government alone cannot keep citizens safe from online harms, and sought to work in close partnership with industry to put in place specific technical solutions to make social media platforms safer.

2.10 Voluntary initiatives between government, industry and civil society are promising in some areas, and the leading companies have taken a number of steps to improve their platforms, for example as set out in Boxes 19-21. We are clear that the progress made on terrorism and CSEA through this voluntary cooperation with the industry must continue, alongside the development of a new regulatory framework.

Existing initiatives to tackle online harms: Global Internet Forum to Counter Terrorism — Box 19

Following the Westminster terrorist attack in March 2017, the government convened a roundtable with major industry players, including Facebook, Twitter, Google and Microsoft to see what more could be done to tackle terrorist content online. This led to these companies setting up the Global Internet Forum to Counter Terrorism (GIFCT) in June 2017.

The GIFCT is leading the cross-industry response to reduce the availability of terrorist content on the internet so that there are no safe spaces for terrorists online. Key objectives for the Forum are to increase the use of automation and machine learning technology to detect and remove terrorist content – ultimately preventing terrorist content being made available to users in the first place – and supporting smaller, less well-resourced companies to tackle these threats on their own platforms.

The Forum has taken some positive steps since its establishment, but there is still much more to do. The government wants to see an ambitious and tangible plan for delivery. Our aims for the GIFCT in 2019 are for the Forum to:

- Expand its membership, securing a greater range and quantity of companies to sign up as members of the Forum.
- Devote greater efforts to targeted interventions with priority platforms, including through the development and sharing of automated technology.
- Put in place a clear programme of activity, providing metrics against which success can be measured.
- Provide greater visibility to drive this agenda forward, including companies having a clearer public voice on the issue.

Existing initiatives to tackle online harms: UK Council for Internet Safety — Box 20

The UK Council for Internet Safety (UKCIS) is a new collaborative forum through which government, the tech community and civil society work together to ensure the UK is the safest place in the world to be online.

Expanding the scope of the former UK Council for Child Internet Safety (UKCCIS), UKCIS works to tackle online harms such as hate crime, extremism and violence against women and girls, in addition to maintaining a focus on the needs of children.

Priority areas of delivery for UKCIS over the next year include:

- Producing a landscape review of research around adult online harms, and regular concise summaries of emerging research.
- Updated guidance to schools on sexting, and evaluation of online safety provision, and for Initial Teacher Training providers to help them upskill new teachers in online safety.

- Promoting the Connected World framework, which describes the digital knowledge and skills that children should have the opportunity to develop at different stages of their lives.
- A digital resilience framework and toolkit to help families, educators, policymakers, frontline service workers and the industry better support users online, across a wide range of harms.

Existing initiatives to tackle online harms: WePROTECT Global Alliance **Box 21**

The WePROTECT Global Alliance (WPGA) was established in recognition that CSEA is a global crime requiring a global response.

The UK government played a key role in establishing WPGA and is its sole financial donor. WPGA aims to protect more children, apprehend more perpetrators of abuse and make the internet free from child sexual exploitation. Eighty-five countries are members of WPGA, along with 20 global technology companies and 25 leading non-governmental organisations.

The success of the UK government funded WPGA is that it has brought together government, law enforcement, industry and civil society to take a stand against online child sexual exploitation.

2.11 In the Government Response to the Internet Safety Strategy Green Paper consultation, we noted that only a relatively small group of the larger companies are engaged with the government's work on online safety, even though online harms can and do occur across many websites. There is also a wide variation in the extent, efficacy and pace of actions by companies to tackle online harms. Some companies rely on user moderation to oversee reported violations of their terms and conditions, such as Reddit; others employ teams of moderators or deploy technology to monitor content, such as Facebook.

2.12 Many companies claim to hold a strong track record on online safety but there is limited transparency about how they implement or enforce their policies, and there is a persistent mismatch with users' experiences – 70% of Britons believe that social media companies do not do enough to prevent illegal or unethical behaviours on their platforms.[64] 60% of respondents to our Internet Safety Strategy Green Paper consultation had witnessed inappropriate or harmful behaviour online; only 41% thought their reported concerns were taken seriously by social media companies.[65]

2.13 At present many online companies rely on using their terms and conditions as the basis by which to judge complaints. In practice however, companies' terms and conditions are often difficult for users to understand, and safety policies are not consistent across different

64 Edelman (2018). Edelman Trust Barometer – UK Findings. Available at: https://www.edelman.co.uk/magazine/posts/edelman-trust-barometer-2018/

65 HM Government (2018). Government Response to the Internet Safety Strategy Green Paper. May 2018. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf

platforms, with take-down times, description of harms and reporting processes varying. A series of investigations have highlighted the risk of serious shortcomings in the training, working conditions and support provided for content moderators.[66]

2.14 There is no mechanism to hold companies to account when they fail to tackle breaches. There is no formal, wide-reaching industry forum to improve coordination on terms and conditions. The absence of clear standards for what companies should do to tackle harms on their services makes it difficult for users to understand or uphold their rights.

2.15 The government believes that voluntary efforts have not led to adequate or consistent steps to protect British citizens online. As highlighted above, users' own experiences confirm a sense of vulnerability online.

An international approach

2.16 The threat posed by harmful and illegal content and activity online is a global one, and many of our international partners are also developing new regulatory approaches to tackle online harms. Box 22 sets out what some other countries are doing in this area.

International approaches to countering online harms — Box 22

Germany adopted its Network Enforcement Act ('NetzDG') in 2017. This law requires online platforms with more than two million registered users in Germany to remove 'manifestly unlawful' content, which contravenes specific elements of the German criminal code, such as holocaust denial and hate speech, within 24 hours of receiving a notification or complaint, and to remove all other 'unlawful' content within seven days of notification. Non-compliance risks a fine of up to €50 million. This law also seeks to increase platform responsibility through imposing greater transparency and significant reporting obligations.

Australia established an eSafety Commissioner through its Enhancing Online Safety for Children Act in 2015. The eSafety Commissioner is responsible for promoting online safety for all Australians. As well as offering a complaints service for young people who experience serious cyber bullying, its remit includes identifying and removing illegal online content and tackling image-based abuse.

The European Commission, led by DG JUST, published in September 2018 a proposal on preventing the dissemination of terrorist content online – Member States agreed a Council version of the text in December 2018. The aim of the proposal is to ensure a consistent approach across industry to the removal of online terrorist content by Hosting Service Providers, for example social media platforms and video sharing sites. There are similarities in the approach taken to the framework proposed in this White Paper – as currently drafted it looks to take a proportionate approach to setting requirements, introduce duties of care on companies, and implementing a transparency framework.

Over 2018, the EU Commission, led by DG CNECT, also published its Action Plan against Disinformation. The Commission collaborated with companies including Facebook, Google and Twitter to produce a code of practice against disinformation. This resulted in commitments to improve the transparency of political advertising, prevent the misuse

66 The Verge (2019). The Trauma Floor. Available at: https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

of automated bots and to invest in tools to amplify diverse perspectives. The UK government has previously indicated its support for the measures and will continue to collaborate internationally on this issue.

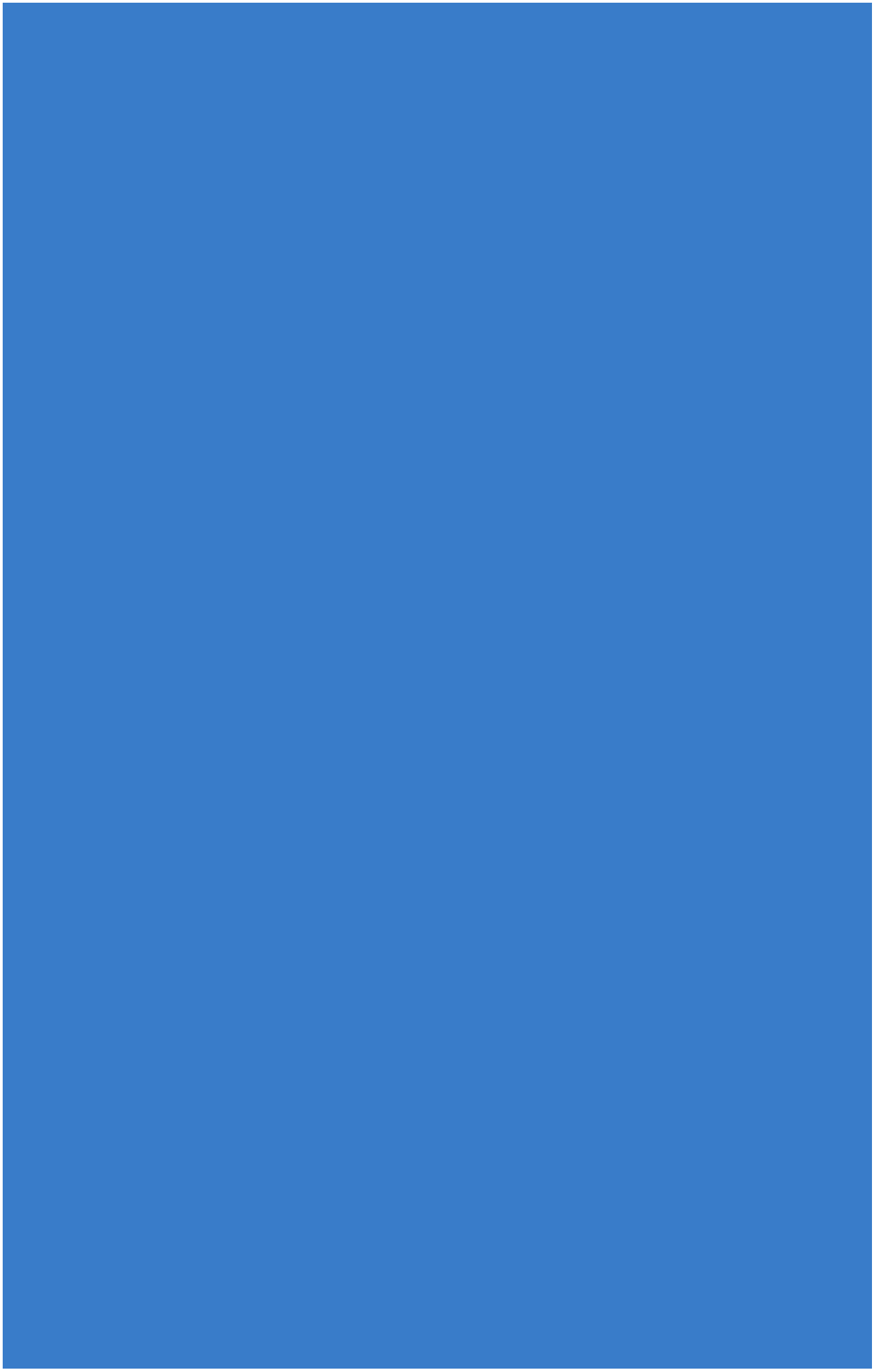
2.17 The government is working closely with international partners as we develop our own approach that reflects our shared values and commitment to a free, open and secure internet. The approach proposed in this White Paper is the first attempt globally to tackle this range of online harms in a coherent, single regulatory framework. We will continue to share experiences and seek to work with international partners. Further details are set out in Chapter 6.

Existing initiatives to tackle online harms: Project Arachnid **Box 23**

The government has invested £600,000 into Project Arachnid, a groundbreaking project that trawls the web to identify web pages with suspected child sexual abuse material. The technology can be deployed across websites, forums, chat services and newsgroups to instantaneously detect illegal content, before sending a take-down notice to service providers so they can quickly protect children from further exploitation.

Project Arachnid is the product of a partnership with the Canadian Centre for Child Protection and the US National Center for Missing and Exploited Children, demonstrating the UK's determination to work with international partners to tackle harmful activity online. To date, Arachnid has trawled 1.5 billion webpages, detected 7.5 million suspected images of child sexual abuse and issued more than 1 million take-down notices for the removal of child abuse material on the open web[67].

67 Live dashboard data, data taken on 21 March 2019.



PART 2: Regulatory model

3. A new regulatory framework

Summary

- The government will establish a new statutory duty of care to make companies take more responsibility for the safety of their users and tackle harm caused by content or activity on their services. Compliance with this duty of care will be overseen and enforced by an independent regulator.
- Companies must fulfil their new legal duties. The regulator will set out how to do this in codes of practice. If companies want to fulfil these duties in a manner not set out in the codes, they will have to explain and justify to the regulator how their alternative approach will effectively deliver the same or greater level of impact.
- Regarding the threat to national security or the physical safety of children, the government will have the power to direct the regulator in relation to codes of practice relating to terrorist activity or CSEA online, and these codes must be signed off by the Home Secretary.
- For all codes of practice relating to illegal harms, including incitement of violence and the sale of illegal goods and weapons, there will be a clear expectation that the regulator will work with law enforcement and other relevant government agencies to ensure the codes adequately keep pace with the threat.
- Developing a culture of transparency, trust and accountability will be a critical element of the new regulatory framework. The regulator will have the power to require annual transparency reports from companies in scope, outlining the prevalence of harmful content on their platforms and what measures they are taking to address this. These reports will be published online by the regulator, so that users and parents can make informed decisions about online use. The regulator will also have powers to require additional information, including about the operation of algorithms.
- The regulator will encourage and oversee the fulfilment of companies' commitments to improve the ability of independent researchers to access their data, subject to appropriate safeguards.

- As part of the new duty of care, we will expect companies, where appropriate, to have an effective and easy-to-access user complaints function. The regulator will require companies to respond to user complaints within an appropriate timeframe and to take action consistent with the expectations set out in the regulatory framework.
- But we also recognise the importance of an independent review mechanism to ensure that users have confidence that their concerns are being treated fairly. We are consulting on allowing designated bodies to make 'super complaints' to defend the needs of users.
- Ahead of the implementation of the new regulatory framework, we will encourage companies to take early action to address online harms. To assist this, the White Paper sets out high-level expectations of companies, including some specific expectations in relation to certain harms. We expect the regulator to reflect these in future codes of practice.
- Where there is a threat to national security or the physical safety of children, such as CSEA and terrorism, we will expect companies to go much further and demonstrate the steps taken to combat the dissemination of associated content and illegal behaviours. We will publish interim codes of practice providing guidance about tackling terrorist activity and online CSEA later this year.

3.1 The government will establish a new statutory duty of care on relevant companies to take reasonable steps to keep their users safe and tackle illegal and harmful activity on their services.

3.2 The fulfilment of this duty will be overseen and enforced by an independent regulator.

3.3 This statutory duty of care will require companies to take reasonable steps to keep users safe, and prevent other persons coming to harm as a direct consequence of activity on their services. This broader application of the duty, beyond simply users of a particular service, recognises that in some cases the victims of harmful activity – victims of the sharing of non-consensual images, for example – may not themselves be users of the service where the harmful activity took place. This duty will apply to all of the harms included in the scope of the White Paper, as set out below.

3.4 A key element of the regulator's approach will be the principle of proportionality. Companies will be required to take action proportionate to the severity and scale of the harm in question. The regulator will be required to assess the action of companies according to their size and resources, and the age of their users.

3.5 The regulatory approach will impose more specific and stringent requirements for those harms which are clearly illegal, than for those harms which may be legal but harmful, depending on the context.

3.6 Companies must fulfil their new legal duties. The regulator will set out how to do this in codes of practice. The codes will outline the systems, procedures, technologies and investment, including in staffing, training and support of human moderators, that companies need to adopt to help demonstrate that they have fulfilled their duty of care to their users.

Companies will still need to be compliant with the overarching duty of care even where a specific code does not exist, for example assessing and responding to the risk associated with emerging harms or technology.

3.7 There will be a strong expectation that companies follow the guidance set out in these codes. If they choose not to do so, companies will have to explain and justify to the regulator how their alternative approach will effectively deliver the same or greater level of impact. This approach is familiar to companies, for example the UK Corporate Governance Code[68] and the ICO's code of practice on data sharing. Though these codes will be developed with the companies and other stakeholders in an open and transparent way, the regulator will ultimately decide on their content.

3.8 The regulator will assess whether companies have fulfilled their duty of care, including by reference to relevant codes of practice, and compliance with the company's own relevant terms and conditions. Failure to meet these obligations may result in enforcement action by the regulator. Further details on enforcement are in Chapter 6.

3.9 The regulator will also expect companies to make clear how they are fulfilling their statutory duty of care. Relevant terms and conditions will be required to be sufficiently clear and accessible, including to children and other vulnerable users. The regulator will assess how effectively these terms are enforced as part of any regulatory action.

Terrorism and CSEA online

3.10 Companies will be required to take particularly robust action to tackle terrorist use of the internet and online CSEA. The government will have the power to issue directions to the regulator regarding the content of the codes of practice for these harms, and will also approve the draft codes before they are brought into effect. Similarly, the regulator will not normally agree to companies adopting proposals which diverge from these two codes of practice, and will require a high burden of proof that alternative proposals will be effective.

3.11 Between the publication of this White Paper and the establishment of a regulator, the government will work with law enforcement and other relevant bodies to produce interim codes of practice for online terrorist content and CSEA. These codes will be published later this year.

General monitoring

3.12 The regulator will not compel companies to undertake general monitoring of all communications on their online services, as this would be a disproportionate burden on companies and would raise concerns about user privacy. The government believes that there is however, a strong case for mandating specific monitoring that targets where there is a threat to national security or the physical safety of children, such as CSEA and terrorism.

Transparency, trust and accountability

3.13 Developing a culture of transparency, trust and accountability, and consistent standards of transparency, will be a critical element of the new regulatory framework.

68 Financial Reporting Council (2018). UK Corporate Governance Code. Available at: https://www.frc.org.uk/directors/corporate-governance-and-stewardship/uk-corporate-governance-code

3.14 In May 2018, the Government Response to the Internet Safety Strategy Green Paper consultation set out the role transparency and reporting must play in building our understanding of the extent of online harms and how effectively companies are tackling breaches in their terms and conditions.

3.15 Alongside this response, we published a draft transparency reporting template and began a series of engagements with industry. This process, which has included discussion with over 20 companies, has provided some helpful insights into current industry action. It is encouraging that more companies have since started publishing their own global transparency reports. We will publish the government's first annual transparency report later this year.

3.16 At the same time, we indicated that transparency reporting was one of the potential areas for new legislation. Greater transparency will ensure:

- The regulator can gain an understanding of the level of harms on online platforms and the mitigating action being taken by companies. This will inform its regulatory priorities and determine the effectiveness of, and compliance with, different regulatory measures.
- Users can gain a greater understanding and awareness of whether and to what extent companies are taking positive steps to keep their users safe, and the processes different companies have in place to prevent harms.
- Companies take responsibility for the impacts of their platforms and products on their users. It will incentivise accountability within the industry.

3.17 To inform its reports and to guide its regulatory action, the regulator will have the power to require annual reports from companies covering the following areas:

- Evidence of effective enforcement of the company's own relevant terms and conditions, which should reflect guidance issued by the regulator in its codes of practice.
- Processes that the company has in place for reporting illegal and harmful content and behaviour, the number of reports received and how many of those reports led to action.
- Proactive use of technological tools, where appropriate, to identify, flag, block or remove illegal or harmful content.
- Measures and safeguards in place to uphold and protect fundamental rights, ensuring decisions to remove content, block and/or delete accounts are well-founded, especially when automated tools are used and that users have an effective route of appeal.
- Where relevant, evidence of cooperation with UK law enforcement and other relevant government agencies, regulatory bodies and public agencies.
- Details of investment to support user education and awareness of online harms, including through collaboration with civil society, small and medium sized enterprises (SMEs) and other companies.

3.18 The regulator will produce and publish an annual transparency report outlining key data on companies' performance against their duty of care and the prevalence of harms on different platforms. It will also publish companies' transparency reports on its website, ensuring these are easily accessible to the public so that users and parents can make

informed decisions about online use. Where the regulator has required companies to produce transparency reports, it will be mandatory to provide them; failure to do so will result in enforcement action (as set out in Chapter 6).

3.19 The regulator will use insight from users, civil society, government, law enforcement and other relevant government agencies, and other regulators to inform its understanding of the prevalence and impact of online harms, and the effectiveness of companies' responses.

3.20 As well as the power to require annual reports from companies, the regulator will have the power to require additional information from them to inform its oversight or enforcement activity, and to establish requirements to disclose information. It may also undertake thematic reviews of areas of concern, for example a review into the treatment of self-harm or suicide related content. The regulator will have the power to require companies to share research that they hold or have commissioned that shows that their activities may cause harm.

3.21 The regulator will build on government engagement with companies to understand how best to establish comparable data-points and reporting between platforms.

3.22 As part of a movement towards greater transparency, companies should also work in conjunction with the regulator to build a shared understanding of the mechanics of their associated platforms or services. Where necessary, to establish that companies are adequately fulfilling the duty of care, the regulator will have the power to request explanations about the way algorithms operate. The regulator may, for example, require companies to demonstrate how algorithms select content for children, and to provide the means for testing the operation of these algorithms.

3.23 In determining where such explanations will be appropriate and what form they should take, the regulator will work closely with the Centre for Data Ethics and Innovation, the expert body that has been set up to advise government on the regulation of data, including algorithmic tools. Appropriate safeguards will be needed to ensure commercial confidentiality, although the regulator is unlikely to require direct access to companies' proprietary codes if necessary explanations have been provided.

3.24 Several of the largest companies have promised access for independent researchers to anonymised information, in line with data protection requirements. This is a positive step, although it is as yet unclear whether these promises have been fulfilled. The government welcomes these steps, and believes that this level of transparency to researchers is a necessary part of developing the increased understanding of online harms. We will task the regulator with encouraging this approach, and ensuring companies make relevant information available.

3.25 We will expect the regulator to foster a culture of cooperation between companies and to encourage companies, especially the larger ones, to share information about online harms. Users perpetrating harm often move between platforms, especially to behave illegally and disseminate illegal content. A greater level of cooperation between platforms by sharing observations and best practices to prevent harms spreading from one provider to another will be essential.

Consultation questions

Question 1: This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

User redress

3.26 Many companies claim a strong track record on online safety, but responses to our Internet Safety Strategy Green Paper showed that this is at odds with users' experiences. To fulfil the new duty of care, we will expect companies, where appropriate, to have an effective and easy-to-access complaints function, allowing users to raise either concerns about specific pieces of harmful content or activity, or wider concerns that the company has breached its duty of care. Users should receive timely, clear and transparent responses to their complaints, and there must be an internal appeals function. The regulator will have oversight of these processes, including through transparency information about the volume and outcome of complaints, and the power to require improvements where necessary. Box 24 explains users' rights under the proposed requirements.

3.27 In addition to the internal appeals processes, we recognise that independent review or resolution mechanisms may be appropriate in some circumstances. This would increase the accountability of companies and help rebuild users' trust. We are consulting on the following option:

- Whether a provision should be made in legislation for designated bodies to bring 'super complaints' to the regulator for consideration, in specific and clearly evidenced circumstances. This could be an important safeguard in the user redress process and we are also consulting on when such complaints would be appropriate and most effective, and on the bodies or groups that may be empowered to bring them.

3.28 We would also welcome views during the consultation process on additional options for redress.

Consultation questions

Question 2: Should designated bodies be able to bring 'super complaints' to the regulator in specific and clearly evidenced circumstances?

Question 2a: If your answer to question 2 is 'yes', in what circumstances should this happen?

Question 3: What, if any, other measures should the government consider for users who wish to raise concerns about specific pieces of harmful content or activity, and/or breaches of the duty of care?

3.29 Under current arrangements, individuals can, in principle, obtain remedies in court against companies where they are negligent or breach their contract with the individual but such legal actions can face difficulties. For example, difficulties in establishing the company's duty of care to the person bringing the claim, showing a causal link between their activities and harm caused, or obtaining factual evidence. Our regulatory model will provide evidence and set standards which may increase the effectiveness of individuals' existing legal remedies.

3.30 The regulator's primary role in the user redress process will be to oversee the requirement on relevant companies to have appropriate and effective internal complaints processes, including consideration of whether there should be an appeals function in certain circumstances. The regulator would also determine any 'super complaints' process and designate bodies. We do not envisage a role for the regulator itself in determining disputes between individuals and companies, but where users raise concerns with the regulator, it will be able to use this information as part of its consideration of whether there may be systemic failings which justify enforcement action. We will also require the regulator to take the interests of users into consideration.

Regulation in practice: User redress: how the regulatory framework will work for individuals Box 24

At the moment, individuals can raise complaints and concerns about harmful online activity with companies, but processes vary and provision is patchy across the industry. Some companies do not have effective means to address user concerns, and it is not always clear what response, if any, a user will receive. Only two in five respondents to the government's consultation on the Internet Safety Strategy felt their concerns were taken seriously by social media companies. The regulatory framework proposed in this White Paper will give individuals new avenues to pursue complaints:

1. Right to an internal complaints procedure that meets standards set out by the regulator. Where appropriate, companies covered by the regulator will be required to have an effective complaints process, and the regulator will set minimum standards for these processes. This means that users will know how they can raise a complaint, how long it will take a company to investigate, and what response they can expect (including appeal rights).
2. Right of redress through an independent process. If the company is unable or unwilling to resolve a complaint, or the user is not satisfied with the response, it may be appropriate for users to be able to seek redress through an independent process. We are seeking views on how this could work in practice, including whether the regulator should run a 'super complaints' scheme, through which designated organisations could raise issues with the regulator on behalf of users.
3. The ability to alert the regulator to an alleged breach of a company's duty of care. While the regulator would not normally adjudicate on individual complaints about companies, users will be able to report concerns to the regulator. This will be an important part of the regulator's horizon scanning to identify where companies might not be fulfilling their duty of care to their users.
4. The scope to use the regulator's findings in any claim against a company in the courts on grounds of negligence or breach of contract. And, if the regulator has found a breach of the statutory duty of care, that decision and the evidence that has led to it will be available to the individual to use in any private legal action.

Role of Parliament

3.31 It will be important to ensure that Parliament is able to scrutinise the regulator's work. Mechanisms for achieving this will depend in part on whether the regulator is a new or existing body but are likely to include, for example, a duty on the regulator to lay its annual report and audited accounts before Parliament. The regulator will also have a general responsibility to provide Parliament with information about its work, as requested.

3.32 In addition, we will consider what role Parliament should have in relation to the regulator's codes of practice. Parliament's role in relation to codes of practice and guidance issued by other regulators varies across different regulatory regimes, ranging from formal approval to no specific role. We will consider options for the role of Parliament as we develop these proposals in more detail.

Consultation questions

Question 4: What role should Parliament play in scrutinising the work of the regulator, including the development of codes of practice?

4: Companies in scope of the regulatory framework

Summary

- The regulatory framework will apply to companies that provide services or tools that allow, enable or facilitate users to share or discover user-generated content, or interact with each other online.
- These services are offered by a wide range of companies, including start-ups and SMEs, and other organisations such as charities.
- The application of the regulatory requirements and the duty of care model will reflect the diversity of organisations in scope and ensure a risk-based and proportionate approach. We will minimise excessive burdens, particularly on small businesses and civil society organisations.
- Reflecting the importance of privacy, any requirements to scan or monitor content for tightly defined categories of illegal content will not apply to private channels. We are consulting on definitions of private communications, and what measures should apply to these services.

4.1 Harmful content and behaviour originates from and migrates across a wide range of online platforms or services, and these cannot readily be categorised by reference to a single business model or sector. Focusing on the services provided by companies, rather than their business model or sector, limits the risk that online harms simply move and proliferate outside of the ambit of the new regulatory framework. We propose that the regulatory framework should apply to companies that allow users to share or discover user-generated content, or interact with each other online.

4.2 There are two main types of online activity that can give rise to the online harms in scope or compound their effects:

- Hosting, sharing and discovery of user-generated content (e.g. a post on a public forum or the sharing of a video).
- Facilitation of public and private online interaction between service users (e.g. instant messaging or comments on posts).

4.3 A wide variety of organisations provide these services to users. This will mean that companies of all sizes will be in scope of the regulatory framework. The scope will include companies from a range of sectors, including social media companies, public discussion forums, retailers that allow users to review products online, along with non-profit organisations, file sharing sites and cloud hosting providers.

4.4 This comprehensive approach is important for the efficacy of the new regulatory framework.

4.5 We also recognise the importance of minimising undue burdens on organisations in scope and of avoiding uncertainty about how regulation will apply. To ensure a proportionate approach and avoid being overly burdensome, the application of the regulatory requirements and the duty of care model will reflect the diversity of organisations in scope, their capacities, and what is technically possible in terms of proactive measures, including for those providing ancillary services such as caching (the process of temporarily storing data in either a

software or hardware 'cache'). While we will minimise excessive burdens according to the size and resources of organisations, all companies will be required to take reasonable and proportionate action to tackle harms on their services. The regulator will ensure that there is clarity about what the regulatory regime means in practice for different company's, and will not impose new requirements where there is no evidence of harm. A range of proposed initiatives to counter regulatory burdens are set out in Chapter 6.

Regulatory approach to private communications

4.6 Defining 'private' and 'public' in the online space is complex from a technical and legal standpoint. For example, there is an obvious difference between one-to-one messaging, and a WhatsApp group of several hundred users. However, users should be protected from harmful content or behaviour wherever it occurs online, and criminals should not be able to exploit the online space to conduct illegal activity. The development of harmful activity online frequently involves a combination of activity taking place on both public and private communication channels. For example, terrorist propaganda is often disseminated over public channels, with activities such as the preparation of terrorist attacks occurring largely on private channels. Such private channels are also widely used to store and share images of CSEA, or to groom young children, with public channels frequently being where initial contact with a child takes place (see Box 25).

4.7 Reflecting the importance of privacy, the framework will also ensure a differentiated approach for private communication, meaning any requirements to scan or monitor content for tightly defined categories of illegal content will not apply to private channels.

4.8 We are consulting on appropriate definitions and what regulatory requirements can and should apply to private communication services alongside this White Paper.

Harm: Child sexual exploitation and abuse – how online grooming moves across different platforms **Box 25**

Evidence shows how grooming activity often migrates across platforms, luring children into less public spaces online:

- Initial contact with a child is often made after they are identified as a potential victim by a groomer on public social media platforms.
- Offenders may target children based on vulnerabilities such as mental health, or by exploiting publicly available information from their social media profiles.
- The grooming process can be extensive; building rapport and manipulating the victim – but it can also move almost immediately into sexual advances.
- This can involve the groomer then sending the child a message, using the same platform's private messaging service or another private or encrypted messaging service, seeking to extort indecent imagery and continue their abuse.

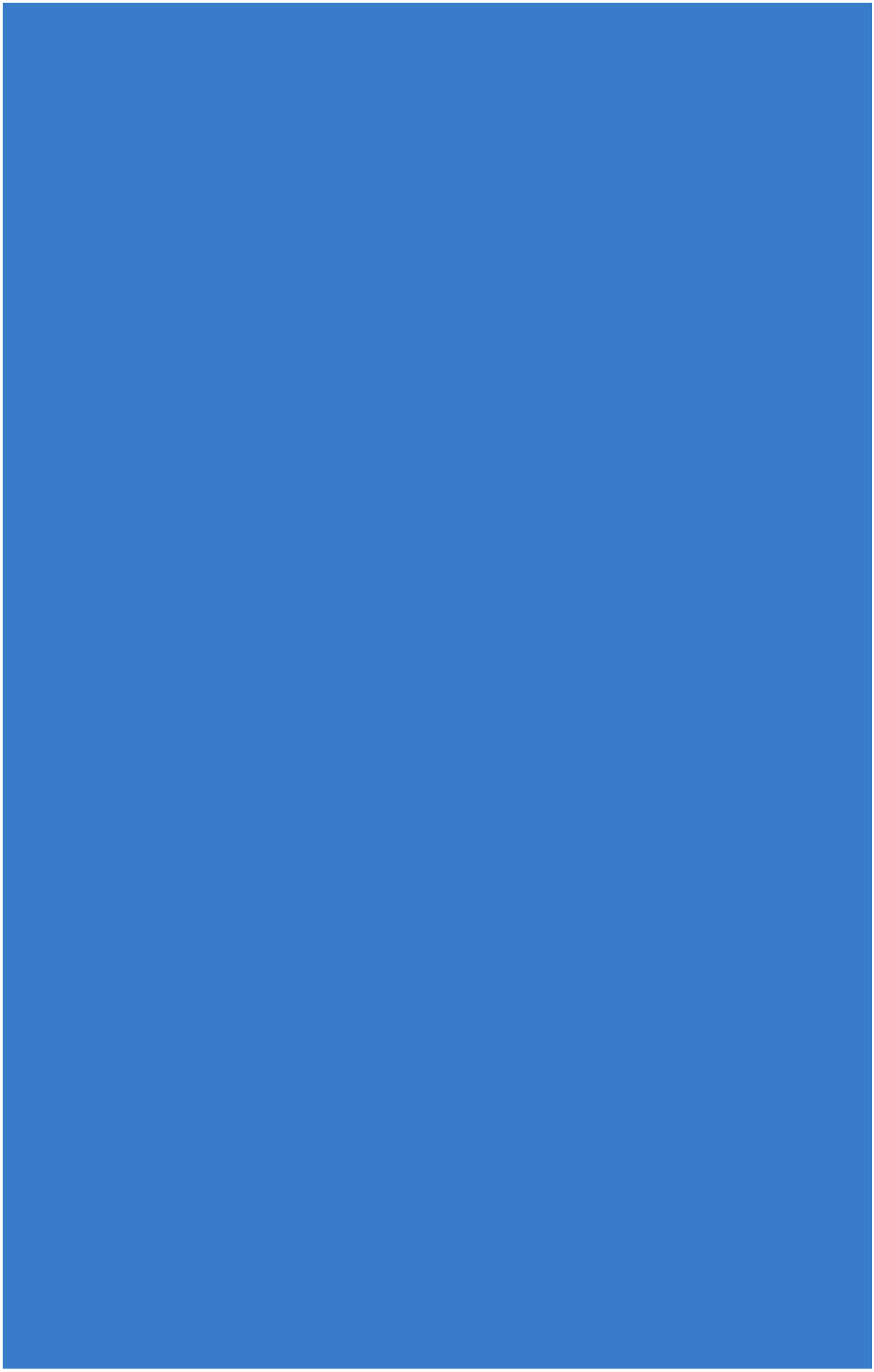
Consultation questions

Question 5: Are proposals for the online platforms and services in scope of the regulatory framework a suitable basis for an effective and proportionate approach?

Question 6: In developing a definition for private communications, what criteria should be considered?

Question 7: Which channels or forums that can be considered private should be in scope of the regulatory framework?

Question 7a: What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?



PART 3: Regulation in practice

5: A regulator for online safety

Summary

- An independent regulator will implement, oversee and enforce the new regulatory framework. It will have sufficient resources and the right expertise and capability to perform its role effectively.
- The regulator will also have broader responsibilities to promote education and awareness-raising about online safety, and to promote the development and adoption of safety technologies to tackle online harms.
- The regulator will take a risk-based approach, prioritising action to tackle activity or content where there is the greatest evidence or threat of harm, or where children or other vulnerable users are at risk.
- To support this, the regulator will undertake and commission research to improve the evidence base, working closely with UK Research and Innovation (UKRI) and other partners.
- The regulator will take a proportionate approach, expecting companies to do what is reasonable, depending on the nature of the harm and the resources and technology available to them.
- The regulator will have a legal duty to pay due regard to innovation, and to protect users' rights online, being particularly mindful to not infringe privacy and freedom of expression.
- The government is consulting on whether the regulator should be a new or existing body. The regulator will be funded by the industry in the medium term, and the government is exploring options such as fees, charges or an industry levy to put it on a sustainable footing.

The functions of the regulator

5.1 The regulatory framework will be implemented, overseen and enforced by an independent regulator. This regulator will be equipped with the powers, resources and expertise it needs to effectively carry out its role.

5.2 The regulator's functions will include:

- Setting out what companies need to do to fulfil the duty of care, including through codes of practice.
- Establishing a transparency, trust and accountability framework, backed by information-gathering powers, to assess companies' compliance with the duty of care and their own relevant terms and conditions.
- Providing support to start-ups and SMEs to help them fulfil their legal obligations in a proportionate and effective manner.
- Overseeing the implementation of user redress mechanisms.
- Taking prompt and effective enforcement action in the event of non-compliance (as set out in Chapter 6).
- Promoting education and awareness-raising about online safety to empower users to stay safe online.
- Promoting the development and adoption of safety technologies to tackle online harms.
- Undertaking and commissioning research to improve our understanding of online harms and their impacts on individuals and society.

A risk-based approach

5.3 The government will require the regulator to adopt a risk-based approach, prioritising regulatory action to tackle harms that have the greatest impact on individuals or wider society. This will shape the development of codes of practice, monitoring and review of online harms, the regulator's work with industry to develop technological solutions, and enforcement action.

5.4 The regulator will also focus on companies where there is the greatest risk of harm, based on factors such as the type of service – for example, services that enable adult users to contact children, services that have large user bases, and services that target or are popular with vulnerable groups of users. It will also use evidence of the actual incidence of harms on different services and the safety track record of different companies to prioritise its resources. The regulator will use its powers to conduct thematic reviews, undertake targeted horizon scanning and investigate specific issues to develop its understanding of the risk landscape.

5.5 This risk-based approach will mean that the regulator's initial focus in the first phase will be on those companies which pose the biggest and most obvious risk of harm to users, either because of the scale of the service's size or because of known issues with serious harms. We expect the regulator to take a proactive approach to assessing compliance in these cases, whereas their approach to the full range of companies in scope would be focused on providing advice and guidance and taking reactive action in response to concerns. This is consistent with the approach in a number of other regulatory regimes, including health and safety and financial services.

5.6 The duty of care approach will also mean companies must improve their understanding of the risks associated with their services and take effective and proportionate steps to mitigate these risks. These steps should be in keeping with the codes of practice set down by the regulator. When assessing compliance, the regulator will need to consider whether the harm was foreseeable, and therefore what is reasonable to expect a company to have done. In the event of a new risk emerging, the company should notify the regulator in order to discuss the best approach to mitigation and to share learning across companies.

A proportionate approach

5.7 The regulator will take account of the capacity of companies to meet regulatory requirements, including the reach of their platforms in terms of user-base and the severity of the harms. This proportionate approach will also be enshrined in the legislation by making clear that companies must do what is 'reasonably practicable' – a test that has underpinned the success of health and safety legislation. However, all companies within scope will be required to take reasonable and proportionate action to tackle harms on their services, and the regulator will set clear expectations of what companies should do to tackle illegal activity and to keep children safe online.

5.8 We expect the regulator to comply with principles of regulatory best practice, which means that its activities will be sensitive to impacts on competition and small and micro-businesses in particular (see Box 26).

5.9 The regulator will also be required to support less well-resourced companies, as part of its work to develop tools to build capacity amongst companies and users. For example, we expect the regulator to work with the industry to encourage the development of technologies that aid compliance, and to facilitate cross-sector collaboration and sharing of expertise. These technologies could be made available to start-up or small companies. This is part of a wider advisory role for the regulator through which it will, for example, provide industry with technical information on best practice content moderation processes, or provide toolkits that explain common patterns of behaviour by cyberstalkers.

5.10 Through the consultation process alongside this White Paper, we intend to work with industry, civil society and the public to look at ways in which we can minimise any excessive burdens and provide additional certainty to businesses, and explore what more the regulator could do to make compliance straightforward and practicable for all businesses.

Regulation in practice: Better regulation principles and the new regulatory framework — Box 26

We need all platforms to take reasonable steps to keep their users safe. Harm can occur on small platforms as well as big ones. There is nowhere on the internet where it is acceptable to host child sexual abuse material or terrorist material.

Regulation can impose a disproportionate burden on smaller companies. Badly designed regulation can stifle innovation by giving an advantage to large companies that can handle compliance more easily. We are determined that this regulatory framework should provide strong protection for our citizens while avoiding placing an impossible burden on smaller companies.

We will take five key steps to achieve this:

1. A proportionate approach. The regulator will take account of the capacity of companies to meet regulatory requirements, including their size and the reach of their platforms in terms of user-base, as well as the risk and prevalence of harms on their service.
2. A duty of innovation. The regulator will have a legal duty to pay due regard to innovation. This will include implementing the framework in a way that does not impose impossible demands on new and challenger companies. This will also ensure that start-ups and those developing innovative new products can work with the regulator, for example through regulatory sandboxes.

3. Making compliance straightforward. The regulator will be tasked with helping start-ups and SMEs fulfil their obligations. We will learn from best practice in other sectors, such as the support provided to companies by the Health and Safety Executive or the ICO.
4. Using technology. Government will work with the regulator to promote effective technological compliance solutions that can be made available to start-ups and small businesses.
5. Minimising compliance costs. We will explore options to streamline compliance, including creating machine executable regulation and facilitating easy, secure data sharing.

A legal obligation to support innovation

5.11 The regulator will have a legal obligation to pay due regard to innovation. A similar obligation was placed on the ICO under the Data Protection Act 2018. This has allowed the regulator to fully implement a robust data protection regime in a pro-innovation way. The ICO currently has plans to establish an initiative that will proactively support organisations to develop innovative products and services that make use of personal data and benefit the public. The ICO will provide this support whether these innovations are at design, proof of concept and testing stages, or as further ongoing development of existing innovative products/services. This is distinct from the legal requirement of data protection by design. We would expect the regulator to explore similar approaches to supporting and encouraging innovation in this space, subject to minimum expectations of user safety. This obligation will encourage the regulator to take a flexible, proportionate and risk-based approach when setting and enforcing expectations and responsibilities for companies.

Protecting users' rights online

5.12 The regulator will also have an obligation to protect users' rights online, particularly rights to privacy and freedom of expression. It will ensure that the new regulatory requirements do not lead to a disproportionately risk averse response from companies that unduly limits freedom of expression, including by limiting participation in public debate. Its regulatory action will be required to be fair, reasonable and transparent.

Empirical approach

5.13 The new regulator will take an evidence-based approach to regulatory activity. It will need to understand the potential impact of technological developments on the companies it regulates, as well as users' experiences of harm. To support this, we expect that it will run a regular programme of user consultation, in-depth research projects, and horizon scanning activity. It will work with companies to ensure that academics have access to company data to undertake research, subject to suitable safeguards. This dynamic approach to evidence gathering will help the regulator to assess the changing nature of harms and the risks associated with them, and of the places and manner in which they manifest online.

5.14 The regulator will work closely with UKRI to ensure support for targeted research into online harms, and to develop the collective understanding of online harms and the evidence base, building on the work of the UKCIS Evidence Group. This will include working with relevant aspects of UKRI's Digital Economy Theme – a partnership between the Engineering and Physical Sciences Research Council (EPSRC), the Arts and Humanities Research Council (AHRC), the Economic and Social Research Council (ESRC) and Innovate UK.

The regulatory body

5.15 An independent regulator could be set up, either by creating a new body, or by altering the remit and functions of an existing organisation. The usual government approval processes would apply to the establishment of a new central government regulatory body. The government is considering:

- A new regulator. Setting up a dedicated new regulator would provide a clear and coherent remit to focus on online safety and provide new leadership of online safety to industry and the public. A new body would, however, be more costly to set up and take longer to become operational and risks further complicating the regulatory landscape.
- An existing regulator. Tasking an existing regulator to assume responsibility for online safety would mean that the new regime would start with regulatory credibility and make the best use of existing experience and expertise. We would assess existing regulators' suitability based on their current responsibilities, the type of regulation they are already responsible for, their track record of working successfully within complex sectors, and their capacity to take on new responsibilities for online safety, including compatibility with their current legal status and operating model.

5.16 If we were to establish a new, dedicated regulator over the long term, we would need to consider options for the interim period, given the time it would take to set up a new body. These include empowering an existing regulator for a limited time period (Ofcom would be a strong candidate, given its experience in upholding its current remit to tackle harmful or offensive content, in the context of TV and radio), or establishing a shadow body that can make the necessary preparations ahead of the new authority. Either approach will require cooperation with other regulators to ensure the new framework complements existing safeguards.

5.17 Alongside these options, the government is carefully considering the remits of existing regulators that may overlap with these new requirements and whether consolidation of these functions, or a broader restructuring of the regulatory landscape, would reduce the risk of duplication and minimise burdens on businesses. It is also important to consider where possible future regulatory functions to tackle other online harms may sit to ensure the institutional structures will endure.

5.18 The government will take steps to ensure that the regulator can command public confidence in its independence, impartiality, capability and effectiveness. For example, we will consider examples from other regulated sectors about how to ensure that any movement of staff between the regulator and companies in scope does not undermine the public's confidence in the regulator's independence, while also ensuring the regulator is able to attract staff with the right skills, knowledge and experience.

Consultation questions

Question 8: What further steps could be taken to ensure the regulator will act in a targeted and proportionate manner?

Question 9: What, if any, advice or support could the regulator provide to help businesses, particularly start-ups and SMEs, comply with the regulatory framework?

Question 10: Should an online harms regulator be: (i) a new public body, or (ii) an existing public body?

Question 10a: If your answer to question 10 is (ii), which body or bodies should it be?

Powers and capabilities of the regulator

5.19 The relationship between companies and regulators is often asymmetric. In regards to online harms this asymmetry can only be overcome if the regulator has real expertise in the technologies, platforms and practices under regulation.

5.20 The new regulator will require the capacity to understand how online technology and platforms operate, and collect, analyse and act upon the relevant data submitted by companies whose services are in scope. It will also require sufficient capacity to undertake research and horizon scanning to ensure the regulatory requirements keep pace with innovation and the emergence of new harms.

5.21 The government intends the new regulator to quickly become cost neutral to the public sector. To recoup the set-up costs and ongoing running costs, the government is considering fees, charges or a levy on companies whose services are in scope. This could fund the full range of the regulator's activity, including setting and enforcing codes of practice, preparing transparency reports, and any education and awareness activities by the regulator.

Consultation questions

Question 11: A new or existing regulator is intended to be cost neutral: on what basis should any funding contributions from industry be determined?

6: Enforcement

Summary

- The regulator will have a range of enforcement powers to take action against companies that fail to fulfil their duty of care. These will include the power to issue substantial fines.
- We are consulting on which enforcement powers the regulator should have at its disposal, particularly to ensure a level playing field between companies that have a legal presence in the UK, and those who operate entirely from overseas.
- In particular, we are consulting on powers that would enable the regulator to disrupt the business activities of a non-compliant company, measures to impose liability on individual members of senior management, and measures to block non-compliant services.
- Companies will continue to be liable for the presence of illegal content or activity on their services, subject to existing protections.

6.1 The regulator will have a suite of powers to take effective enforcement action against companies that have breached their statutory duty of care. While the primary objective will be to drive rapid remedial action, when companies do not cooperate there will be serious consequences.

The regulator's enforcement powers

6.2 To be effective, the regulator must have enforcement powers that both incentivise companies to comply and are technically possible to implement. The regulator will use these powers in a proportionate manner, taking the impact on the economy into account. These powers must also be designed and used in a way that creates a level playing field, so that companies with a presence in the UK are not disproportionately penalised.

6.3 The potential sanctions for non-compliance need to:

- Incentivise companies to fulfil their obligations quickly and effectively.
- Apply effectively across different types of online companies, which vary enormously in size and revenue and may be based overseas.
- Be proportionate to potential or actual damage caused and the size and revenue of the company.

6.4 There are a number of enforcement powers that will be an essential part of the new regulator's toolkit. These powers have been well tested in numerous other regulatory regimes. These core powers will include:

- Issuing civil fines for proven failures in clearly defined circumstances. Civil fines can be tied into metrics such as annual turnover, volume of illegal material, volume of views of illegal material, and time taken to respond to the regulator.
- Serving a notice to a company that is alleged to have breached standards, and setting a timeframe to respond with an action plan to rectify the issue.
- Requiring additional information from the company regarding the alleged breach.

- Publishing public notices about the proven failure of the company to comply with standards.

6.5 However, because of the particularly serious nature of some of the harms in scope, the global nature of many online services and the weak economic incentives for companies to change their behaviour, we think it is likely the regulator will need additional powers at its disposal. These measures will be more contentious because of either challenges around their technical feasibility or the potential impact on companies and the wider economy. We are therefore consulting on these options alongside this White Paper:

- Disruption of business activities. In the event of extremely serious breaches, such as a company failing to take action to stop terrorist use of their services, it may be appropriate to force third party companies to withdraw any service they provide that directly or indirectly facilitates access to the services of the first company, such as search results, app stores, or links on social media posts. These measures would need to be compatible with the European Convention on Human Rights.
- ISP blocking. Internet Service Provider (ISP) blocking of non-compliant websites or apps – essentially blocking companies' platforms from being accessible in the UK – could be an enforcement option of last resort. This option would only be considered where a company has committed serious, repeated and egregious violations of the outcome requirements for illegal harms, failing to maintain basic standards after repeated warnings and notices of improvement. Deploying such an option would be a decision for the independent regulator alone. While we recognise that this would have technical limitations, it could have sufficient impact to act as a powerful deterrent. The British Board of Film Classification (BBFC) will have this power to address non-compliance when the requirements for age verification on online pornography sites come into force. We are exploring a range of options in this space, from a requirement on ISPs to block websites or apps following notification by the regulator, through to the regulator issuing a list of companies that have committed serious, repeated and egregious violations, which ISPs could choose to block on a voluntary basis.
- Senior management liability. We are exploring possible options to create new liability for individual senior managers. This would mean certain individuals would be held personally accountable in the event of a major breach of the statutory duty of care. This could involve personal liability for civil fines, or could even extend to criminal liability. In financial services, the introduction of the Senior Managers & Certification Regime has driven a culture change in risk management in the sector. Another recent example of government action is establishing corporate offences of failure to prevent the criminal facilitation of tax evasion. Recent changes to the Privacy and Electronic Communications Regulations (PECR) provide powers to assign liability to a specific person or position within an organisation. However, this is as yet largely untested. There are a range of options for how this could be applied to companies in scope of the online harms framework, and a number of challenges, such as identifying which roles should be prescribed and whether this can be proportionate for small companies.

Consultation questions

Question 12: Should the regulator be empowered to i) disrupt business activities, or ii) undertake ISP blocking, or iii) implement a regime for senior management liability? What, if any, further powers should be available to the regulator?

Working with law enforcement and other relevant government agencies

6.6 As previously set out, the regulator will set a spectrum of expectations for companies, reflecting the nature of online harms and the company concerned. The expectations placed on firms, as set out in codes of practice, will vary according to the category of harm.

6.7 The codes of practice for specific illegal harms (e.g. terrorism and CSEA) will seek to set out expectations that keep pace with criminal behaviours and activities. They will establish requirements and processes, where appropriate and proportionate, for referring illegal content and activities to law enforcement and other relevant government agencies to aid investigations.

6.8 In formulating the codes of practice for other illegal harms, the regulator will be expected to incorporate insights from law enforcement and other relevant government agencies to ensure the codes are adequately addressing the threat. The regulator will also be required to ensure its wider actions are not detrimental to matters of national security.

Enforcement in an international context

6.9 The new regulatory regime will need to handle the global nature of both the digital economy and many of the companies in scope. The law will apply to companies that provide services to UK users. We will design the regulator's powers to ensure that it can take action against companies without a legal presence in the UK, including blocking platforms from being accessible in the UK as a last resort. Where companies do not have a legal presence in the UK, close collaboration between government bodies, regulators and law enforcement overseas, in the EU and further afield, will be required.

6.10 We are also considering options for the regulator, in certain circumstances, to require companies which are based outside the UK to appoint a UK or EEA-based nominated representative. This is similar to the concept of nominated representatives within the EU's GDPR. Under GDPR, if an organisation is based outside of the EEA but serves users in the EEA, they are required to nominate an EEA-based representative, notionally helping to enforce compliance in respect of companies established outside the EEA. This may be done by appointing a representative under a simple service contract, and for the information to be easily accessible to the regulator by publishing on the company's website. The regulator will also ensure that such representatives are tightly linked to their own genuine knowledge and ability to control the situation. However, under GDPR, the extent of compliance by companies based outside the EEA is still relatively untested – last year, the ICO launched the first extra-EEA enforcement case under the GDPR against AggregateIQ Data Services Ltd (AIQ) based in Canada. The regulator will take a company's failure to comply with such a requirement into consideration when making decisions about appropriate enforcement action.

6.11 It is vital that the regulator takes an international approach. Where similar regulators and legal systems are in place in other countries, the regulator will lead engagement with its international counterparts. Having these relationships will support the UK's ability to put pressure on companies whose primary base is overseas.

6.12 As part of our global strategy for tackling online harms, the government will seek to work with international partners to build consensus and identify common approaches to keep citizens safe online.

Appeals

6.13 Companies and others must have confidence that the regulator is acting fairly and within its powers. They will have the ability to seek judicial review of the regulator's actions and decisions through the High Court. We will also seek views through the consultation about whether there should be another statutory mechanism of review, which would allow the use of a tribunal other than the High Court, and what bar should be set for appeals through this route.

Consultation questions

Question 13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

Question 14: In addition to judicial review, should there be a statutory mechanism for companies to appeal against a decision of the regulator, as exists in relation to Ofcom under sections 192-196 of the Communications Act 2003?

> Question 14a: If your answer to question 14 is 'yes', in what circumstances should companies be able to use this statutory mechanism?
>
> Question 14b: If your answer to question 14 is 'yes', should the appeal be decided on the basis of the principles that would be applied on an application for judicial review or on the merits of the case?

Current liability for illegal content

6.14 Under the current liability regime, which is derived from the EU's e-Commerce Directive, platforms are protected from legal liability for any illegal content they 'host' (rather than create) until they have either actual knowledge of it or are aware of facts or circumstances from which it would have been apparent that it was unlawful, and have failed to act 'expeditiously' to remove or disable access to it. In other words, they are not liable for a piece of user-generated illegal content until they have received a notification of its existence, or if their technology has identified such content, and have subsequently failed to remove it from their services in good time.

6.15 In 2018, the Prime Minister announced the government's intention to look at how existing frameworks and definitions can be made to work better, with a view to ensuring companies take greater responsibility for removal of illegal content on their services. The Prime Minister noted that applying 'publisher' levels of liability to companies would not be proportionate; such an approach would force companies to check every piece of content before upload to ensure it was legal, with implications for freedom of expression, and it would be difficult to reconcile with platforms hosting large amounts of user generated content.

6.16 Our review found that, while it is important to ensure that companies have the right level of liability for illegal content, this is not the most effective mechanism for driving behavioural change by companies. The existing liability regime only forces companies to take action against illegal content once they have been notified of its existence. It therefore does not provide a mechanism to ensure proactive action to identify and remove content. In addition, even if reforms to the liability regime successfully addressed the problem of illegal content, they would not address the full range of harmful activity or harmful behaviour in scope. More fundamentally, the focus on liability for the presence of illegal content does not incentivise the systemic improvements in governance and risk management processes that we think are

necessary. We concluded that standalone changes to the liability regime would be insufficient. Instead, the new regulatory framework takes a more thorough approach. It will increase the responsibility that services have in relation to online harms, in line with the existing law that enables platforms to operate. In particular, companies will be required to ensure that they have effective and proportionate processes and governance in place to reduce the risk of illegal and harmful activity on their platforms, as well as to take appropriate and proportionate action when issues arise. The new regulatory regime will also ensure effective oversight of the take-down of illegal content, and will introduce specific monitoring requirements for tightly defined categories of illegal content.

7. Fulfilling the duty of care

Summary

- Ahead of the implementation of the new regulatory framework, we will encourage companies to take early action to address online harms.
- To assist this, the White Paper sets out high-level expectations of companies, including some specific expectations in relation to certain harms. We expect the regulator to reflect these in future codes of practice.

7.1 While it will be for the new regulator to produce codes of practice when it becomes operational, the government expects companies to take action now to tackle harmful content or activity on their services. For those harms where there is a risk to national security or to the physical safety of children, the government will publish interim codes of practice.

7.2 To support early action from companies, and to guide the initial priorities of the regulator, we have set out high-level expectations of companies below. Some of these apply to all harms in scope, and others apply to specific issues where a tailored response is more appropriate.

7.3 Given the range of services in scope of the regulatory framework, some of the expectations below may not be applicable to every company. However, each company in scope will be required to build an understanding of the risk associated with its service(s) and take reasonable steps to guard against the risk of harm in order to fulfil its duty of care.

The duty of care

7.4 As indication of their compliance with their overarching duty of care to keep users safe, we envisage that, where relevant, companies in scope will:

- Ensure their relevant terms and conditions meet standards set by the regulator and reflect the codes of practice as appropriate.
- Enforce their own relevant terms and conditions effectively and consistently.
- Prevent known terrorist or CSEA content being made available to users.
- Take prompt, transparent and effective action following user reporting.
- Support law enforcement investigations to bring criminals who break the law online to justice.
- Direct users who have suffered harm to support.
- Regularly review their efforts in tackling harm and adapt their internal processes to drive continuous improvement.

7.5 To help achieve these outcomes, we expect the regulator to develop codes of practice that set out:

- Steps to ensure products and services are safe by design.
- Guidance about how to ensure terms of use are adequate and are understood by users when they sign up to use the service.
- Measures to ensure that reporting processes and processes for moderating content and activity are transparent and effective.
- Steps to ensure harmful content or activity is dealt with rapidly.

- Processes that allow users to appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to ensure that users who have experienced harm are directed to, and receive, adequate support.
- Steps to monitor, evaluate and improve the effectiveness of their processes.

7.6 The rest of this chapter sets out more specific outcomes for harms in scope, as well as further examples of how companies will be expected to fulfil their duty of care.

CSEA

7.7 CSEA online poses a severe threat to the physical safety and emotional wellbeing of children. Companies will be required to take stringent action – proactive and reactive – to monitor and address the growing and evolving threat and to tackle all manifestations of CSEA activity, including bearing down on the proliferation of imagery and taking necessary steps to target grooming and live streaming.

7.8 We will also expect the regulator to set expectations around imagery that may not be visibly illegal, but linked to CSEA, for example, a series of images, some of which were taken prior to or after the act of abuse itself. We are continuing to work with partners to understand the impact of this abusive content on victims.

7.9 Existing legal requirements and voluntary industry initiatives are set out earlier in this White Paper.

CSEA: Fulfilling the duty of care

7.10 Some of the areas we expect the regulator to include in a code of practice are:

- The reasonable steps companies should take to proactively prevent new and known CSEA content, and links to such material, being made available to users.
- The reasonable steps companies should take to proactively identify and act upon CSEA activity such as grooming.
- The reasonable steps companies should take to proactively identify and act upon CSEA activity alongside, or within, live streams.
- The reasonable steps companies should take to proactively identify accounts showing indicators of CSEA activity and ensure children are protected from them, including disabling accounts and informing law enforcement where appropriate.
- The reasonable steps companies should take to prevent searches linking to CSEA activity and content, including automatic suggestions for CSEA content not being made and users being directed towards alternative sources of information or support.
- The reasonable steps companies should take to ensure services are safe by design.
- The reasonable steps companies should take to provide effective systems for child users, and their parents or carers, to report, remove and prevent further circulation of images of themselves which may fall below the illegal threshold, but which leave them vulnerable to abuse.
- The reasonable steps companies should take to implement effective measures to identify which users are children, and adopt enhanced safety measures for these users.

- The reasonable steps companies should take to promptly inform law enforcement where there is information about a CSEA offence, including provision of sufficient identifying information about victims and perpetrators.
- Steps companies should take to continually review their efforts in tackling CSEA, to adapt their internal processes and technology, and to continue to keep sufficiently up to date with the threat landscape; ensuring that their identification and response continually improves.
- Guidance on the CSEA content and activity companies should proactively prevent, identify and act upon, which will help inform the design and implementation of technological tools.
- Thresholds for the types of content companies should preserve following removal, for how long they should keep it and when/with whom such information should proactively be shared.
- Steps to ensure that users who are affected by CSEA content and activity are directed to, and are able to access, adequate support.

Terrorist use of the internet

7.11 Our aim is to ensure there is no safe space online for terrorists to operate, and to prevent the dissemination of terrorist content online. Such material can have significant real-world ramifications and poses a severe threat to national security. Given this, the regulator will require companies to take robust action to tackle terrorist content and activity on their services, and ideally prevent this content from reaching users in the first place.

7.12 We set out some of the existing measures to tackle terrorist use of the internet in Part 1. The establishment of the GIFCT and voluntary cooperation between the government and the industry has led to the positive creation and adoption of automated technologies by the biggest companies to proactively detect and remove terrorist content. This is essential if the threat from terrorists is to be prevented. It is also essential that smaller companies receive sufficient support to successfully prevent their platforms from being exploited, and that all relevant platforms support the role of law enforcement and other relevant government agencies.

Preventing terrorist use of the internet: Fulfilling the duty of care

7.13 Some of the areas we expect the regulator to include in a code of practice are:

- The reasonable steps companies should take to prevent new and known terrorist content, and links to content, being made available to users. This should include guidance on proactive use of technological tools, where appropriate, to identify, flag, block or remove terrorist content.
- Guidance on the content and/or activity companies should proactively prevent from being made available to users, which will help inform the design of technological tools.
- Clarification as to what constitutes an expedient timeframe for the removal of terrorist content where either it is not known that it is terrorist content at the point of upload, or it is not possible to prevent it from being made available to users.
- Guidance about the requirements for how companies should inform and support law enforcement and other relevant government agencies' investigations and prosecution of criminal offences in the UK. This will include specific guidance

about the content companies should preserve following removal and for how long, and when companies should proactively alert law enforcement and other relevant government agencies to this content.

- The reasonable steps companies should undertake when dealing with accounts that have uploaded, engaged with or disseminated terrorist content, including disabling accounts.
- The reasonable steps companies should take to identify and act upon terrorist activity or content, including within live streams.
- The reasonable steps we expect services to take to prevent searches which lead to terrorist activity and/or content, including automatic suggestions for terrorist content not being made and users being directed towards alternative sources of information or support.
- Steps companies should take to ensure that services are safe by design.
- Steps companies should take to continually review their efforts in tackling terrorist material, to adapt their internal processes and technology, and to continue to keep sufficiently up to date with the threat landscape; ensuring that their identification and response continually improves.

Serious violence

7.14 Violent content ranges from content which directly depicts or incites acts of violence, through to content which is violent with additional contextual understanding or which is harmful to users through the glamorisation of weapons and gang life.

Serious violence: Fulfilling the duty of care

7.15 Some of the areas we expect the regulator to include in a code of practice are:

- Guidance to companies to outline what activity and material constitutes violent or violence related content, including that which is explicitly criminal and how to report it.
- Guidance on the content and/or activity companies should proactively identify, to either prevent it being made publicly available or prevent further sharing and to ensure that users will not receive recommendations to violent or violence related content.
- Clarification as to what constitutes an expedient timeframe for the referral and removal of content when it is either proactively identified or referred.
- Guidance about the requirements for how companies should inform and support law enforcement and other relevant government agencies' investigations and prosecution of criminal offences in the UK. This should include specific guidance about the content companies should preserve following removal and for how long, and when companies should proactively alert law enforcement and other relevant government agencies to this content.
- The reasonable steps companies should take when dealing with accounts that have uploaded, engaged with or disseminated violent or violence related content, including disabling accounts.
- Measures to ensure that reporting processes are fit for purpose to tackle this harm and are clear, visible and easy to use. Users should receive clear explanations of decisions taken.

- Steps to ensure that services have effective and transparent processes for moderating this type of content and users are kept up to date with the progress of their report.
- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to ensure that users who have been exposed to violent or violence related material are directed to, and are able to access, adequate support.
- Steps companies should take to ensure that services are safe by design.
- Steps companies should take to continually review their efforts in tackling this content, and to continue to keep sufficiently up to date with the threat landscape, adapting their internal processes accordingly to ensure that their identification and response continually improves.

Hate crime

7.16 Hate crimes include crimes demonstrating hostility on the grounds of an individual's actual or perceived race, religion, sexual orientation, disability or transgender identity. In Action Against Hate, the government's plan for tackling hate crime (2016), and Action Against Hate Two Years On (2018), jointly led by the Ministry of Housing, Communities and Local Government (MHCLG) and the Home Office, the government has made clear that offending online is just as serious as that occurring offline and perpetrators of hateful attacks should be held accountable for their actions. Companies should create platforms where people – whatever their identity or background – can work, learn and socialise together, with shared rights, responsibilities and opportunities.

7.17 A number of third party organisations are providing support to users to report instances of hate crime. The government supports True Vision, the police hate crime reporting portal, which helps encourage victims of hate crime to report instances online through their website report-it.org.uk. In Action Against Hate Two Years On (2018), we committed to supporting the National Police Chiefs' Council (NPCC) to refresh the True Vision website.

7.18 MHCLG and the Home Office also support and engage with third party organisations such as the Community Security Trust, Tell MAMA and Stop Hate UK, who have Trusted Flagger status with social media platforms to provide greater support to users to report experiences of hate crime online. We support the continued close cooperation of these organisations with government and social media platforms.

Hate crime: Fulfilling the duty of care

7.19 Some of the areas we expect the regulator to include in a code of practice are:

- Guidance to companies to outline what activity and material constitutes hateful content, including that which is a hate crime, or where not necessarily illegal, content that may directly or indirectly cause harm to other users – for example, in some cases of bullying, or offensive material.
- Guidance on the content and/or activity companies should proactively identify, to either prevent it being made publicly available or prevent further sharing.
- Steps companies should take to ensure their services are safe by design.
- Expectations around clear and accessible guidance to users on what constitutes hate crime and how to report it.

- Measures to ensure that reporting processes are fit for purpose to tackle this harm and are clear, visible and easy to use. Users should receive clear explanations of decisions taken.
- Steps to ensure that services have effective and transparent processes for moderating this type of content and users are kept up to date with the progress of their report.
- Clarification as to what constitutes an expedient time frame for the removal of (or temporarily limiting access to) hateful content.
- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Reasonable steps to take to ensure that users will not receive recommendations to hateful or inappropriate content.
- Steps to ensure that users who have been exposed to hateful material are directed to, and are able to access, adequate support.
- Guidance on the requirements for how companies should support law enforcement and other relevant bodies' investigations where appropriate.
- An expectation that companies will continually review their efforts in tackling hateful material and adapt their internal processes accordingly, to drive continuous improvement.

Harassment

7.20 Being harassed online can be upsetting and frightening, and online harassment can amount to a criminal offence. Far too many people, from public figures to schoolchildren, have experienced this kind of behaviour. A poll conducted for Amnesty International found that 21% of the women surveyed in the UK (504 women) had experienced online harassment or abuse, with 17% having experienced this on social media.[69] There are many forms of abuse and some evidence suggests differences in the type of abuse experienced between men and women. Research suggests more women than men experience sexual forms of verbal abuse (21% compared to 9% of men), while more men than women experience offensive name calling (30% compared to 23%) and physical threats (12% compared to 8%).[70]

7.21 The cumulative impact of online misogyny undermines women's and girls' digital contributions, silencing their voices and reducing their visibility. As a result of abuse or harassment, 67% of women in the UK experienced a feeling of apprehension when thinking about using the internet or social media.[71]

Harassment: Fulfilling the duty of care

7.22 Companies will need to take robust action when there is evidence that users are being harassed or abused on their services. Companies will also need to respond quickly and proportionately if this activity emerges.

69 Amnesty International (2018). Toxic Twitter – Women's Experiences of Violence and Abuse on Twitter. Available at: https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-3/

70 Pew Research Centre (2017). Online Harassment. Available at: http://www.pewInternet.org/2017/07/11/online-harassment-2017/

71 NewStatesman (2017). Social media and the silencing effect: why misogyny online is a human rights issue. Available at: https://www.newstatesman.com/2017/11/social-media-and-silencing-effect-why-misogyny-online-human-rights-issue

7.23 Current measures taken by companies to tackle online harassment include:

- Tools to report incidents of harassment.
- Tools to block or stay hidden from other users.
- Removal of content which is illegal or violates acceptable use.

7.24 Some of the areas we expect the regulator to include in a code of practice are:

- Steps companies should take to ensure that their services are safe by design. For victims of harassment, it is important that there are easy-to-use tools that allow them to take control over the privacy and visibility of their account and who is able to contact them.
- Tools companies can provide to help users experiencing harassment, such as the ability to mute, block or stay hidden from other users, and to manage and control access to particular services and content.
- Guidance about how to ensure it is easy for users to understand these tools, and the company's terms of use in relation to this harm, when they sign up to use the service.
- Measures to ensure that reporting processes are fit for purpose to tackle this harm, such as the ability to report a high volume of messages in bulk to reduce the burden on victims suffering from a campaign of harassment, and a prompt to use the tools to block the other user while the report is being investigated.
- Services have effective and transparent processes for moderating this type of content and activity. Users are kept up to date with the progress of their report.
- Steps companies should take to ensure harms are dealt with rapidly, such as removing content which is illegal, blocking users responsible for illegal activity and, where appropriate, supporting law enforcement efforts.
- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to prevent banned users creating new accounts to continue the harassment.
- Steps to limit anonymised users abusing their services, including harassing others.
- Steps to ensure that users who have experienced harassment are directed to, and are able to access, adequate support.

Disinformation

7.25 When the internet is deliberately used to spread false or misleading information, it can harm us in many different ways, encouraging us to make decisions that could damage our health, undermining our respect and tolerance for each other and confusing our understanding of what is happening in the wider world. It can also damage our trust in our democratic institutions, including Parliament.

7.26 Current initiatives that companies are exploring to tackle the spread of disinformation include:

- Terms of service that require users not to misrepresent their identity on social media in order to disseminate or amplify disinformation.
- Tools to report suspicious, fake or spam accounts on some social media platforms.
- Use of automated AI techniques to detect and remove fake and spam accounts.
- Partnerships between platforms and independent fact-checking services.

- Tools to provide users with more context about the content they view on platforms, including enhanced transparency about the origins of political and electoral adverts.

Disinformation: Fulfilling the duty of care

7.27 Companies will need to take proportionate and proactive measures to help users understand the nature and reliability of the information they are receiving, to minimise the spread of misleading and harmful disinformation and to increase the accessibility of trustworthy and varied news content.

7.28 Some of the areas we expect the regulator to include in a code of practice are:

- The steps companies should take in their terms of service to make clear what constitutes disinformation, the expectations they have of users, and the penalties for violating those terms of service.
- Steps that companies should take in relation to users who deliberately misrepresent their identity to spread and strengthen disinformation.
- Making content which has been disputed by reputable fact-checking services less visible to users.
- Using fact-checking services, particularly during election periods.
- Promoting authoritative news sources.
- Promoting diverse news content, countering the 'echo chamber' in which people are only exposed to information which reinforces their existing views.
- Ensuring that it is clear to users when they are dealing with automated accounts, and that automated dissemination of content is not abused.
- Improving the transparency of political advertising, helping meet any requirements in electoral law.
- Reporting processes which companies should put in place to ensure that users can easily flag content that they suspect or know to be false, and which enable users to understand what actions have been taken and why.
- Processes for publishing data that will enable the public to assess the overall effectiveness of the actions companies are taking, and for supporting research into the nature of online disinformation activity.
- Steps that services should take to monitor and evaluate the effectiveness of their processes for tackling disinformation and adapt processes accordingly.

7.29 Maintaining a news environment where accurate content can prevail and high quality news has a sustainable future is vital to healthy social and democratic engagement, and key to long-term success in tackling disinformation. In March 2018, the government commissioned Dame Frances Cairncross to conduct her independent review into the sustainability of high quality journalism. In her detailed and considered report (published in February 2019), Dame Frances proposed that a 'news quality obligation' be imposed upon social media companies, which would require these companies to improve how their users understand the origin of a news article and the trustworthiness of its source. This recommendation is very much in line with our aim to strengthen the online environment and relates closely to our expectations for social media companies (as set out above). The government is now considering this proposal and Dame Frances' other recommendations, and we will look to take action where appropriate.

7.30 Companies will be required to ensure that algorithms selecting content do not skew towards extreme and unreliable material in the pursuit of sustained user engagement.

7.31 Importantly, the code of practice that addresses disinformation will ensure the focus is on protecting users from harm, not judging what is true or not. There will be difficult judgement calls associated with this. The government and the future regulator will engage extensively with civil society, industry and other groups to ensure action is as effective as possible, and does not detract from freedom of speech online.

Encouragement of self-harm and suicide

7.32 Users should be able to talk online about sensitive topics such as suicide and self-harm, but more needs to be done to protect vulnerable users and tackle content and behaviour which encourages suicide and self-harm.[72]

7.33 Current measures to tackle the encouragement of self-harm and suicide include:

- Arrangements between individual companies and charities to improve the identification and removal of this content when it is reported.
- Services that signpost help and promote supportive content to their users.

Encouragement of self-harm and suicide: Fulfilling the duty of care

7.34 Companies will be required to take robust action to address harmful suicidal and self-harm content that provides graphic details of suicide methods and self-harming, including encouragement of self-harm and suicide. Services must also respond quickly to identify and remove content which is illegal or violates terms of use, and act swiftly and proportionately when this content is reported to them by users.

7.35 Some of the areas we expect the regulator to include in a code of practice are:

- Steps to ensure that vulnerable users and users who actively search for or have been exposed to this content, including content that encourages eating disorders, are directed to, and able to access, adequate support.
- Ensuring that companies work with experts in suicide prevention to ensure that their policies and practices are designed to protect the most vulnerable (and to ensure that moderators receive appropriate training).
- Steps companies should take to ensure that their services are safe by design, including tools to help users avoid material or behaviour which encourages suicide or self-harm, and measures to block content and block, mute and stay hidden from other users.
- Guidance about how to ensure it is easy for users to understand these tools, and the company's terms of use in relation to these harms, when they sign up to use the service.
- Processes to stop algorithms promoting self-harm or suicide content to users.
- Measures to ensure that reporting processes and processes for moderating content and activity are transparent and effective at tackling the encouragement of self-harm and suicide and measures to ensure that users are kept up to date with the progress of their report.

72 Encouraging and assisting suicide is a criminal offence. Self-harm is not a criminal offence.

- Steps services should take to ensure they engage sufficiently with civil society groups and law enforcement, so that moderators are educated about what constitutes self-harm or suicide encouragement and how it can be prevented and tackled.
- Steps companies should take to ensure harm is tackled rapidly, such as removing content which is illegal or violates acceptable use, and blocking users responsible for activity which violates terms and conditions, as well as steps that services can take to ensure that these measures are conducted sensitively.
- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to prevent banned users creating new accounts to continue to encourage suicide or self-harm.

Online abuse of public figures: Fulfilling the duty of care

7.36 As set out in Box 14, those involved in public life in the UK experience regular and sustained abuse online, which goes beyond free speech and impedes individuals' rights to participate. As well as being upsetting and frightening for the individual involved, this abuse corrodes our democratic values and dissuades good people from entering public life.

7.37 The steps we expect the regulator to include in codes of practice relating to all forms of abusive behaviour online, including harassment and cyber-bullying, will also help address this problem, and include:

- Steps companies should take to ensure that their services are safe by design. For all users, including public figures, it is important that there are easy-to-use tools that allow them to take control over the privacy and visibility of their account and who is able to contact them.
- Tools companies can provide to help users experiencing abuse, such as the ability to mute, block or stay hidden from other users, and to manage and control access to particular services and content.
- Clear guidance in the company's terms of use on the type of activity which will be treated as unacceptable and the actions the company will take in response to such activity, which is available to users when they sign up to use the service.
- Measures to ensure that reporting processes are fit for purpose to tackle this harm, such as the ability to report a high volume of messages in bulk to reduce the burden on victims suffering from a campaign of online abuse, and a prompt to use the tools to block the other user while the report is being investigated.
- Services have effective and transparent processes for moderating this type of content and activity. Users are kept up to date with the progress of their report.
- Steps companies should take to ensure harms are dealt with rapidly, such as removing content which is illegal, blocking users responsible for illegal activity, enforcing and upholding the service's relevant terms and conditions and, where appropriate, supporting law enforcement efforts.
- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps companies should take to limit anonymised users using their services to abuse others.
- Steps to prevent banned users creating new accounts to continue the abuse.

- Steps to ensure that users who are affected by abusive comments and activity are directed to, and are able to access, adequate support.

Interference with legal proceedings

7.38 Activity that can impede a person's right to a fair trial, or that breaches a person's legal right to anonymity, such as communications that may amount to a breach of a court order or a statutory prohibition, or that may prejudice a jury, may amount to contempt of court or be a criminal offence.

7.39 Current measures to tackle this problem include:

- Action taken by law enforcement and the criminal justice system in relation to publishing information online and exposing the identity of protected individuals which could jeopardise legal proceedings.
- The bringing of contempt proceedings against those who create a substantial risk of serious prejudice.

7.40 Furthermore, in its Response to the Call for Evidence on the Impact of Social Media on the Administration of Justice,[73] the Attorney General's Office has:

- Set out plans to promote the safe use of social media as part of a public legal education campaign, which will include a GOV.UK webpage.
- Highlighted that the Judicial Office are working to develop clear, accessible, and comprehensive guidance on contempt.
- Agreed points of contact with a number of social media companies so that relevant material can be flagged and, if necessary, removed.
- Set out plans to work with cross-government partners to improve the enforcement of the law on anonymity online.

7.41 Companies will be required to take robust action when there is evidence that a risk of interference with criminal trials or other legal proceedings is present. Companies will also be required to respond quickly and proportionately where new risks emerge.

Interference with legal proceedings: Fulfilling the duty of care

7.42 Some of the areas we expect the regulator to include in a code of practice are:

- Tools companies can provide to help users report possible interference with legal proceedings, such as the ability to report anonymously.
- Measures to ensure that reporting processes and processes for moderating content and activity are transparent and effective at tackling interference with legal proceedings and measures, to ensure that users are kept up to date with the progress of their report.
- Steps companies should take to ensure harms are dealt with rapidly, to ensure that posts that are in contempt of court or that breach anonymity orders are removed as soon as possible once they have been reported. User guidance setting this out should be incorporated into the company's terms and conditions to ensure clarity when users sign up to use the service.

73 Attorney General's Office (2019). Response to Call for Evidence on the Impact of Social Media on the Administration of Justice. Available at: https://www.gov.uk/government/publications/response-to-call-for-evidence-on-the-impact-of-social-media-on-the-administration-of-justice

- Processes companies should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to prevent banned users creating new accounts to continue or repeat the interference with legal proceedings.

Cyberbullying

7.43 Cyberbullying, including trolling, is unacceptable. Being bullied online can be a deeply upsetting experience, particularly for children or other vulnerable users.

7.44 Current measures to tackle cyberbullying and trolling include:

- Provision of information and resources on bullying and other online safety issues.
- Tools to report incidents of bullying.
- Tools to block or stay hidden from other users.
- Removal of content which violates acceptable use.

Cyberbullying: Fulfilling the duty of care

7.45 The regulator will set out steps that should be taken to tackle cyberbullying, such as ensuring that those who have suffered from this harm are directed to, and are able to access, adequate support.

7.46 In the meantime, the statutory Social Media Code of Practice, published alongside this White Paper in line with the DCMS Secretary of State's duty under section 103 of the Digital Economy Act 2017, sets out non-binding principles that companies should adhere to in order to tackle bullying, insulting, intimidating and humiliating conduct online. It also explains good practice ways to implement these principles. We expect all social media companies to adhere to this code of practice, ahead of the new regulatory requirements. We expect the regulator to consider this guidance when drawing up future codes of practice.

7.47 These principles are:

- Social media providers should maintain a clear and accessible reporting process to enable individuals to notify social media providers of harmful conduct.
- Social media providers should maintain efficient processes for dealing with notifications from users about harmful conduct.
- Social media providers should have clear and accessible information about reporting processes in their terms and conditions.
- Social media providers should give clear information to the public about action they take against harmful conduct.

Children accessing inappropriate content

7.48 Some online content that is lawful and appropriate for adults, such as dating apps or pornography, may cause significant harm to children who either access it intentionally or stumble across it. The Chief Medical Officers for England, Wales and Scotland recently advocated a precautionary approach to protecting children from harmful content because of its possible impact on their mental health or development.

7.49 Current measures to tackle children accessing inappropriate content include:

- Forthcoming compulsory age verification for commercial online pornography sites.
- Family friendly filters to filter inappropriate material.

- Content warnings for inappropriate content.

7.50 The designated classification authorities for offline content, the BBFC and the Video Standards Council (VSC), have clear standards based on their evaluation of likely harm and use these to allocate BBFC or PEGI age suitability ratings to inform viewing decisions and protect children and vulnerable adults. These age ratings are applied voluntarily to online content by some publishers and platforms. The new regulatory framework is not intended to impact the existing classification of offline and online content by BBFC and VSC.

Children accessing inappropriate content: Fulfilling the duty of care

7.51 Companies will be required to take robust action when there is evidence that children are accessing inappropriate content. Companies will also be required to respond quickly and proportionately where new risks emerge.

7.52 Some of the areas we expect the regulator to include in a code of practice are:

- Steps companies should take to ensure that their services are safe by design. This could include the provision of accounts with different settings for children.
- Terms of service should make clear what behaviour and activity is tolerated on the service and the measures that are in place to prevent children accessing inappropriate content and they should be easy for children and parents to understand.
- Steps companies should take to ensure children are unable to access inappropriate content, including guidance on age verification, content warnings and measures to filter and block inappropriate content.
- Measures to ensure that reporting processes are fit for purpose to tackle this harm and are clear, visible and easy for children and parents to understand. Users should receive clear explanations of decisions taken.
- Services have effective and transparent processes for moderating this type of content and activity. Users are kept up to date with the progress of their report.
- Steps companies should take to ensure harms are rapidly dealt with, such as removing content which violates terms of service.
- Processes services should have in place to ensure that users can appeal the removal of content or other responses, in order to protect users' rights online.
- Steps to prevent banned users creating new accounts in order to continue to make inappropriate content which violates terms of service.

PART 4: Technology, education and awareness

8: Technology as part of the solution

Summary

- Companies should invest in the development of safety technologies to reduce the burden on users to stay safe online.
- In November 2018, the Home Secretary co-hosted a hackathon with five major technology companies to develop a new tool to identify online grooming, to be licensed for free to other companies, but more of these innovative and collaborative efforts are needed.
- The government and the new regulator will work with leading industry bodies and other regulators to support innovation and growth in this area and encourage the adoption of safety technologies.
- The government will also work with the industry and civil society to develop a safety by design framework, linking up with existing legal obligations around data protection by design and secure by design principles, to make it easier for start-ups and small businesses to embed safety during the development or update of products and services.

8.1 Technology can play a crucial role in keeping users safe online. By designing safer and more secure online products and services, the tech sector can equip all companies and users with better tools to tackle online harms. We want the UK to be a world-leader in the development of online safety technology and to ensure companies of all sizes have access to, and adopt, innovative solutions to improve the safety of their users.

Existing initiatives

8.2 In the UK, a dynamic and innovative market has sprung up around online safety, developing tools for business to protect their users from harms. For example:

- SuperAwesome, one of the fastest-growing technology SMEs in the UK, provides tools and technology that protect the digital privacy of children.

- Crisp, an SME with complex AI-based tools to support moderation and monitoring of content, helps hundreds of companies worldwide run safer platforms – every month its systems assess billions of pieces of content for illegal or harmful content, and help to identify repeat offenders who are continually posting inappropriate content.
- Yoti, a digital identity provider, is partnering with the Yubo social network to use machine learning age estimation to detect whether website users are in the right age band for their platform – an important step in helping safeguard children online.

8.3 The government is supporting the development of this emerging safety tech ecosystem in the UK:

- The Home Office worked with Faculty (formerly ASI Data Science) to develop technology that can identify the official Daesh propaganda videos that are a key part of the terrorist groups efforts to radicalise, recruit and inspire acts of terrorism in the UK and abroad.
- The government launched a challenge fund, through the GovTech Catalyst scheme, to develop technology that can automate the detection of terrorist still imagery. Five UK companies have been awarded £50,000 to work on proposals, and this year the leading proposals will receive up to £500,000 to develop and test a prototype.
- The government is investing £300,000 to fund up to five innovative projects that use new technologies to disrupt live online CSEA.
- Through National Cyber Security Centre (NCSC) Accelerator programmes, the government is ensuring the rapid development of solutions to cyber security challenges such as authentication, mobile device security and identity management, that work to reduce online harms through increasing the security of users' online environments, giving them more control over their interactions and making access harder for those who seek to use technology to facilitate abuse.

Online safety apps: BBC Own It — Box 27

BBC Own It, launching later in 2019, is a new wellbeing app aimed at children aged 8-13 receiving their first smartphone. The app is part of the BBC's commitment to supporting young people in today's changing media environment and follows the successful launch of the Own It website in 2018.

The app combines state-of-the-art machine-learning technology to track children's activity on their smartphone with the ability for children to self-report their emotional state. It uses this information to deliver tailored content and interventions to help children stay happy and healthy online, offering friendly and supportive nudges when their behaviour strays outside the norm. Users can access the app when they're looking for help but it will always be on-hand to give instant, on-screen advice and support when they need it, via a specially-developed keyboard. Features include:

- Reminding them to think twice before sharing personal details like mobile numbers on social media.
- Helping them understand how messages could be perceived by others, before they hit send.

- Tracking their mood over time and offering guidance on how to improve the situation if needed.
- Information on topics like using phones late at night and the impact on their wellbeing.

The app features specially commissioned content from across the BBC. It provides useful material and resources to help young people get the most out of their time online, and build healthy online behaviours and habits. The app will help young people and parents have more constructive conversations about their experiences online, but won't provide reports or feedback to parents and no data will leave their device.

Tackling online grooming: industry hackathon Box 28

The challenge of online grooming of children for exploitation and abuse crosses borders and platforms. Addressing it requires collaboration between companies to develop innovative solutions that can be shared in a joint effort to eradicate grooming from digital space.

- In November 2018, the Home Secretary co-hosted a 'hackathon' event in the US with Microsoft and a range of other tech companies, where they worked to develop a new AI product to detect online grooming of children. Hackathon participants analysed tens of thousands of conversations to understand patterns used by predators. This enabled engineers to develop technology to automatically and accurately detect these patterns.
- During 2019, this anti-grooming tool will be licensed free of charge to smaller and medium-sized technology companies worldwide – and government will work closely with industry to help ensure high rates of adoption.

Boosting innovation in safety technology

8.4 The online safety ecosystem incorporates a number of distinct markets including third party technical solutions, human moderation services, hashing and finger-printing technologies and AI/machine learning solutions for the automated detection of harmful content. The government will work with the tech sector to make the UK a world-leader in innovative safety solutions across these markets.

8.5 The new regulator will use its unique position in the market to drive development of new technologies and encourage the sharing of tools and best practice amongst companies.

8.6 In the meantime, the government will work with partners across industry, academia and civil society to support innovation in safety technologies. In particular, we will:

- Assess the capability and potential of the UK online safety sector.
- Work with Tech Nation, TechUK and other industry partners to help companies more effectively detect and respond to online harms by promoting the rapid innovation, development and scale-up of safety products.
- Work with UKRI to support research into understanding online harms and the development of innovative technological solutions that meet the challenge of protecting citizens online.

- Support the development of scalable privacy-enhancing technologies to allow companies to access training data to develop AI solutions, without compromising highly sensitive or illegal datasets.

8.7 We are bringing in external expertise to help government provide further direction. This includes the Digital Charter Fellowship programme, which DCMS is running in partnership with the Alan Turing Institute (Turing). Fellows will develop – with the lead government departments – policy responses to key challenges posed by the internet and new technologies. This will include experimenting with a range of processes and tools, including convening small groups of experts (from industry, academia and government) to work intensively on the issue. Manipulation, disinformation and online safety are key areas of focus for the first phase of the Fellowship programme.

8.8 We will work further with research organisations to understand how AI can best be used to detect, measure and counter online harms, while ensuring its deployment remains safe and ethical.

Online safety initiatives: Analysing and countering hate speech: The role of AI **Box 29**

Hateful content on digital platforms is a growing problem in the UK, inflicting harm on victims, creating and exacerbating social divisions, and eroding trust in the host platforms.

- However, despite the harm caused by hate content, we lack adequate data on its scale and scope, limiting our ability to develop more sophisticated and effective responses. Part of the challenge is that online hate takes many forms and is directed against many different targets, including ethnic minorities and women.
- A new project led by Turing is setting out to address this issue. The 'Hate Speech: Measures and Counter-measures' project will use a mix of natural language processing techniques and qualitative analyses to create tools which identify and categorise different strengths and types of online hate speech.
- The aim is to make these tools open and accessible to the public, and ultimately for them to be used to support a broad range of commercial and public sector providers to detect and address harmful and undesirable content. The project also aims to release annotated training datasets, enabling other researchers to further build on their work.
- Turing is planning work more broadly to study the influence of algorithmic systems on humans, as part of its initiative on safe and ethical AI.

Interventions to boost adoption and use of technologies

8.9 Many of the leading companies already use their resource and expertise to support the development of shared platforms and technologies that can be adopted by wider industry. These include Microsoft's PhotoDNA, a shared system for detecting and responding to images of child sexual abuse, and Google's Perspective API, which uses machine learning to flag potentially harmful or 'toxic' content to moderators. In November 2018, Microsoft and other companies came together in a 'hackathon' to develop anti-grooming technology, which will be licensed free of charge to smaller companies worldwide (Box 28).

8.10 It is crucial that we continue to drive the adoption of safety products so that users receive consistent levels of protection online. To achieve this we will:

- Work with industry to encourage the development and take-up of free or low-cost shared platforms for safety, such as the Home Office anti-grooming tool.
- Fund research by Doteveryone, an independent thinktank, into barriers to the adoption of technologies and working practices that promote user safety and wellbeing, and the practical guidance and techniques needed to overcome these barriers.
- Support the development by Innovate UK and BSI of a publicly available standard (PAS) for responsible innovation, to help companies think through and identify any potential issues raised by their proposed innovations.

Safety by Design

8.11 Creating a safer user experience on online products and platforms requires more than the use of new technology. Decisions made throughout the product development life cycle – around privacy and data protection, cybersecurity, moderation, reporting and support mechanisms for users, clarity of terms and conditions – all combine to shape the overall safety and security of a user's experience.

8.12 To prepare for the new regulatory framework, it should be as easy as possible for designers of products and platforms to understand what standards are expected of them, and to be able to incorporate existing good practices into their products from the earliest stages of product development to ensure that their products are safe by design.

8.13 Box 30 provides an example of good practice safe design that aims to protect children online. Across the industry as a whole, however, standards remain inconsistent, and frequently do not prioritise users' rights – in particular, it is commonplace for design to encourage addictive behaviour rather than wellbeing, or for collecting user-data to be prioritised over privacy. This results in an unacceptable burden on users to manage their online safety without sufficient support from the companies that they rely on. This is a particular concern for vulnerable users.

Online safety apps: Lego Life Box 30

In 2017, the LEGO Group launched a social-themed app, LEGO® Life. The ambition behind the app is to inspire younger children to build and share their creations in a high-safety, high-trust environment. LEGO® Life embeds safety by design principles, as well as introducing children to positive elements of social platforms, such as being able to share moments with family and friends.

They have recently strengthened this approach by using innovative solutions, such as the anthropomorphic advice engine, Captain Safety. The character provides a safety tutorial from the beginning and becomes the child's guide throughout the experience, delivering empowering safety messages at certain critical points, such as before sharing certain data or commenting on public posts.

8.14 To drive up standards, the government will work with industry and civil society to develop a Safety by Design framework to help companies incorporate online safety throughout the development or update of online services. This framework will set out clear principles and practical guidance on how to include online safety features in new applications

and platforms from the start, targeted at digital product teams, including designers, developers and user researchers. This could include guidance which highlights the need for providers to:

- Make it clear to users what forms of content are acceptable, as part of the terms of service and throughout their journey.
- Have effective systems for detecting and responding to illegal or harmful content, including the use of AI-based technology and trained moderators.
- Make it easy for users to report problem content, and design an efficient triage system to deal with reports.
- Give users control of their experience by collecting the minimum amount of personal data and giving them informed choices about how their personal information, including geolocation data, is used.

8.15 In developing the framework, we will pay special attention to the needs of start-ups and scale-up businesses, which can lack the capacity and expertise to ensure their products and services are safe by design. We expect the framework will be complementary with existing privacy by design and security by design standards. For example, it will reflect and signpost the forthcoming Age-appropriate design code (see Box 31) – and the Code of Practice for Consumer Internet of Things Security (Box 32).

8.16 We anticipate that the new regulator will build on this framework and use it to inform its approach to issuing codes of practice and guidance to companies about how to fulfil their new legal duties. We envisage that this framework will support companies to take practical action to tackle harm and meet the high-level expectations set out in Chapter 7.

8.17 The DfE is also planning to publish its Education Technology Strategy in the spring which will highlight the importance of privacy, security and safety. The strategy will include clarity on the guidelines that EdTech suppliers should adhere to and the guidance available for schools and colleges to support their procurement and use of safety technology.

Consultation questions

Question 15: What are the greatest opportunities and barriers for (i) innovation and (ii) adoption of safety technologies by UK organisations, and what role should government play in addressing these?

Question 16: What, if any, are the most significant areas in which organisations need practical guidance to build products that are safe by design?

Online safety initiatives: Age-appropriate design code: Safer design standards for children **Box 31**

Whether playing online games, watching and sharing videos or interacting with friends via apps and social media, children today grow up with digital technology as a fundamental part of their daily lives. This can enrich their lives but it can also pose risks. To help guarantee a better digital future for our children, we need to have world-leading standards that provide proper safeguards for our children when online.

- The government included provisions in the Data Protection Act 2018 to help ensure this is the case. These require the Information Commissioner to produce an 'age-appropriate design code of practice', to provide guidance on the privacy standards that organisations should adopt where they are offering online services and apps that children are likely to access and which will process their data.
- These standards in the code will be backed by legally enforceable data protection laws, which empower the Information Commissioner to take action and impose tough penalties under GDPR, including enforcement orders and fines of up to 4% of global turnover.
- The code will focus on the best interests of a child. It will ensure nothing is left to chance and that a 'data protection by design' approach is adopted. Companies will be held accountable for their actions living up to their promises on how they handle children's information.
- The code will address the need to implement high privacy settings by default and use language that is clear and easy to understand for youngsters at different stages of their development. It will also focus on key safeguards around the automated profiling of children, the use of geolocation data, and the transparency of marketing techniques.
- It will also address practices such as those used by sites and apps to personalise a child's experience to encourage them to stay online longer, such as auto-play videos and the timing of social media notifications.
- Work on developing the code is well advanced, with calls for evidence and commissioned research already concluded. A formal public consultation will follow in the coming months.

Online safety initiatives: Internet of Things: Security Code of Practice Box 32

Recent years have seen huge growth in the number of 'Internet of Things' (IoT) products, consumer-facing internet connected 'smart' devices that people use in their homes such as smart appliances, personal assistants, children's toys, web cameras and baby monitors. However, across the IoT, there are many instances of insecure products that make consumers vulnerable to cyber attacks, which can lead to physical and emotional harm.

- To combat this, in October 2018 the government published the Code of Practice for Consumer IoT Security. This code of practice consists of thirteen outcome-focused guidelines that clearly describe the steps that IoT producers need to take to ensure their products and services are secure by design.
- We believe the next stage in this work is for appropriate aspects of the code of practice to become legally enforceable, therefore offering consumers greater protection from the online harms associated with these products. We have commenced work to consider which aspects of regulatory change are necessary.

- The code of practice has also been used to create the first globally-applicable industry standard for IoT consumer devices, the ETSI 103 645 Technical Standard (ETSI TS). We will work to drive adoption of this standard, setting in place a harmonised technical approach that protects citizens across the world.

9. Empowering users

Summary

- Users want to be empowered to manage their online safety, and that of their children, but there is insufficient support in place and they currently feel vulnerable online.
- Government has taken steps to address digital literacy in the relevant areas of the school curriculum.
- The Government will develop a new online media literacy strategy, through broad consultation with stakeholders.
- While companies are supporting a range of positive initiatives, there is insufficient transparency about the level of investment and the effectiveness of different interventions. The new regulator will have the power to require companies to report on their education and awareness raising activities.

9.1 All users, children and adults, should be empowered to understand and manage risks so that they can stay safe online. The government is ensuring that children get high quality education at school to develop their digital literacy. Adult users should act in an acceptable manner, challenge unacceptable behaviour when they witness it, and use tools available to them to manage their online experience. They have a responsibility to manage their own online safety, and to support children in their care. In this rapidly changing environment, it can take time to learn how to evaluate what is and is not risky, and to acquire the skills to avoid harm.

9.2 Many companies have invested in education and awareness activities, often in partnership with civil society, and created tools to empower their users, such as software from Apple and Google that produces reports for users that help them to assess and control their online activity (see Box 33). While such industry initiatives are welcome, there continues to be a lack of transparency about their scale and effectiveness, and a real risk of duplication in the absence of strategic coordination. While we recognise the concerns of civil society about the risks of disrupting these existing positive working relationships with industry, we want to work with all stakeholders to ensure that there is sufficient industry investment in education and preventative activity, and that there is independent evaluation of its effectiveness.

9.3 The technical complexity and pace of innovation of the online world means that there is a constant need to improve the tools available to users so that they are able to manage and address risks online. A number of recent independent reports have also highlighted the specific need for improved digital literacy, including the DCMS Select Committee's report into disinformation[74] and the Cairncross report on A sustainable future for journalism.[75] Children have also told us that they want more education about online safety, as well as more support from tech companies to keep them safe (see Box 34).

74 Digital, Culture, Media and Sport Committee (2019). Disinformation and 'fake news': Final Report. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf

75 The Cairncross Review (2019). A sustainable future for journalism. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf

Online safety apps: Apple Screen Time and Google Family Link — Box 33

In June 2018, Apple launched updates to its mobile operating system that help customers reduce interruptions and manage screen time for themselves and their families.

These included:

- Daily and weekly reports that inform users of the total time spent in each app, usage across categories of apps, how many notifications are received and how often a person picks up their device.
- Tools to set specific limits on the amount of time spent in an app, and a notification that displays when a time limit is about to expire.
- Settings that allow users to more closely control their notifications, including a mode to help people get a better night's sleep.
- Parents can access information about their child's activity on their own devices to understand where their child spends their time and can manage and set limits for them.

Similarly, Google's Family Link app allows parents to:

- View how long their children spend on different apps.
- Approve or block apps their children want to download, or recommend specific apps.
- Set limits on screen time, and remotely lock a child's device for a break.

Furthermore, some gaming consoles, such as Xbox One, Playstation4 and Nintendo Switch have tools which allow parents to control access to content and place limits on screen time.

Online safety initiatives: Understanding children's needs online — Box 34

The UK Council for Child Internet Safety (now the UK Council for Internet Safety) published the Children's online activities, risks and safety research review in October 2017. This included evidence that while many feel able to cope with general or random negative comments online, personal or targeted behaviour was more distressing and they were likely to seek help from friends or family, or report abuse to the relevant social media platform. However, when experiences are persistent and extreme, children can find it difficult to tell anyone, and this often makes the experience worse.

- Children and young people are much more likely to confide in friends than parents or carers about upsetting or embarrassing incidents. The review notes a range of reasons children and young people don't talk to parents, including feeling uncomfortable talking to parents or worries that devices and internet access will be taken away from them.

- Alongside the Internet Safety Strategy Green Paper, the government worked with the British Computing Society (BCS), The Chartered Institute for IT to carry out a survey of 6,500 children and young people about online safety.[76]

The survey highlighted that:

- Two thirds of children aged 12 and under (67%) and nearly half of 13 to 18 year olds (46%) would welcome more education in schools about online safety.
- Children have low expectations of social media platforms in relation to their privacy, safety and security online and would like to be better protected against abusive content.
- Nearly half (42% of under 13s, 41% of 13 to 18 year olds) said tech companies don't think about the online safety of people their age when they're making websites or apps.
- Nearly two thirds (66% of under 13s, 63% of 13 to 18 year olds) thought tech firms should proactively delete abusive messages before complaints are made.

Existing initiatives to empower users to stay safe online

9.4 DfE continues to incorporate online safety into the school curriculum, to help children and young people understand healthy relationships online, and to improve their digital literacy to equip them to manage the different and escalating risks that young people face.

9.5 As part of this, DfE is making Relationships Education compulsory for all primary pupils, Relationships and Sex Education compulsory for all secondary pupils and Health Education compulsory for all pupils in all primary and secondary state-funded schools in England. The Department recently consulted[77] on draft guidance for these subjects which includes teaching about respectful relationships, including online, as well as health and mental wellbeing. This will include:

- How to stay safe online.
- Critically considering information and how people present themselves online.
- Rights and responsibilities.
- How data is gathered, shared and used.
- The benefits of rationing time spent online.

9.6 In the government response to the above consultation, we also set out that we intend to produce supporting information for schools on how to teach about all aspects of internet safety, not just those relating to relationships, sex and health, to help schools deliver this in a coordinated and coherent way across their curriculum.

9.7 Schools will be encouraged to teach the new subjects from September 2019 – many of them are already doing this and will be able to adapt to the new guidance quite quickly. The requirement to teach the new subjects will then follow from September 2020.

76 BCS (2018). Young people want more from social media giants over online safety. Available at: https://www.bcs.org/more/about-us/press-office/press-releases/young-people-want-more-from-social-media-giants-over-online-safety-survey-by-bcs-reveals/

77 Department for Education (2019). Consultation outcome – Relationships (and sex) education and health education. Available at: https://www.gov.uk/government/consultations/relationships-and-sex-education-and-health-education

9.8 The new computing curriculum, introduced in September 2014, includes the principles of e-safety at all key stages, with progression in the content to reflect the different and escalating risks that young people face. This includes how to use technology safely, responsibly, respectfully and securely, how to keep personal information private, and where to go for help and support when they have concerns about content or contact on the internet or other online technologies.

9.9 These changes to the curriculum are part of a broader strategy to ensure that schools are supporting young people to stay safe online. Some of the measures included in this strategy include:

- A National Centre for Computing Education that will develop, curate and disseminate a central repository of free, high quality, knowledge-rich resources for teachers to cover the whole computing national curriculum (from key stages 1-4).
- Strengthened statutory safeguarding guidance for schools in England, Keeping Children Safe in Education (KCSIE), including guidance on how to keep children safe online. The revised guidance came into effect on 3 September 2018.
- A new online safety working group, established by the Minister for Children and Families and made up of online safety and education experts, to advise the department on future iterations of the safeguarding guidance.

Wider initiatives to empower users to stay safe online

9.10 In Chapter 2, we set out the significant body of work being led by the UK Council for Internet Safety to ensure that children and vulnerable adults are taught about online safety, and that parents have access to appropriate advice, including an online resilience toolkit and driving the adoption of the Education for a Connected World framework[78] in schools (see Box 35). The government has also funded the UK Safer Internet Centre to develop cyberbullying guidance which provides advice for schools on understanding, preventing and responding to cyberbullying, and an online safety toolkit to help schools deliver sessions through PSHE about cyberbullying, peer pressure and sexting. There are a number of civil society organisations that have also made valuable contributions to online safety education and awareness, such as 5Rights (see Box 36).

9.11 The Information Commissioner's Office has also developed a public facing campaign to enable the public to better understand their data protection rights called 'Your Data Matters'.[79] The ICO has also produced teaching materials to support and empower children to understand their data rights.[80]

78 UKCIS (2018). Education for a Connected World framework. Available at:https://www.gov.uk/government/publications/education-for-a-connected-world

79 ICO. Your data matters – building confidence and trust. Available at: https://ico.org.uk/for-organisations/resources-and-support/your-data-matters-campaign/

80 ICO. Resources for schools. Tailored lesson plans for children and young people. Available at: https://ico.org.uk/for-organisations/education/resources-for-schools/

Online safety initiatives: Education for a Connected World Box 35

The UK Council for Internet Safety's (UKCIS) Education for a Connected World framework describes the digital knowledge and skills that children and young people should have the opportunity to develop at different ages and stages of their lives.

Designed to help educators engage in a meaningful dialogue with their students about their lives online, the tool covers a wide range of issues, including self-image and identity, privacy and security, online relationships and online bullying.

Online safety initiatives: 5Rights Box 36

The 5Rights Foundation is a registered charity that produces child-led and co-designed policies and works towards an online environment that meets the needs and protects the rights of children and young people online.

The 5Rights framework takes existing children's rights and applies them to the digital world:

- The Right to Remove.
- The Right to Know.
- The Right to Safety and Support.
- The Right to Informed and Conscious Use.
- The Right to Digital Literacy.

Their report Disrupted Childhood: The Cost of Persuasive Design (July 2018) highlights how persuasive design strategies deployed to maximise the collection of personal data impact on children's social, mental and physical development, and calls for better protections for children and young people.

The 5Rights report Towards an Internet Safety Strategy (January 2019) sets out a framework to prevent harm to children from digital products and services. It sets out seven priorities for the development of online safety strategies: parity of protection, design standards, accountability, enforcement, leadership, education, and evidence-based interventions.

The need for greater online media and digital literacy

9.12 Online media and digital literacy can equip users with the skills they need to spot dangers online, critically appraise information and take steps to keep themselves and others safe online. It can also have wider benefits, including for the functioning of democracy by giving users a better understanding of online content and enabling them to distinguish between facts and opinions online. In recent months, there have been several reports that recognise the importance of online media and digital literacy, calling for action at all levels. Box 37 summarises the recommendations of some of these reports.

Stakeholder calls for action to improve media and digital literacy — Box 37

- The House of Commons DCMS Select Committee has called for digital literacy to be the fourth pillar of education, alongside reading, writing and maths in its report Disinformation and 'Fake News'. The Committee also noted the role of Ofcom, the ICO, the Electoral Commission and the Advertising Standards Authority in promoting digital literacy, and recommended that the government ensures that the four main regulators produce a more united strategy in relation to digital literacy.
- The Cairncross review, A sustainable future for journalism, published in February 2019,[81] recommended that the government should develop a media literacy strategy, working with Ofcom (which has a statutory duty to promote media literacy), the online platforms, news publishers and broadcasters, voluntary organisations and academics, to identify gaps in provision and opportunities for more collaborative working.
- In 2018, the House of Lords Select Committee on Political Polling and Digital Media stressed the need to teach critical literacy skills in schools to limit the spread of misinformation online and its potential impact on democratic debate.
- The Children's Commissioner's report Growing up Digital, published in 2017, called for the creation of a compulsory digital citizenship programme for pupils aged 4 to 14, to improve children's digital literacy skills and digital resilience and to broaden digital literacy education beyond safety messages.[82]

9.13 There has been significant work in this area, with several organisations in tech, media and civil society developing resources for use in school and at home, to equip children and young people with the skills to critically assess information and keep themselves safe online.

Online safety initiatives: Examples of news literacy initiatives for children and young people — Box 38

NewsWise is a free, cross-curricular news literacy project for 9-11 year olds across the UK.

- The project, a partnership between the Guardian Foundation, National Literacy Trust and the PSHE Association, officially launched in September 2018 and Google is funding its first year.
- The project aims to strengthen children's critical thinking skills before they start using social media, and aims to deepen children and young people's understanding of why and how the news is produced, with sessions on selecting facts, checking sources and news analysis to develop children's skills of informed questioning and verification.

81 The Cairncross Review (2019). A sustainable future for journalism. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf

82 Children's Commissioner (2017). Growing up Digital: A report of the Growing Up Digital Taskforce. Available at: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2017/06/Growing-Up-Digital-Taskforce-Report-January-2017_0.pdf

The BBC's 'Young Reporter' project works with young people aged 11-18 in schools, colleges and youth organisations to help them navigate news and current affairs and give them the skills to produce their own reports and share story ideas about what matters to them.

Through the project, all schools have been given access to free online materials including classroom activities, video tutorials and the BBC iReporter game. This interactive game uses the principles of journalism to guide critical thinking by putting the player in the heart of the newsroom on the day of a breaking news story.

9.14 While we welcome these initiatives, we believe that there are notable gaps in provision and that adults need support too – for themselves but also as parents. Ofcom's Children and parents: Media use and attitudes report 2018 (January 2019)[83] notes that parents and carers are increasingly worried about the internet, and are finding controlling screen time harder. Fifty per cent of parents are concerned about the data companies are collecting on children and young people's online activities. They also worry about children damaging their reputations, the pressures of children to spend money and the possibility of children being radicalised online.

9.15 Government is committed to continuing to support parents in preventing and dealing with online harms. Government welcomes the support provided by the UK Safer Internet Centre. They produce a free Safer Internet Day resource pack for parents and carers, most recently in February 2019, helping parents and carers understand the facts about online risks and have positive conversations with their children about staying safe online.

9.16 However, for adults, there is insufficient messaging or resources covering online media literacy. There is a need for further work to address issues such as the sharing of disinformation, catfishing (i.e. luring someone into a relationship by means of a fictional online persona), attacks on women online (particularly public figures), and the differing needs of people with disabilities when navigating information. We also recognise the need for improved coordination of activity. Ofcom is working with a number of partners to assess existing research and evidence about people's attitudes and understanding of being online. This will assist policy-makers to identify gaps and opportunities.

Online safety apps: NewsGuard Box 39

NewsGuard was first launched in the US in March 2018 by journalists to tackle the problem of disinformation online and is now available to UK users.

- NewsGuard rates and reviews news and information websites using nine standards of credibility and transparency, allocating a 'nutrition label' review which provides information on the site's ownership, financing, content, credibility, transparency and history.
- The NewsGuard desktop browser extension displays these 'nutrition labels' next to headlines in social media feeds and search results. This has been rolled out in libraries in the US, and Microsoft now offers the extension as an optional setting in the desktop and mobile versions of its Edge browser.

83 Ofcom (2018). Children and parents: Media use and attitudes report 2018. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/134907/Children-and-Parents-Media-Use-and-Attitudes-2018.pdf

An online media literacy strategy

9.17 Industry and government have a shared responsibility to empower users to manage their online safety. The new regulator will have oversight of industry activity and spend, and a responsibility to promote online media literacy.

9.18 Ahead of the new regulator, the government will develop an online media literacy strategy. The media literacy field is a broad one, and we will therefore consult widely, possibly through a new taskforce, in order to ensure its objectives are well informed by evidence and take account of existing work.

9.19 The first step will be a comprehensive mapping exercise to identify what actions are already underway, and to determine the objectives of an online media literacy strategy. This process will involve convening representatives from tech companies, regulators, libraries, civil society, academics and government to identify ways to strengthen existing provisions, as well as to identify what additional activity is needed to make progress against key objectives, which may include:

- Ensuring that users can be more resilient in dealing with mis- and disinformation, including in relation to democratic processes and representation.
- Equipping people to recognise and deal with a range of deceptive and malicious behaviours online, including catfishing, grooming and extremism.
- Ensuring people with disabilities are not excluded from digital literacy education and support.
- Developing media literacy approaches to tackling violence against women and girls online.

9.20 The strategy will also reflect the government's commitment to look at how to give the public confidence in online information so they are equipped to make their own decisions about the issues that matter. The government has already invested over £1 million in 2018/19 to deliver two initiatives in support of this:

- The new 'RESIST' counter-disinformation toolkit equips government, public service and partner country communicators with the knowledge and skills they need to identify, assess and respond to disinformation. It will help develop a strategic and consistent counter-disinformation capability, and help reduce the impact of disinformation campaigns on UK society and our national interests, in line with our democratic values.
- Government has launched a pilot public disinformation communications campaign.[84] This campaign provides the public with the skills they need to recognise and respond to disinformation, showing people how it can affect them and what they can do about it.

The role of the tech sector in empowering users

9.21 As set out above, we recognise that companies fund a range of valuable education and awareness activities. However, we believe there needs to be greater transparency about the level of investment, that all activity needs to be evaluated to ensure resources are directed at the most impactful initiatives, and that there should be greater coordination across industry to avoid duplication.

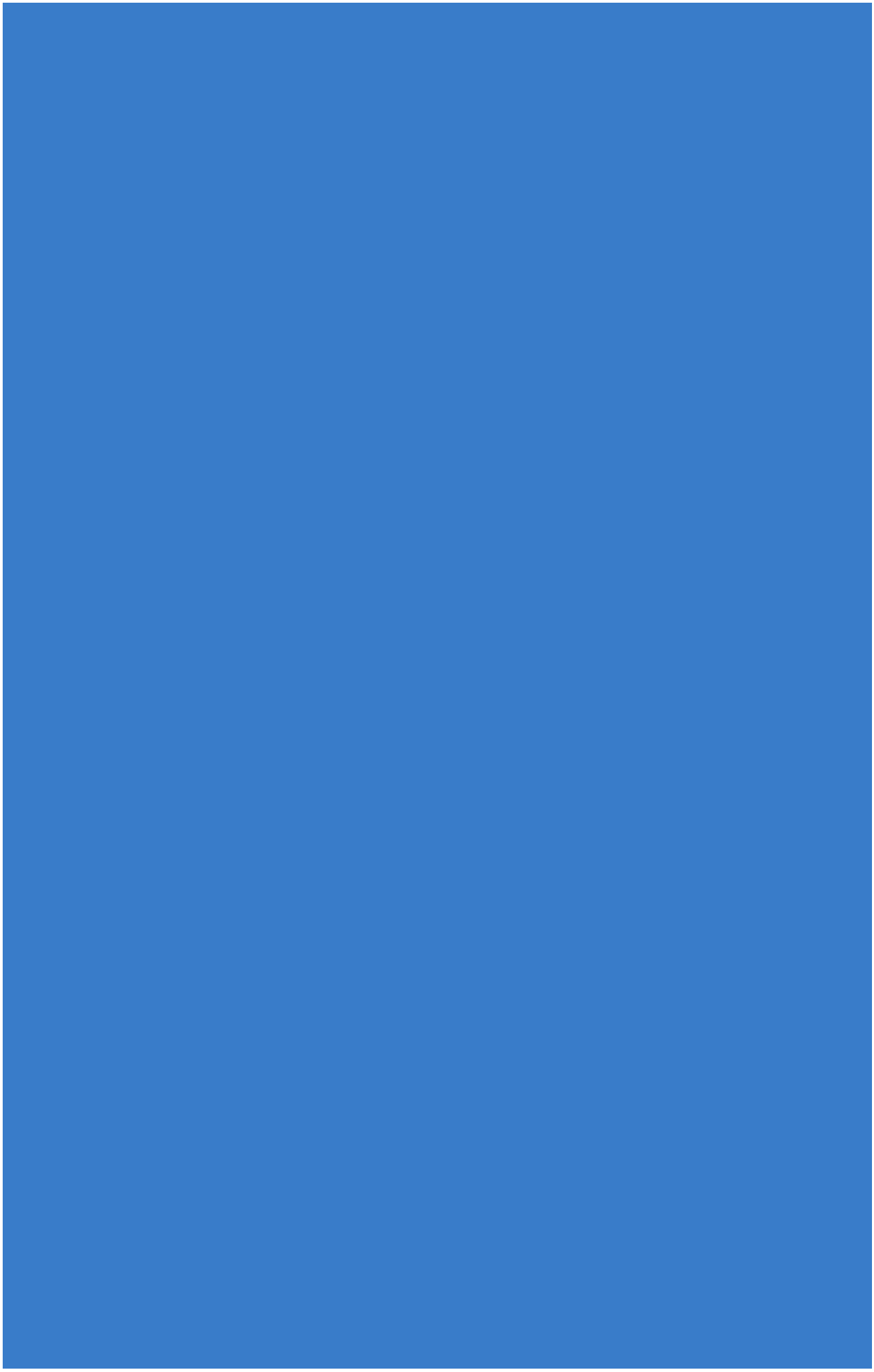
84 https://sharechecklist.gov.uk/

9.22 The new regulator will have the power to require companies to report on their education and awareness raising activity. We are consulting on appropriate powers for the regulator in this area.

Consultation questions

Question 17: Should the government be doing more to help people manage their own and their children's online safety and, if so, what?

Question 18: What, if any, role should the regulator have in relation to education and awareness activity?



Part 5: Conclusion and next steps

10: Conclusion and next steps

10.1 This White Paper sets out the UK's ambitious vision for online safety, including a new regulatory framework to tackle a broad range of harms; the development of a safety by design framework and support for innovation in safety technologies; and a new online media literacy strategy.

10.2 The measures outlined in this White Paper are novel and ambitious, with potentially far reaching effects for how our society engages with the internet. The UK remains committed to a multi-stakeholder model of internet governance as the best way to ensure a free, open and secure internet. All stakeholders from industry, civil society and government have a responsibility to help address legitimate online harms.

10.3 Given this, we want to engage with the widest possible audience on our proposals, and in particular invite views from industry, civil society, think tanks, campaigners and representatives. A series of consultation questions are posed throughout this document and they act as a basis for a formal consultation. We encourage respondents to provide not just their opinions, but also the supporting facts and reasoning to inform the evidence base for the development of our final proposals.

10.4 The consultation begins on 8 April 2019 and will close 12 weeks after it opens on 1 July 2019. We will then publish the government's response to this consultation on the GOV.UK website, summarising the responses received and setting out the action we will take, or have taken, in respect of them in developing our final proposals for legislation. Further information on responding to this consultation can be found in annex A.

10.5 DCMS and the Home Office will also run a series of engagement workshops to convene civil society actors and user groups. This will focus in particular on groups which are disproportionately affected by online harm and abuse. Given the formal and technical nature of the consultation, this will allow us to facilitate engagement with a wider audience.

10.6 Alongside this, we will continue to draw on advice from legal, regulatory, technical, online safety and law enforcement experts, to inform the further development of these proposals.

10.7 Finally, we are committed to continuing to build the evidence base for our proposals and will continue to work across government and with other stakeholders, including UKCIS, to commission a suitable programme of research.

Legislation

10.8 Following the publication of the Government Response to the consultation, we will bring forward legislation when parliamentary time allows.

Note on territorial scope

10.9 Internet services and their regulation is a reserved issue, therefore we intend for our proposed framework to apply on a UK wide basis. While some of the harms in the scope of this White Paper relate to offences in Scots or Northern Ireland Law, and therefore involve devolved competencies (such as child protection), we are not seeking to change the law in relation to these offences but rather to clarify the responsibility of companies to tackle this activity on their services. Education policy is devolved in Wales, Scotland and Northern Ireland.

10.10 As part of the wider process of consultation, we will engage with the Devolved Administrations on the proposals in this White Paper. This consultation will consider in particular the implications for law enforcement, and explore how we can advance a cohesive UK-wide approach to educating children and adults about online safety.

Annex A: How to respond to the consultation

We are inviting individuals and organisations to provide their views by responding to the questions set out throughout this White Paper. The questions are listed below.

The consultation will be open for 12 weeks, from 8 April 2019 to 23:59 1 July 2019.

You can respond online via the following link:
https://dcms.eu.qualtrics.com/jfe/form/SV_5nm7sPoxilSoTg9

If you prefer, you can also email your response to:
onlineharmsconsultation@culture.gov.uk

Or you can write to us at:

Online Harms Team
DCMS
100 Parliament Street
London
SW1A 2BQ

Consultation Questions:

Question 1: This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

Question 2: Should designated bodies be able to bring 'super complaints' to the regulator in specific and clearly evidenced circumstances?

> Question 2a: If your answer to question 2 is 'yes', in what circumstances should this happen?

Question 3: What, if any, other measures should the government consider for users who wish to raise concerns about specific pieces of harmful content or activity, and/or breaches of the duty of care?

Question 4: What role should Parliament play in scrutinising the work of the regulator, including the development of codes of practice?

Question 5: Are proposals for the online platforms and services in scope of the regulatory framework a suitable basis for an effective and proportionate approach?

Question 6 In developing a definition for private communications, what criteria should be considered?

Question 7: Which channels or forums that can be considered private should be in scope of the regulatory framework?

> Question 7a: What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?

Question 8: What further steps could be taken to ensure the regulator will act in a targeted and proportionate manner?

Question 9: What, if any, advice or support could the regulator provide to businesses, particularly start-ups and SMEs, comply with the regulatory framework?

Question 10: Should an online harms regulator be: (i) a new public body, or (ii) an existing public body?

> Question 10a: If your answer to question 10 is (ii), which body or bodies should it be?

Question 11: A new or existing regulator is intended to be cost neutral: on what basis should any funding contributions from industry be determined?

Question 12: Should the regulator be empowered to i) disrupt business activities, or ii) undertake ISP blocking, or iii) implement a regime for senior management liability? What, if any, further powers should be available to the regulator?

Question 13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

Question 14: In addition to judicial review should there be a statutory mechanism for companies to appeal against a decision of the regulator, as exists in relation to Ofcom under sections 192-196 of the Communications Act 2003?

> Question 14a: If your answer to question 14 is 'yes', in what circumstances should companies be able to use this statutory mechanism?
>
> Question 14b: If your answer to question 14 is 'yes', should the appeal be decided on the basis of the principles that would be applied on an application for judicial review or on the merits of the case?

Question 15: What are the greatest opportunities and barriers for (i) innovation and (ii) adoption of safety technologies by UK organisations, and what role should government play in addressing these?

Question 16: What, if any, are the most significant areas in which organisations need practical guidance to build products that are safe by design?

Question 17: Should the government be doing more to help people manage their own and their children's online safety and, if so, what?

Question 18: What, if any, role should the regulator have in relation to education and awareness activity?

HM Government

ISBN 978-1-5286-1080-3
CCS0219683420