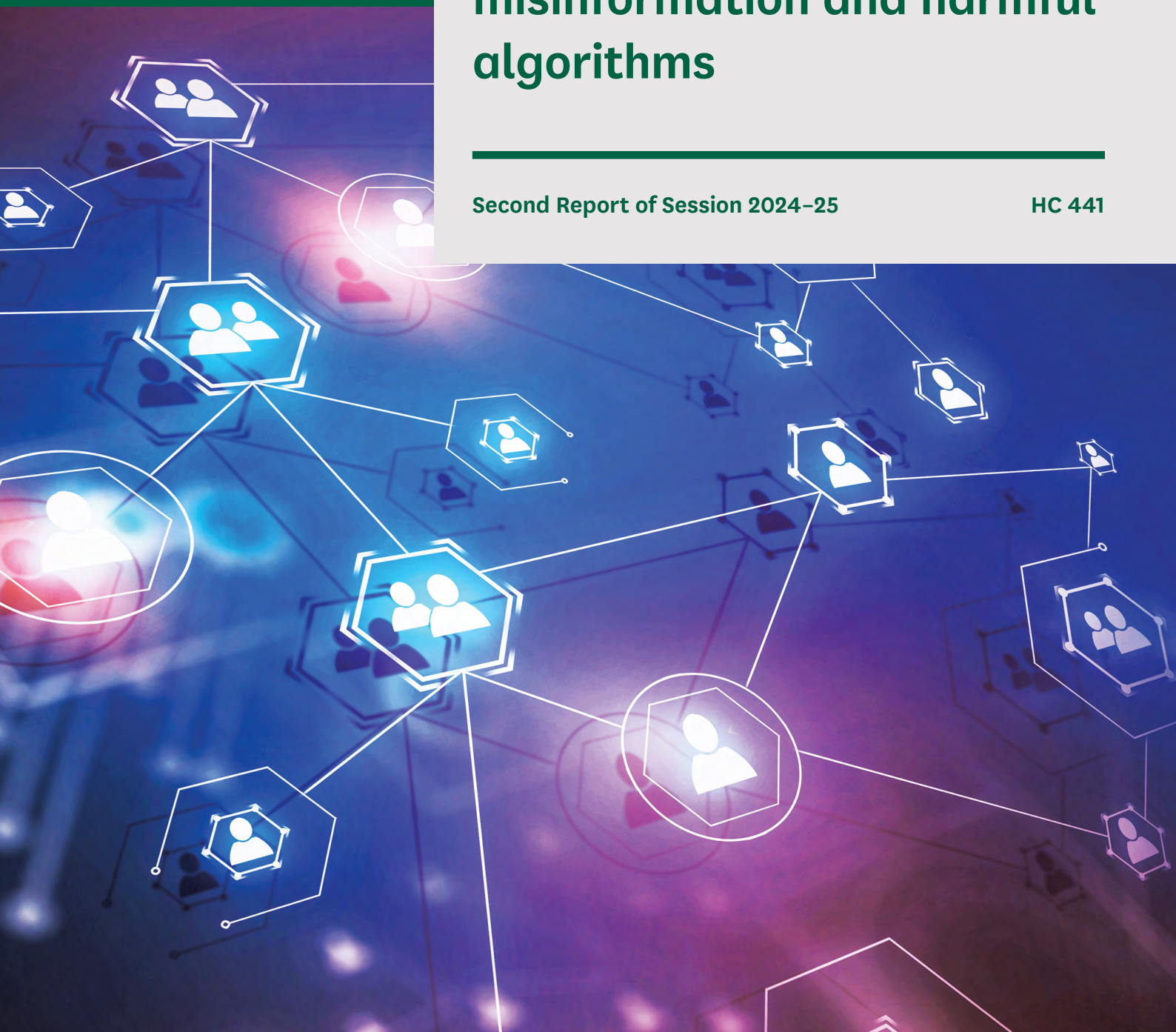
House of Commons

Science, Innovation and Technology Committee

# Social media, misinformation and harmful algorithms

Second Report of Session 2024–25 HC 441



# Science, Innovation and Technology Committee

The Science, Innovation and Technology Committee is appointed by the House of Commons to examine the expenditure, administration, and policy of the Department for Science, Innovation and Technology and its associated public bodies. It also exists to ensure that Government policies and decision-making across departments are based on solid scientific evidence and advice.

## Current membership

Dame Chi Onwurah (Labour; Newcastle upon Tyne Central and West) (Chair)

Emily Darlington (Labour; Milton Keynes Central)

George Freeman (Conservative; Mid Norfolk)

Dr Allison Gardner (Labour; Stoke-on-Trent South)

Tom Gordon (Liberal Democrat; Harrogate and Knaresborough)

Kit Malthouse (Conservative; North West Hampshire)

Jon Pearce (Labour; High Peak)

Steve Race (Labour; Exeter)

Dr Lauren Sullivan (Labour; Gravesham)

Adam Thompson (Labour; Erewash)

Martin Wrigley (Liberal Democrat; Newton Abbot)

## Powers

The Committee is one of the departmental select committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No. 152. These are available on the internet via www.parliament.uk.

## Publication

This Report, together with formal minutes relating to the report, was Ordered by the House of Commons, on 24 June 2025, to be printed. It was published on 11 July 2025 by authority of the House of Commons.


Committee reports are published on the Committee's website at www.parliament.uk/CommonsSITC and in print by Order of the House.

## Contacts

All correspondence should be addressed to the Clerk of the Science, Innovation and Technology Committee, House of Commons, London SW1A 0AA. The telephone number for general enquiries is 020 7219 5023; the Committee's email address is commonssitc@parliament.uk. You can follow the Committee on X (formerly Twitter) using @CommonsSITC.

# Contents

# Summary

The UK government—like its counterparts around the world—is facing the challenge of attempting to regulate hugely powerful technology companies that operate across national borders, providing technologies that transform societies, with bigger budgets than many countries. It is essential that their impact on our society be understood, effectively scrutinised and, where necessary, regulated in the public interest by Parliament. We have experienced some of the challenges of this in the course of the present inquiry. We were reassured by the statements from Google, Meta, TikTok and X in our evidence session that they accepted their responsibility to be accountable to Parliament. We hope to see this in practice as we continue our work in this area.

After the horrific murders in Southport on 29 July 2024, misleading and hateful messaging proliferated rapidly online, amplified by the recommendation algorithms of social media companies. Protests turned violent, often targeting Muslim and migrant communities, driven in part by the spread of these messages.

These events provide a snapshot of how online activity can contribute to real world violence and hate. Many parts of the long-awaited Online Safety Act were not fully in force at the time of the unrest, but we found little evidence that they would have made a difference if they were. Moreover, the Act is already out of date, failing to adequately address generative AI—a technology evolving faster than governments can legislate—which could make the next misinformation crisis even more dangerous. Regulating technology alone is not sufficient—our online safety regime should be based on principles that remain sound in the face of technological development.

Social media has made important positive contributions, helping to democratise access to a public voice, but it comes with huge risks. The advertisement-based business models of most social media companies mean that they promote engaging content, often regardless of its safety or authenticity. This spills out across the entire internet, via the opaque, underregulated digital advertising market, incentivising the creation of content that will perform well on social media—as we saw in the 2024 unrest.

Our concerns were exacerbated when we questioned representatives of regulators and the government, as we were met with confusion and contradiction at high levels. It became clear that the UK's online safety regime has some major holes.

The Online Safety Act was a first step. However, more is needed to protect UK citizens from online harms. In the course of this inquiry, we identified five key principles that we believe are crucial for regulation of social media and related technologies. They are set out below, with the key specific recommendations that follow from each one:

1. **Public safety**: Algorithmically accelerated misinformation is a danger that companies, government—both national and local, law enforcement, and security services need to work together to address.
    - Platforms should algorithmically demote fact-checked misinformation, with established processes setting out more stringent measures to take during crises.
    - More research is needed into how platforms should tackle misinformation, and how far recommendation algorithms spread harm.
    - All AI-generated content should be visibly labelled.
2. **Free and safe expression**: Steps to tackle amplified misinformation should be in line with the fundamental right to free expression.
    - Measures to meet misinformation must be aligned with the right to free expression.
3. **Responsibility**: Users should be held liable for what they post online, but the platforms they post on are also responsible.
    - Platforms should be held accountable for the impact from amplification of harmful content.
    - Platforms should undertake risk assessments and report on content that is legal but harmful.
    - New regulatory oversight, clear and enforceable standards, and proportionate penalties are needed to cover the process of digital advertising.
4. **Control**: Users should have control over both their personal data and what they see online.
    - Users should have a right to reset the data used by platform recommendation algorithms.

5. **Transparency**: The technology used by platform companies should be transparent, accessible and explainable to public authorities.

   - Recommendation algorithms and generative AI should be fully transparent and explainable to public authorities and independent researchers.
   - Transparency is needed for participants in the digital advertising market.

# 1 Introduction

**1.** In the days following the horrific murders in Southport on 29 July 2024, demonstrations and riots took place across the UK. Many were violent, targeting mosques and migrants.[1] Following the attacks, false or unfounded information about the suspect—that he was a Muslim and/or an asylum seeker—spread rapidly online alongside anti-Muslim and anti-migrant rhetoric.[2] Calls to violence were posted across major platforms, in some cases seemingly amplified by recommendation algorithms.[3] Social media and encrypted private messaging platforms were used to organise protests and riots.[4] Generative AI was used to spread misleading content and to boost the algorithmic profile of certain posts.[5] The Home Office cited the "online environment" as a significant factor in inciting violence.[6]

**2.** This viral spread of harmful misinformation took place despite the UK's recent effort to pass legislation tackling online harms. The Online Safety Act received Royal Assent in October 2023 after a six-year passage from green paper to statute book.[7] The Act places an independent regulator—Ofcom—in charge of overseeing social media and its providers, with powers to require information and impose penalties for non-compliance.[8]

**3.** Shortly after Parliament established this committee, we launched an inquiry into the online spread of misinformation following the Southport attack, the role of social media business models, recommendation algorithms and other technologies, and how the Online Safety Act relates to it. We also looked ahead to how the rapid development of these technologies could influence future, similar crises, and what government is doing to address this. This report focuses on the technological aspects of online services and markets that can lead to the amplification of false, unfounded or harmful

1 Home Affairs Committee, Second Report of Session 2024–25, 'Police response to the 2024 summer disorder', HC 381, para 3, 5
2 Center for Countering Digital Hate (SMH0009); Institute for Strategic Dialogue (SMH0062); Marc Owen-Jones (SMH0071)
3 Online Safety Act Network (SMH0031); Marc Owen-Jones (SMH0071)
4 Center for Countering Digital Hate (SMH0009); Clean Up The Internet (SMH0023); Institute for Strategic Dialogue (SMH0062)
5 Dr Mihaela Popa-Wyatt (SMH0045); Marc Owen-Jones (SMH0071)
6 Written evidence received for the Home Affairs Committee's inquiry into Police response to the 2024 summer disorder, Home Office (SDR0015)
7 Department for Science, Innovation and Technology, Home Office, Ministry of Justice, UK children and adults to be safer online as world-leading bill becomes law, 26 October 2023
8 Ss 6, 91, 139–143, Online Safety Act 2023

information. We are limited in the scope of issues we can address in this report, but we recognise the complexity of ethical debates surrounding free expression online and offline, and the moral and legal limitations that can be placed on it.

**4.** In choosing to focus on the part that social media played in the disorder last summer, we have not lost sight of the tragic events that preceded it. Our thoughts remain with those harmed by the appalling attack in Southport, as well as in the unrest that followed.

**5.** In the course of this inquiry, we held four evidence sessions, a private roundtable, received an expert briefing on social media algorithms, and visited the BBC to hear about its approach to these topics. The evidence sessions included representatives of:

- Groups affected by the riots, local government in an affected area, and experts on online narratives and disinformation;
- Major tech companies;
- Experts on digital advertising, fact-checking and online safety;
- Regulators and the government.

We received more than 80 pieces of written evidence, including from experts, campaigners and members of the public. We are grateful to all those who engaged with our inquiry.

**6.**

**CONCLUSION**

In the course of this inquiry, we identified five key principles that we believe are crucial for regulation of social media and related technologies:

1) Public safety: Algorithmically accelerated misinformation is a danger that companies and government need to address—the government and platform companies should work together to protect the public from it.

2) Free and safe expression: Neither government nor private companies should be arbiters of truth. Steps to tackle amplified misinformation should be in line with the fundamental right of free expression, with restrictions where proportionate and necessary to protect national security, public safety, health, or to prevent disorder and crime.

3) Responsibility: Users should be held liable for what they post online, but the platforms they post on are also responsible, especially with regard to the systems used to moderate, circulate or amplify content.

4) Control: Users should have control over both their personal data and what they see online. This includes the right to delete the data stored by platforms and services which is used to drive content and advertisement recommendation algorithms.

5) Transparency: The technology used by platform companies, including social media algorithms, has huge public safety implications, and should be transparent and accessible to public authorities.

## Note on definitions

### Harmful content

The Online Safety Act defines "harm" as physical or psychological harm. Content that is harmful includes harm arising from the nature of the content and the fact or manner of its dissemination.[9] In this report, we use "harmful content" as a broad category that can include content that is hateful, extreme or dangerous, as well as misinformation.[10]

### Misleading content

"Misinformation" can be defined as verifiably false information that is shared without an intent to mislead, and "disinformation" as verifiably false information that is shared with an intent to deceive.[11] In this report, we consider disinformation to be a sub-category of misinformation, and use terms such as "false" or "misleading" to describe different types of misinformation.

### Platforms and companies

This report uses the term "platform" to describe an online service that hosts content. The term "company" is used to describe the business that owns and runs that platform. The term "platform company" is used to describe an online service that facilitates interactions between multiple groups, such as users, advertisers and third-party developers.

9 S 234, Online Safety Act 2023

10 Ofcom's Online Experience Tracker questionnaire identifies the following categories of potentially harmful online content: hate and abuse, sexual content and exploitation, violence and extremism, child safety risks, mental health and body image harms, illegal or unsafe products and services, privacy violations and scams, manipulative or misleading content, and offensive or inappropriate language. Ofcom, Questionnaire—Online Experience Tracker (accessed June 2025), pp 13–14

11 Government Communication Service, RESIST 2 Counter Disinformation Toolkit, accessed June 2025

## Technology

An “algorithm” is set of instructions a computer follows to perform tasks or solve problems. A recommendation algorithm is a type of algorithm designed to suggest data or content based on patterns in user behaviour or preferences. A social media recommendation algorithm selects and promotes content or accounts that it predicts will engage users, shaping individual feeds and influencing what users see, usually aiming to increase engagement and time spent on the platform.

“Generative AI” refers to algorithms that can create new content. Large language models (LLMs) are a form of generative AI that produce text-based content, trained on large sets of data to create the most statistically likely textual answer. LLM-based assistants or conversational agents, often referred to as ‘chatbots’, are software applications built using LLMs in order to simulate human conversation.

# 2 Misleading and harmful content on social media

## Online activity and the 2024 unrest

### Misinformation and harmful messaging

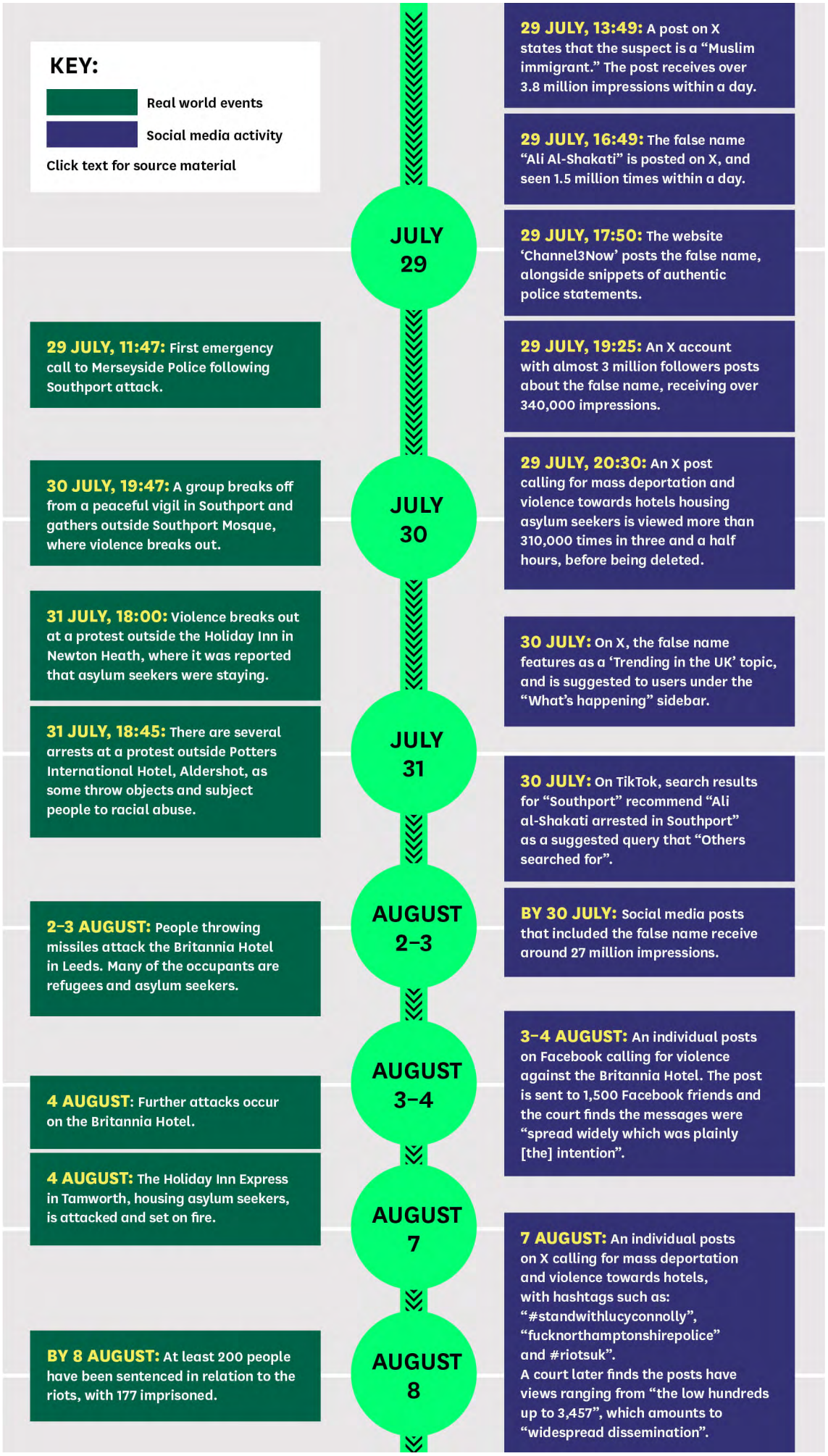
**7.** Within hours of the Southport attack, misleading and unfounded information started to circulate online, stating that the killer was a Muslim asylum seeker, with the name "Ali Al-Shakati".[12] The following day, the police issued a statement saying that this was not the name of the suspect, but were limited in how much information they could release.[13] The Home Affairs Committee concluded that the lack of information "created a vacuum where misinformation was able to grow."[14]

12 Marc Owen Jones (SMH0071)

13 Identity of the subject could not be published under Section 49, Children and Young People Act 1933. Information such as the suspect's previous conviction and Prevent referrals could not be published under Section 2, Contempt of Court Act 1981. CPS and Merseyside Police discussed the potential to release information on the suspect's religious background, due to hate crimes against Muslim communities. This was ultimately not released. Merseyside Police, Update on major incident in Southport, 30 July 2024; Letter from the Crown Prosecution Service to the Chair of the Home Affairs Select Committee regarding process and guidance around publication of prosecution information, 21 February 2025

14 Home Affairs Committee, Second Report of Session 2024–25, 'Police response to the 2024 summer disorder', HC 381, para 17

## Timeline of information published after Southport attack:



**8.** The claims achieved a viral reach on social media. Between 29 July and 9 August, false or unfounded claims about the Southport attacker achieved 155 million impressions on X.[15] Across social media, the false name was seen 420,000 times, with a potential reach of 1.7 billion people.[16] It was directly promoted by social media algorithmic tools, featuring on X's 'Trending in the UK' and TikTok's 'Others searched for' features.[17] The number of social media posts with the keyword 'Muslim' rose by 242%,[18] while social media posts about migrants increased threefold.[19] On X, anti-migrant content received an estimated 31.1 million impressions between 29 July and 9 August.[20]

**9.** The Joint Council for the Welfare of Immigrants (JCWI) told us that misinformation acts like a virus on digital platforms, aided by individuals who act as "superspreaders" as it mutates and cuts through to more people.[21] However, they, and others, warned against blaming social media alone for the violence:[22]

> In this case, X was the message carrier. If pigeons, like in the olden days, were the message carrier, would we blame the pigeons? [ … ] We need to look at society as a whole.[23]

15 Marc Owen Jones (SMH0071)

16 "Potential reach" refers to the number of followers and/or page likes where the key word appeared. Independent, How fake claims over Southport suspect spread like wildfire with false name seen more than 420,000 times, 17 August 2024

17 Institute for Strategic Dialogue (SMH0062)

18 Independent, Fake claims over Southport suspect, 17 August 2024

19 Posts about migrants refers to 'immigration keywords' such as 'migration' 'immigrant' 'migrant' or 'asylum seeker.' The Muslim Council of Britain told us about the "collective fear and trauma" that Muslims experienced in the period, citing the algorithms that boosted "more of what you hated, as well as more of what was targeting you." The charity Tell Mama estimated that in 2024, anti-Muslim hate rose by 43% in the UK. Independent, Fake claims over Southport suspect, August 2024; Q1 [Zara Mohammed]; Tell Mama, The New Norm of Anti-Muslim Hate, 19 February 2025

20 Marc Owen Jones (SMH0071)

21 Marc Owen Jones: "Certain accounts repeatedly engage in false and xenophobic or hateful narratives [...] A component of Algorithm manipulation involves highly followed and influential accounts amplifying false content. This directly impacts what is seen, and how popular it is." Mathematic modelling has been used to suggest that misinformation spreads across the internet similar to a virus spreading through a population. Q2 [Ravishaan Muthiah]; Marc Owen Jones (SMH0071); S. van der Linden, D. R. Grimes, Misinformation really does spread like a virus, suggest mathematical models drawn from epidemiology, 26 November 2024

22 Qq2–5 [Zara Mohammed; Ravishaan Muthiah]; Computer and Communications Industry Association (SMH0029); Oxford Internet Institute (SMH0057)

23 Q2 [Ravishaan Muthiah]

**10.** Social media was used to organise the unrest, with accounts created for the riots including far-right symbols and calls for "mass deportation."[24] Closed groups on encrypted platforms such as Telegram and WhatsApp were used to coordinate and incite violence.[25]

## Social media platforms' responses

**11.** Meta, TikTok, Google and X told us they each set up crisis protocols and internal working groups to monitor the unrest and related online activity, engaged with government and law enforcement, and removed posts and suspended accounts that violated their policies on hateful speech, inauthentic behaviour or dangerous organisations.[26] DSIT stated it worked with major platforms to tackle content "contributing" to the disorder in this period.[27] The companies said they took steps, including:

- Meta: Removed 24,000 posts for rules on violence and incitement, 12,000 for hate speech rules, and 2,700 for rules on hate organisations. Reduced visibility of misleading content as marked by fact-checkers and set up a "trending event" for fact-checkers to find related content.[28]

- TikTok: Established a worldwide command centre with over 100 people working 24/7, added search interventions to block harmful queries and direct users to further resources, and "oversaw the removal of tens of thousands of videos and comments." TikTok said the false name of the attacker was removed from the suggested search feature the evening it appeared and removed as a search result the next day. The company said it would have would have liked the response to have been faster, but that its focus was on moderating video content.[29]

---

24 Institute for Strategic Dialogue (SMH0062)

25 Center for Countering Digital Hate (SMH0009); Clean Up The Internet (SMH0023)

26 Meta designated the riots under its Crisis Policy Protocol on August 5 2025, neither X nor TikTok provided a date for when their protocols were triggered; Meta (SMH0037); X (SMH0064); Google (SMH0065); TikTok (SMH0068); Meta (SMH0080)

27 UK Government (SMH0061)

28 Meta stated it "did not experience the same level of riot organisation or coordination" as other platforms. Separately, Meta's Oversight Board (an independent body set up by Meta to make binding decisions on content moderation across its platforms) found that the company erroneously left three posts up during the unrest that were reported for hate speech and incitement to violence, and noted "strong concerns about Meta's ability to accurately moderate hateful and violent imagery." Meta (SMH0037); Meta (SMH0080); Oversight Board, New Cases Involve Posts Shared in Support of the UK Riots, 3 December 2024; Oversight Board, Posts Supporting UK Riots, 23 April 2025

29 TikTok stated that the majority of content during the protests was documentary or bystander footage, but there was "some rise in violations, particularly around hate and misinformation". TikTok (SMH0068); TikTok (SMH0081); Qq93, 101 [Ali Law]

- X: Had a "vigilant" team monitoring platform conversation and on-ground developments, including "proactive action" on "tens of thousands" of pieces of content, and users adding 'Community Notes' context and fact-checking to certain posts. X told us it had been "widely reported" that advertisers paused activity during the period, but declined to share further details.[30]
- Google/YouTube: The Trust and Safety Team began monitoring YouTube on July 29, and by 13 August had removed two channels under violent extremism or criminal organisations policy, and one under spam and deceptive practices policies.[31]

**12.** Ofcom told us, of the unrest:

> "we do not think that companies are sufficiently, consistently or effectively responding to events of this kind."[32]

We received evidence that, during the unrest, platforms were often slow to react, unwilling or unable to moderate algorithmic amplification of harmful content, allowed false content and hate speech to be published, and failed to uphold their terms of service.[33] Some evidence criticised X's Community Notes system for failing to mark certain posts as misleading in this period.[34] Ofcom found that "Most online services took rapid action in response to the situation, but responses were uneven."[35] It assured us that it would require companies to implement crisis response protocols and would expect them "to be much more accountable for their response than they have been in the past."[36]

30 When asked to clarify whether advertisers paused activity during the unrest, X stated that "details regarding X's customers and clients are sensitive business information and in the interest of our customer's data and business privacy, we cannot disclose any further details." X (SMH0064); X (SMH0082); Qq90, 99–100, 159, 162 [Wifredo Fernandez]; ACM Europe Technology Policy Committee (SMH0035); Center for Countering Digital Hate, X ran ads on five accounts pushing lies and hate during UK riots, 19 August 2024

31 Google (SMH0065)

32 Q258 [Mark Bunting]

33 Center for Countering Digital Hate (SMH0009); Clean Up the Internet (SMH0023); Institute for Strategic Dialogue (SMH0062); Ofcom (SMH0078)

34 Only 0.1% of 1,060 posts from five key accounts (Andrew Tate, Laurence Fox, Calvin Robinson, Ashlea Simon and Paul Golding) posting harmful or misleading content during the summer unrest displayed a 'Community Note'. Harmful and misleading posts included "An illegal migrant arrived on a boat one month ago" and "We need to permanently remove Islam from Great Britain", among others. Center for Countering Digital Hate, X ran ads on five accounts pushing lies and hate during UK riots, 19 August 2024. X's system of 'Community Notes' is discussed further below, under How platforms moderate harmful and misleading content, Misleading content

35 Ofcom, Letter from Dame Melanie Dawes to the Secretary of State, 22 October 2024, p 2

36 Q258

## Advertising and the profit incentive

**13.** We heard evidence that many social media platforms likely profited from the increased engagement at the time of the unrest, and that they are "less likely to moderate [ … ] high engagement content."[37] As advertising is by far the dominant revenue stream for most social media companies, engagement is key.[38] The Center for Countering Digital Hate (CCDH) found that certain right-wing figures amassed 38.9 million ad impressions (views) on X in the week following the attack, which could have generated £27,976 in daily ad revenue.[39] The organisation's CEO Imran Ahmed told us that "the problem is that Southport was profitable for X and the other platforms. We need to make sure that in the future it is costly".[40] When we put this to X, Meta and TikTok, they all denied profiting from the unrest.[41] Alphabet, the parent company of Google, also made nearly 78% of its revenue from advertising in 2024.[42]

**14.** **CONCLUSION**

We launched this inquiry in the wake of the riots that followed the horrific attack in Southport in 2024. We received overwhelming evidence that online activity, including social media recommendation algorithms amplifying harmful and misleading content, played a key part in driving the unrest and violence. Social media companies' responses were inconsistent and inadequate, often enabling, if not encouraging, this viral spread, with evidence that they may have profited due to the heightened engagement. The evidence supports the conclusion that social media business models incentivise the spread of content that is damaging and dangerous, and did so in a manner that endangered public safety in the hours and days following the Southport murders.

37 Ravishaan Muthiah: "social media companies [...] that rely on engagement and impressions to boost their own advertising revenue are less likely to moderate this high engagement content. That ultimately leads to the loudest and most polarising views receiving the most digital airtime". REPHRAIN: "These algorithms prioritise content that elicits strong emotional responses—such as outrage, anger, or surprise—because such reactions keep users engaged for longer [...] in cases of public unrest or riots, misinformation that exploits stereotypes or entrenched biases often goes viral, further escalating tensions. This was the case for the 2024 summer riots in the UK, a direct consequence of platforms optimising for content likely to trigger strong emotional reactions and engagement." Q2; REPHRAIN (SMH0033)

38 The business models of social media platforms are discussed further in Chapter 4, *Digital advertising market*

39 Center for Countering Digital Hate, X ran ads on five accounts pushing lies and hate during UK riots, 19 August 2024

40 Q25

41 Q90 [Wifredo Fernandez, Chris Yiu, Ali Law]

42 Statista, Distribution of Google segment revenues from 2017 to 2024 (accessed June 2025); Google's role in online advertising and the summer unrest is discussed further in Chapter 4, *Digital advertising market*

## The Online Safety Act

**15.** We heard that much of the misleading or harmful content that drove the unrest last summer would not have been covered by the Online Safety Act even if it had been in full force.[43] When we questioned representatives of Meta, TikTok, and X, they were unable to say if or how the Act would have changed their response to the unrest.[44]

**16.** Ofcom confirmed that the Act is not designed to tackle the spread of "legal but harmful" content such as misinformation but said that, if it had been in place, platforms would have had to answer "a number of questions" about risk assessments and crisis response mechanisms.[45] Baroness Jones, the minister responsible for online safety, argued the Act would have made a "real" and "material" difference, as it would have allowed Ofcom to insist that illegal posts be taken down.[46]

**17.** We heard calls for platforms to be compelled to include "demotion" and "de-amplification" measures to automatically limit the reach of misinformation—where it has been verified as such through fact-checking—without taking it down.[47] Some contributors to this inquiry pointed to the EU's 2022 Digital Services Act, which compels platforms to take measures during "extraordinary circumstances", including adapting their algorithmic

43 Illegal content duties would have been engaged for priority illegal offences such as explicit calls for violence, attacks on property, and racial hatred. Immigration status is not a protected characteristic and would not be covered by the illegal content duties. One X user was arrested under Section 179 (False Communications Offence) for spreading false information, however was ultimately not charged due to insufficient evidence. Misleading or unfounded content surrounding the attacker's identity would not have been covered, nor would certain legal but harmful content surrounding Muslims or migrants. The Act also includes no measures related to algorithmic demotion in times of crisis. Center for Countering Digital Hate (CCDH) (SMH0009); Professor Jeffrey Howard, and Dr Maxime Lepoutre (SMH0013); Online Safety Act Network (SMH0031); Full Fact (SMH0047); Institute for Strategic Dialogue (SMH0062); BBC News, No charge over spreading of Southport misinformation, 18 September 2024; Qq259, 304–5 [Mark Bunting; Baroness Jones]

44 Qq170–73 [Chris Yiu, Ali Law, Wifredo Fernandez]

45 Mark Bunting: "[…] but should anything similar happen again, you are right, of course, that the previous Government made a decision to remove legal material that might be harmful to adults from the scope of the Act, including other forms of misinformation" Qq258–9

46 Qq 304–5, 308

47 Antisemitism Policy Trust (SMH0005); Professor Jeffrey Howard, Dr Maxime Lepoutre (SMH0013)

and advertising systems.[48] Ofcom announced in December 2024 that it would consult on "Crisis response protocols for emergency events" (such as last summer's riots).[49]

**18.** **CONCLUSION**
The Online Safety Act was not designed to tackle misinformation—we heard that even if it had been fully implemented, it would have made little difference to the spread of misleading content that drove violence and hate in summer 2024. Therefore, the Act fails to keep UK citizens safe from a core and pervasive online harm.

**19.** **RECOMMENDATION**
We welcome Ofcom's consultation on a 'crisis response protocol' for companies to follow in response to events such as the 2024 unrest. The protocol should directly address misinformation by including all online services at risk of contributing to the spread of false or harmful information, including large online social media, search and messaging services; those with smaller user numbers but high-risk profiles; and others, such as generative AI platforms. In establishing the mechanism, Ofcom should acknowledge the different ways in which different services operate. Following our Principle 2, it should hold platforms responsible for: decelerating the spread of harmful misinformation without censoring lawful speech; ensuring substantial and continuous engagement with law enforcement and government bodies; giving users control over the content they see; and providing transparency around their actions.

## Misleading and harmful content—beyond the 2024 unrest

**20.** The viral spread of misinformation and hateful content at the time of the Southport riots forms part of a wider pattern. We received compelling evidence setting out the high levels of misleading and harmful content that many internet users encounter online, outside of crisis situations.[50] Technology such as recommendation algorithms can create "echo

48 Other measures include allocating more resources to content moderation, modifying terms and conditions, increased collaboration with fact-checkers, and promoting trusted information. Online Safety Act Network (SMH0031); Beatriz Kira, Zoe Asser, Phoebe Li and Julie Weeds (SMH0056); EU, Article 36, Digital Services Act 2022

49 Ofcom, Overview, 15 December 2024

50 Ofcom found that as of June 2024, 68% of adult online users have experienced potentially harmful content, with 26% reporting seeing hateful, offensive or discriminatory content, and 39% of UK users aged 13+ have reported encountering misinformation online. Evidence raised that often individuals accept information online at face value, without

chambers" that normalise extreme content and behaviour.[51] Young people are particularly susceptible to misleading and harmful content. There is evidence that children are vulnerable to online radicalisation due to the stage of their cognitive development, and that they spend more time online, drawn in by a "feedback loop" of algorithmically reinforced content.[52] This can have devastating effects–the Molly Rose Foundation estimates that technology plays a role almost one-quarter of deaths by suicide among those aged 10 to 19.[53] Marianna Spring of the BBC told us:

> You see a video or post that has had thousands of views—millions, in some cases. It is where you take your social cues from, so you start to think that a particular narrative or rhetoric is normal and widespread, even if that is not necessarily the case.[54]

## Social media recommendation algorithms and harm

**21.** We heard that social media algorithms can play a major role in promoting misinformation and harmful content. The design principle of maximising engagement for profit means that algorithms can amplify content regardless of accuracy or potential for harm. Indeed, harmful and false content is often designed to be engaging, so may be promoted more than other types of content.[55] Examples include mis/disinformation, violence, extremism, prejudiced views, suicide and self-harm content.[56]

---

proper scrutiny, leading to a lack of trust in communications from official sources or public institutions. Ofcom, Online Nation 2024 Report, 28 November 2024, p 7; Dr Aine MacDermott (SMH0010); Faculty of Public Health (SMH0011)

51 Dr Aine MacDermott (SMH0010); Faculty of Public Health (SMH0011); Professor Keith Hyams and Dr Jessica Sutherland (SMH0015); The Electoral Commission (SMH0021); Andreu Casas, Georgia Dagher, and Ben O'Loughlin (SMH0030)

52 Dr Áine MacDermott (SMH0010); Office of the Children's Commissioner for England (SMH0014); 5Rights Foundation (SMH0024); NSPCC (SMH0032): UK Safer Internet Centre (SMH0044)

53 Molly Rose Foundation (SMH0016)

54 Q16

55 Faculty of Public Health (SMH0011); Swansea University Cyber Threats Research Centre (SMH0018); Clean Up The Internet (SMH0023); Digital Mental Health Programme at the University of Cambridge (SMH0027); Andreu Casas, Georgia Dagher, and Ben O'Loughlin (SMH0030); Atlantic Council's Democracy + Tech Initiative (SMH0034); Minderoo Centre for Technology and Democracy, University of Cambridge (SMH0051); Qq12, 22 [Marianna Spring, Imran Ahmed, Dr Whittaker]

56 The Youth Select Committee found that social media companies' business models and algorithms "may result in companies inadvertently promoting violent content or content that incites violence." Youth Select Committee, Youth Violence and Social Media, 27 March 2025, para 14; Marc, Chaslot, H Farid, A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos, 6 March 2020, The Washington Post, Misinformation on Facebook got six times more clicks than factual news during the 2020 election, study says, 4 September 2021; TrustLab, Code of Practice on Disinformation: A Comparative Analysis of the Prevalence and Sources of Disinformation across Major Social Media Platforms in Poland,

**22.** When we put this to major tech companies, they told us that there is no business incentive to allow harmful content on their platforms, as it can damage the brand, repelling advertisers. They said that their recommendation algorithms are explicitly designed to reduce exposure to content that is harmful and violates their terms of service.[57] Indeed, there is some evidence that recommendation algorithms can provide beneficial content to users.[58]

**23.** Little is known about the inner workings of social media recommendation algorithms, as it is closely-guarded intellectual property.[59] We asked several tech companies to provide high-level representations of their recommendation algorithms to the committee, but they did not.[60] As a result of this confidentiality, independent third-party research generally relies on output-based "black box" studies to assess harm.[61] This shortfall in transparency, as well as methodological issues, such as difficulties in establishing clear causal links between recommendations and harm, make it difficult to assess the true extent of algorithmic amplification of harmful content.[62] Baroness Jones confirmed to us that policymaking in this space has lacked a full evidence base.[63]

---

Slovakia, and Spain, September 2023; University of Chicago Biological Sciences Division, Health information on TikTok: The good, the bad and the ugly, 23 April 2024; Antisemitism Policy Trust (SMH0005); Molly Rose Foundation (SMH0016); Swansea University Cyber Threats Research Centre (SMH0018); Clean Up The Internet (SMH0023); 5Rights Foundation (SMH0024); Andreu Casas, Georgia Dagher, and Ben O'Loughlin (SMH0030); NSPCC (SMH0032); UK Safer Internet Centre (SMH0044); Logically (SMH0049); Professor Sander van der Linden, Dr Jon Roozenbeek, and Professor Stephan Lewandowsky (SMH0052); Institute for Strategic Dialogue (SMH0062); Foxglove (SMH0066); Oral evidence taken on 25 October 2021, Q155 [Frances Haugen]; Q16 [Marianna Spring]

57 Meta (SMH0037); X (SMH0064); Google (SMH0065); TikTok (SMH0068)

58 Professor Martin Hilbert (SMH0008)

59 Swansea University Cyber Threats Research Centre (SMH0018); Glitch (SMH0028)

60 Meta gave some written detail on how its recommendation algorithms work and pointed us to other Meta pages giving explanations; TikTok provided an infographic but stated that certain details about its trust and safety systems and algorithm cannot be publicly disclosed due to commercial confidentiality and the risk of enabling malicious actors to bypass protections; X (formerly Twitter) made a portion of its source code of its recommendation algorithm (at the time) public in March 2023. Meta (SMH0080); TikTok (SMH0081); X (formerly known as Twitter) (SMH0082); Github, Source code for Twitter's Recommendation Algorithm (accessed June 2025)

61 Swansea University Cyber Threats Research Centre (SMH0018)

62 Professor Martin Hilbert (SMH0008); Swansea University Cyber Threats Research Centre (SMH0018)

63 Q330

### User control over recommendation algorithms and data collection

**24.** The Online Safety Act includes measures to give users more control over what is recommended to them.[64] Platforms told us they provide various ways for users to control algorithmic content recommendation. On X, Meta and TikTok users can block keywords and provide feedback on individual posts to influence future recommendations.[65] Meta and TikTok offer tools that reset the recommendation algorithm.[66] However, privacy and control options on platforms can be difficult to find.[67]

**25.** **CONCLUSION**
Social media and other online platforms have huge power and reach into our lives, with positive and negative impacts. They can democratise knowledge and access to the public sphere, and help to build social connections and global communities. Generative AI provides further opportunities in terms of productivity, creativity and content moderation. For these reasons, it is imperative that we regulate and legislate these technologies based on the principles set out in this report, harnessing the digital world in a way that protects and empowers citizens.

**26.** **CONCLUSION**
Internet users are exposed to large volumes of harmful and misleading content which can deceive, damage mental health, normalise extremist views, undermine democracy, and fuel violence. We are concerned by the evidence that recommendation algorithms—integral to the advertisement- and engagement-driven business models of social media companies—play a role in this. Young people are particularly vulnerable to these harms, and those born today will never have known a world without AI—we plan to explore in detail the impact the online world has on their developing brains in our future work.

**27.** **CONCLUSION**
The technology used by social media companies should be transparent, explainable and accessible to public authorities, as stated in our Principle 5. This is currently not the case: when we asked, major platforms did not give us detailed, transparent up-to-date representations of their recommendation algorithms.

64 Ss 15–16, Online Safety Act 2023
65 Meta (SMH0037); X (SMH0064); TikTok (SMH0068)
66 Meta (SMH0037); TikTok (SMH0068)
67 European Data Protection Board, Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognise and avoid them, 24 February 2023, pp 65–66. The UK GDPR offers individuals the right to erase personal data following a verbal or written request. Article 17, UK GDPR; ICO, Right to erasure (accessed June 2025)

28. **CONCLUSION**

Social media companies have often argued that they are not publishers but platforms, abdicating responsibility for the content they put online. We believe that these services, with sophisticated recommendation algorithms that directly amplify and push content to users, are not merely platforms but curators of content. As we have seen, the amplification and spread of this content can have serious, large-scale impacts.
We recognise that this is a complex area of law and that defining social media companies as publishers would have major consequences, but the current situation is deeply unsatisfactory. We call on the government to set out its position on this question in its response to this report.

29. **RECOMMENDATION**

There is a shortfall in data needed to accurately analyse the scale of the problem and identify policy solutions. In line with our Principle 4, the government should commission a large-scale research project into how far social media recommendation systems spread, amplify or prioritise harmful content. This should be undertaken by a group of credible independent researchers, bringing diverse perspectives, with full access to the inner functions of the systems that major platforms use to algorithmically recommend content, including the private, external, and third party data used to train their systems; the user, content and engagement attributes the algorithms rely on and how these are weighted, and the objectives the algorithms are optimised for; where user interactions reinforce future recommendations; and any curation rules or interventions that influence promotion or suppression of content. We expect full cooperation from all major services that employ recommendation algorithms.

30. **RECOMMENDATION**

Based on the research described above, the government should publish conclusions on the level and nature of harm that these platforms promote through their recommendation systems. Following our Principle 3, if significant harm is found, the responsible online services should publish the actions they will take to address these harms. Ofcom should be given the power to serve penalty notices to services that fail to comply, either 10% of the company's worldwide revenue, or £18 million, whichever is higher.

**31.** **RECOMMENDATION**
Following our Principles 2 and 3, the government should compel social media platforms to embed tools within their systems that identify and algorithmically deprioritise fact-checked misleading content, or content that cites unreliable sources, where it has the potential to cause significant harm. It is vital that these measures do not censor legal free expression, but apply justified and proportionate restrictions to the spread of information to protect national security, public safety or health, or prevent disorder or crime.

**32.** **RECOMMENDATION**
As per Principle 4, users should have more control over the content that is pushed to them online. Government should mandate all online services with a content recommendation algorithm to give the user a 'right to reset', which would delete all data stored by their recommendation algorithm, in the manner that users can clear their cookie history. This option should be displayed prominently on the platform's main feed or homepage.

## How platforms moderate misleading and harmful content

### Harmful content

**33.** The major platforms use AI-based moderation systems to automatically flag and remove content that clearly violates their terms of service, and use human moderators where the question is more complicated. Meta told us its automated systems removed 99% of 8.2 million pieces of terrorist content before it was reported to them, TikTok that 98.2% of 140 million pieces of violative content were removed through automated methods before being reported by a user, and YouTube that roughly 96% of 8.6 million removed videos were first detected by an automated classifier.[68]

**34.** However, evidence to this inquiry accused the platforms of "marking their own homework" in this area.[69] The Center for Countering Digital Hate argued that, as the internal data of platforms is secret, there is a lack of true transparency, oversight or accountability over their moderation practices:

68 Meta (SMH0037); Google (SMH0065); TikTok (SMH0068); TikTok (SMH0081); Google Transparency Report, YouTube Community Guidelines enforcement (accessed June 2025); Meta Transparency Center, Dangerous Organizations: Terrorism and Organized Hate (accessed June 2025)

69 Q17 [Imran Ahmed]; Q235 [Dr Abdul Rahman]; Q309 [Tabitha Rowland]

> [Social media companies can] gaslight you by just saying, "Look, we take down 99%, guys. We're great." [ … ] we have the serious problem that they mark their own homework; they provide their own data. We have had no transparency. Meaningful accountability requires that transparency [ … ].[70]

Some major social media platforms have weakened their content moderation policies in recent years. Elon Musk acquired Twitter in 2022, renamed it 'X', and implemented changes such as disbanding its Trust and Safety Council.[71] We heard that there had been a fall in effective content moderation since the takeover, with action taken against 1 million accounts for posting hateful material in 2021, which fell to only 2,361 in 2024, out of 66.9 million reports.[72] Despite these developments, X's representative told us he thinks that X is a safer online environment since Musk's takeover.[73]

**35.** In January 2025, Meta announced that it would lift restrictions on its platforms in the US around topics such as immigration, gender and politics.[74] Reported leaked internal documents gave examples of posts that would now be allowed, including statements such as "immigrants are grubby, filthy pieces of shit."[75] When we asked Meta's representative about these reports, he appeared to confirm their accuracy, stating some

70 Qq12, 17 [Imran Ahmed]

71 Other changes include: replacing identity verification with a paid system that offers algorithmic prioritisation, financial rewards for high engagement to 'Premium' users, reinstating previously banned users, allowing blocked accounts to continue viewing posts of accounts that blocked them. [former Twitter Head of Trust and Safety, Yoel Roth] stated that Musk's laissez-faire approach to content moderation, and his lack of a transparent process for making and enforcing platform policies, made Twitter less safe. It is reported that hate speech on X rose by 50–273% in the months following Musk's takeover. The New York times, Elon Musk Completes $44 Billion Deal to Own Twitter, October 27 2022; NBC News, Twitter rebrands to 'X' as Elon Musk loses iconic bird logo, July 24 2023; Dr Áine MacDermott(SMH0010); Sky News, Elon Musk's Twitter dissolves Trust and Safety Council, 13 December 2022; DMR News, Creators on X Will Now Earn Based on Engagement, Not Ads, 11 October 2024; Associated Press, You may have blocked someone on X but now they can see your public posts anyway, November 2024; CNN Business, Twitter is less safe due to Elon Musk's management style, says former top official, November 30, 2022; Euro News, Hate speech on X now 50% higher under Elon Musk's leadership, new study finds, February 13 2025; Center for Countering Digital Hate, The Musk Bump: Quantifying the rise in hate speech under Elon Musk, 2 December 2022

72 Q234 [Dr Abdul Rahman]

73 Q175 [Wifredo Fernandez]

74 Meta, More Speech and Fewer Mistakes, 7 January, 2025

75 Wired, Meta Now Lets Users Say Gay and Trans People Have 'Mental Illness', 7 January 2025

conversations, "while challenging, should have a space to be discussed."[76] Meta's Oversight Board—an independent body that was set up by Meta and reviews the company's decisions—expressed concerns over these changes.[77]

### Misleading content

**36.** Major platforms differ in measures to address misleading content where the veracity or harmfulness is not immediately clear. Meta, TikTok and Google emphasised their collaboration with independent third-party fact-checking (3PFC) organisations.[78] In contrast, X uses 'Community Notes', a crowd-sourced and volunteer-led fact-checking system, where users contribute suggested corrections or context to be appended to posts.[79] X told us this system uses a 'bridging algorithm': for a Note to be shown, it needs to have been found helpful by people who have tended to disagree in their past ratings, which can help to "identify content that is healthier and higher quality, and reduce the risk of elevating polarizing content."[80]

### Moves to weaken moderation policies

**37.** Major platforms have also announced measures weakening their policies on harmful content. In December, 2024, Meta told us that its 3PFC programme was a "key part" of its approach to combating misinformation.[81] However, the following month it announced it would end its relationship with fact-checkers in the US in favour of a 'Community Notes' model, on the grounds

76 Meta later stated that leaked documents are "not a reliable source of information about content policy". Q133 [Chris Yiu]; Meta (SMH0080)

77 Meta Oversight Board: "On the broader policy and enforcement changes hastily announced by Meta in January, the Board is concerned that Meta has not publicly shared what, if any, prior human rights due diligence it performed in line with its commitments under the UN Guiding Principles on Business and Human Rights. It is vital Meta ensures any adverse impacts on human rights globally are identified and prevented." Meta Oversight Board, Posts supporting UK riots, 23 April 2025

78 These are usually certified through the International Fact-Checking Network and follow its Code of Principles. Meta (SMH0037); Google (SMH0065); TikTok (SMH0068)

79 Contributors must have no recent policy violations and a verified phone number. When a Community Note is posted, neither X nor the original poster have power to remove it. There are currently 1 million contributors globally, with over 70,000 in the UK. X (SMH0064)

80 X (formerly known as Twitter) (SMH0082); In 2022, the Harvard Kennedy School recommended that to minimise polarising and harmful content, social media platforms should adopt bridge algorithms to rank contents on mutual understanding and trust across divides rather than on engagement. Harvard Kennedy School, Bridging-Based Ranking, May 17 2022

81 Meta (SMH0037); This was criticised by the Director of the International Fact-Checking Network, prominent fact-checking organisations, and other disinformation experts. Linkedin, Angie Drobnic Holan, January 2025; Full Fact (SMH0070); BBC News, Meta is ditching fact checkers for X-style community notes. Will they work?, 26 January 2025; France 24, Disinformation experts slam Meta decision to end US fact-checking, 8 January 2025

that fact-checkers could become a "tool to censor."[82] Meta told us it has no plans to end 3PFC or introduce Community Notes in the UK "at this stage", and that it would carefully consider its obligations, including under the Online Safety Act, before making changes.[83]

**38.** 3PFC and Community Notes have advantages and drawbacks. 3PFC has been found to stop users interacting with misinformation, but has been criticised for inconsistency and slow response times.[84] It is also far more expensive. Community Notes has been praised for scalability and trustworthiness, but criticised for a reliance on consensus over factual accuracy, and a lack of specialist nuance.[85]

82 Meta, More Speech and Fewer Mistakes, January 7, 2025

83 Meta (SMH0080)

84 Lyric Jain: "If we look at the work that 120-odd accredited fact-checking organisations around the world have done as part of Meta's programme or TikTok's programme—full disclosure: we happen to be part of those programmes—it has been work that has been powerful in helping to downrank certain types of harmful content." Internal Meta data from the 2023 EU parliamentary elections showed that when content was labelled false, 95% of people did not view it. 2021 research on fact-checking in Argentina, Nigeria, South Africa and the UK found that fact-checks reduced false beliefs in all countries. However, fact-checks have been criticised for slow response times, inconsistency across fact-checking organisations, and a lack of reach in comparison to the initial false claim. Qq237–39 [Lyric Jain]; Jeffrey Howard, and Maxime Lepoutre (SMH0013); Full Fact (SMH0070); Meta, How Meta Is Preparing for the EU's 2024 Parliament Elections, 25 February 2024; Free Speech Union (SMH0059); Harvard Kennedy School, "Fact-checking" fact checkers: A data-driven approach, 26 October 2023

85 Lyric Jain: "There are some really good aspects about community notes: more people can participate through crowdsourcing; people with greater local context or subject matter expertise may be able to contribute to assessments; and it is certainly a lot more economical." X and Meta both cite evidence that posts with Community Notes are shared 50–61% less, and deleted 80% more than posts without. Full Fact: "An additional problem with the Community Notes model, as used on X, is that it relies on establishing consensus rather than emphasising factual accuracy. The need to find consensus means that most proposed Notes are never published." Dr Abdul Rahman: "This is the problem with community notes. […] If you look at disinformation, it is always inflected with humour, and that is where community notes on X fail, because it becomes a reductive binary approach of yes or no, when humour can be weaponised, and we all know that." Qq237, 243 [Lyric Jain; Dr Abdul Rahman]; X (SMH0064); Meta (SMH0080); Full Fact (SMH0070); Logically (SMH0049); FactCheckHub, Navigating information space: experts debate the role of community notes in fact-checking, 5 April 2025

39. **CONCLUSION**
The UK government—like its counterparts around the world—is facing the challenge of attempting to regulate hugely powerful technology companies that operate across the world, providing technologies that transform societies, with bigger budgets than many countries. It is essential that their impact on our society be understood, effectively scrutinised and, where necessary, regulated in the public interest by Parliament. The committee has experienced some of the challenges of this in its engagement with these companies. We were reassured by the statements from Google, Meta, TikTok and X in our evidence session that they accepted their responsibility to be accountable to Parliament, and we hope that this will be put into practice.

40. **CONCLUSION**
We are concerned by disjointed approaches from platforms to false and harmful content; in particular by recent moves from X and Meta to water down their Terms of Service and approach to content moderation. While there are merits to crowd-sourcing models of context provision and fact-checking—as part of a wider policy on misleading and harmful content—these platforms seem to be prioritising this method over third-party fact checking without clear evidence on whether it will adequately protect users from algorithmically amplified harm and misinformation.

41. **RECOMMENDATION**
In line with our Principle 1 of tackling amplified misinformation, the government should compel platforms to put in place minimum standards for addressing the spread of misleading content online. More information is needed on the merits of different approaches to this. The government should commission research into the relative benefits of independent third-party fact-checkers, crowd-sourced context provision, and AI driven detection of misinformation, using researchers who are independent, bring diverse perspectives on the issue, and have full access to the data of these systems. The research should enable Ofcom to offer guidance on the most effective method, or combination of methods, to address misinformation.

# The Online Safety Act and the spread of misleading and harmful content

## Harmful and misleading content

**42.** The Online Safety Act focuses on tackling content that is illegal, or harmful to children. The Act and Ofcom's Codes of Practice do not designate "misinformation" as a harm that platforms need to address, except when it falls into these categories.[86] Some limited measures relevant to misinformation were included in the Act:

- **Introduction of 'False Communications' offence.** This makes it an offence to send a message that "conveys information that the person knows to be false", intended to cause "non-trivial psychological or physical harm" to a "likely audience."[87] It has been criticised for being vaguely worded with an overly wide margin for interpretation, which could fail to protect both user safety and free expression.[88] Ofcom declined a request from X to provide clear examples of this offence, on the grounds that this would lead to an "insufficiently nuanced" approach from platforms.[89] Ofcom told us the offence "will not be easy for a company to identify", and confirmed that it would not fall within the remit of the Online Information Advisory Committee (see below).[90]

- **Establishment of "Advisory Committee on Disinformation and Misinformation".**[91] This was formed in April 2025, with a new name: "The Online Information Advisory Committee".[92] Full Fact criticised the name as "bland, vague and disappointing", noting changes to the terms of reference that de-emphasise mis/disinformation.[93] Ofcom

86 Online Safety Act 2023; Ofcom, Statement: Protecting people from illegal harms online, 16 December 2024; Ofcom, Statement: Protecting children from harms online, 24 April 2025

87 S 179, Online Safety Act 2023

88 Big Brother Watch (SMH0043); Logically (SMH0049); Dr Elena Abrusci (SMH0050); The Free Speech Union (SMH0059)

89 Ofcom, Protecting People from illegal harms online, December 2024, p 71

90 Q282 [Mark Bunting]; Ofcom (SMH0085)

91 S 152, Online Safety Act 2023

92 Ofcom, Ofcom establishes Online Information Advisory Committee, 28 April 2025

93 The November 2024 Terms of Reference in its "functions and duties" section quotes Section 51 in full, leading to nine mentions of disinformation and misinformation, whereas the April 2025 version just refers readers to that section, resulting in one explicit mention. Ofcom, Advisory Committee on Disinformation and Misinformation, 2 December 2024, archived at https://web.archive.org/web/20241202203336/https://www.ofcom.org.uk/siteassets/resources/documents/about-ofcom/how-ofcom-is-run/mis-and-dis-information-committee/advisory-committee-on-disinformation-and-misinformation-

told us the previous name had not described the "totality" of the committee's role, but that mis/disinformation would still be its main focus.[94]

- **Publication of 'Media Literacy Strategy'.** The Act tasked Ofcom with publishing a strategy to raise public awareness on online safety, which it did in October 2024.[95] We have heard evidence about the importance of media literacy in helping the public identify and avoid being influenced by mis/disinformation.[96] However, the strategy's indicators of success seem limited.[97]

**43.** We received evidence calling for the OSA to follow the model of the EU's Digital Services Act (DSA) in addressing misinformation. Under the DSA, large social media and search platforms must address "systemic risks" that extend beyond illegal content, which could include misinformation, and must consider whether the design of their recommendation algorithms influences systemic risks.[98]

**44.** The Act does not mandate standards for platforms to enforce with regard to misinformation. Platforms are considered compliant if they meet Ofcom's illegal harms or children's harms Codes of Practice, which do not designate misinformation as a specific harm to be assessed or met. Ofcom's risk assessment and safety-by-design provisions do not apply to misinformation.[99]

---

terms-of-reference.pdf?v=386330 (accessed June 2025); Ofcom, Online Information Advisory Committee, 28 April 2025; Full Fact, Disinformation and misinformation must remain the primary focus of Ofcom's Committee, 28 April 2025

94 Q277 [Mark Bunting]; Oral evidence taken on 20 May 2025, Q30 [Melanie Dawes]

95 S 166, Online Safety Act 2023; Objectives and priorities included promoting protective technologies, signposting helpful resources, focusing on vulnerable groups, and including a media literacy statement in Ofcom's annual report. Ofcom's evidence states that each of the three sections of the strategy – Research, Evidence and Evaluation, Engaging Platforms and People and Partnerships – has an explicit focus on understanding what works in supporting people to identify and build resilience to misinformation and disinformation. Ofcom, Ofcom's three-year media literacy strategy (accessed June 2025); Ofcom (SMH0078)

96 The charity Glitch have argued that Ofcom are underfunded in this regard. Dr Aine MacDermott (SMH0010); Faculty of Public Health (SMH0011); Dr Hossein Dabbagh (PhD) (SMH0017); Glitch (SMH0028); Atlantic Council's Democracy + Tech Initiative (SMH0034); Meta (SMH0037); Dr Kumar (SMH0040); Prof van der Linden (SMH0052); Internet User Behaviour Lab (SMH0058). The House of Lords Communications and Digital Committee have recently opened an inquiry into Media literacy. House of Lords Communications and Digital Committee, Media literacy Inquiry (accessed June 2025)

97 Ofcom, Ofcom's three-year media literacy strategy (accessed June 2025)

98 EU, Article 34, Digital Services Act 2022; The Electoral Commission (SMH0021); Global Witness (SMH0048); Beatriz Kira, Zoe Asser, Phoebe Li and Julie Weeds (SMH0056)

99 Ofcom, Statement: Protecting people from illegal harms online, 16 December 2024; Ofcom, Statement: Protecting children from harms online, 24 April 2025

**45.** **CONCLUSION**
The Online Safety Act will lead to some improvements, but is designed only to protect users from harm that is illegal or affects children. The decision not to include measures related to the algorithmic amplification of "legal but harmful" content, such as misinformation, means that full enforcement of the Act would have made little difference to the online environment that helped to incite the violence of last summer.

**46.** **CONCLUSION**
It is vital that platforms are held responsible for the algorithmic spread of misleading or deceptive content that can radicalise and harm users. The few measures in the Act that address misinformation fall short. The False Communications offence is vaguely worded and will be difficult to implement; the advisory committee on mis/disinformation has been renamed, suggesting a change of emphasis; and the media literacy strategy does not set ambitious goals.

**47.** **RECOMMENDATION**
The broad scale—and serious impact—of misinformation online requires greater transparency and accountability from the government. In line with our Principle 1, the government should submit an annual report to Parliament on the state of misinformation online, tracking trends and issues from the year, and setting out successes and failures in addressing them.

**48.** **RECOMMENDATION**
In line with Principle 5, transparency, the government should introduce duties for platforms to undertake risk assessments and reporting requirements on legal but harmful content, such as potentially harmful misinformation, with a focus on the role of recommendation algorithms in its spread.

**49.** **RECOMMENDATION**
To ensure true responsibility from platform companies, as per Principle 3, Ofcom and DSIT should confirm that services are required to act on all risks identified in risk assessments, regardless of whether they are included in Ofcom's Codes of Practice.

## Safe by design

**50.** The Online Safety Act states that regulated services must be "safe by design."[100] It includes some measures targeting the recommendation algorithms of social media platforms, focused on control of content that is illegal or harmful to children.[101] The Act also commissions Ofcom to publish Codes of Practice on Illegal Harms and on Protecting Children. These contain some measures on algorithms, such as compelling platforms to analyse the risk that design adjustments to content recommender systems could lead to the recommendation of illegal content; and requiring user-to-user services to "configure their algorithms to filter out or reduce the prominence of harmful content in children's feeds."[102]

**51.** Stakeholders have criticised the Act for a focus on reactive content moderation measures as opposed to proactive design measures that could create true safety by design in platforms.[103] These could include addressing the features of platforms that allow the algorithmic amplification of misleading and harmful content, through independent audits of platforms' recommendation algorithms and the imposition of safety duties, as well as demotion or de-amplification to reduce the reach of content that is legal but potentially harmful.[104] Independent researchers could be given greater access to the internal data of social media recommendation algorithms in order to build a systemic understanding of emerging developments and the spread of false and harmful content.[105] We heard that regulating at content

100 S 1, Online Safety Act 2023

101 Ss 9—12, 14—15, 77—79, Online Safety Act 2023

102 Ofcom, Quick guide to Protection of Children Codes, 24 April 2025; Ofcom, Statement: Protecting people from illegal harms online, 16 December 2024; Ofcom, Illegal content Codes of Practice for user-to-user services, 24 February 2025, p 42; Ofcom, Protecting children from harms online Volume 4: What should services do to mitigate the risks of online harms to children?, pp 6, 14; Ofcom, Statement: Protecting children from harms online, 24 April 2025

103 Office of the Children's Commissioner for England (SMH0014)

104 Antisemitism Policy Trust (SMH0005); Professor Jeffrey Howard, Dr Maxime Lepoutre (SMH0013); Global Witness (SMH0048); Minderoo Centre for Technology and Democracy (SMH0051); Beatriz Kira, Zoe Asser, Phoebe Li and Julie Weeds (SMH0056); Institute for Strategic Dialogue (SMH0062)

105 Center for Countering Digital Hate (SMH0009); Swansea University Cyber Threats Research Centre (SMH0018); Minderoo Centre for Technology and Democracy (SMH0051); Institute for Strategic Dialogue (SMH0062); Ofcom consulted on researcher access to information from regulated online services under the Online Safety Act from October 2024—January 2025. Clause 125 of the Data (Use and Access) Act includes the following amendment to the Online Safety Act, inserting the new Section 154A "Information for research about online safety matters": "The Secretary of State may by regulations require providers of regulated services to provide information for purposes related to the carrying out of independent research into online safety matters." Ofcom, Call for evidence: Researchers' access to information from regulated online services, 28 October 2024 (accessed June 2025); Data (Use and Access) Act 2025; S 154A, Online Safety Act 2023

level puts the responsibility for harm on the individual posters or sharers, whereas regulating at a systematic design level would place responsibility on the platforms that approve, host and algorithmically recommend and spread harmful content.[106]

## 'Small but risky' platforms

**52.** The Online Safety Act requires Ofcom to place different requirements on user-to-user and search services, depending on their number of users.[107] The previous government had amended the bill to categorise services based on number of users or level of risk, but, following Ofcom advice, it reversed this and categorised services based on number of users alone.[108] Platforms with a smaller userbase therefore do not face the same transparency or safety requirements.[109]

**53.** We received evidence that smaller platforms can still present a high level of risk, spreading misleading or harmful content that is then amplified on bigger platforms.[110] This was demonstrated in the summer unrest[111]—TikTok told us that much of the misinformation and riot coordination content on its platform was created elsewhere before spreading to TikTok.[112] Targeted disinformation campaigns and influence operations are often developed on fringe sites, before being spread through major sites.[113]

106 Dr Abdul Rahman: "It is really interesting because a lot of the oral proceedings that I have gone through in January and February focus on content. That really protects the companies under section 230 of the US Communications Decency Act [...]" Q225

107 Schedule 11, Online Safety Act 2023

108 Baroness Morgan of Cotes' amendment, Schedule 11, Online Safety Act 2023; Ofcom, Categorisation Advice submitted to the Secretary of State, 25 March 2024; Online Safety Act Implementation HCWS312, 16 December 2024. The categories are: Category 1: User-to-user services with content recommender systems that either: have over 34 million UK users; or have over 7 million UK users and allow resharing of user-generated content. Category 2A: General search services with over 7 million UK users. Category 2B: Applies to services allowing direct messaging with over 3 million UK users. The Online Safety Act 2023 (Category 1, Category 2A and Category 2B Threshold Conditions) Regulations 2025, SI 2025/226

109 Office of the Children's Commissioner for England (SMH0014); Dr Beatriz Kira, Professor Julie Weeds, Professor Phoebe Li, and Zoe Asser (SMH0056)

110 Harms hosted on small platforms include hosting suicide ideation and spreading misleading, harmful or hateful content that is then amplified on bigger platforms. Children's Commissioner for England's Office (SMH0014); Dr Beatriz Kira, Professor Julie Weeds, Professor Phoebe Li, and Zoe Asser (SMH0056); Mental Health Foundation, Our joint letter to Sir Keir Starmer about the Online Safety Act, 15 October 2024

111 Institute for Strategic Dialogue (SMH0062)

112 TikTok (SMH0068); Q93 [Ali Law]

113 Clean Up the Internet (SMH0024); Marc Owen Jones (SMH0071)

**54.** Ofcom has set up a 'Small but Risky supervision task force' to identify and manage these services, and take action against them in the case of non-compliance.[114] It began work in summer 2024, looking at "low reach" services with under or around 1% of the UK population as active monthly users, that have "high risk features of functionality."[115] The taskforce has engaged with more than 25 high risk services, including fringe sites such as Gab, Kiwifarms, Bitchute and a suicide forum, and has "taken steps" to prevent UK users accessing the latter site."[116]

**55.** **RECOMMENDATION**

The Online Safety Act does not do enough to address the risks posed by small platforms due to its exclusive focus on size. Ofcom should create an additional category to cover 'small but risky' platforms, based on analysis of the role that harmful smaller platforms can play in the online ecosystem, interacting with the recommendation algorithms of large platforms to spread harms such as misinformation, and disinformation campaigns. This regulation of small platforms should be in line with our Principles 1, 3 and 5.

# Disinformation campaigns

## Foreign interference

**56.** Evidence to this inquiry stated that foreign influence operations may have played a role in last summer's unrest, while the head of counter terrorism policing stated that foreign bots had "turbo-charged" the spread of misinformation.[117] Some state actors, such as Russia and China, invest heavily in online information campaigns and influence operations, disseminating false or polarising content to widen social divides and influence political behaviour.[118] Technology such as bots is used to amplify messages through social media recommendation algorithms.[119] We heard that the website Channel3Now, which published Southport misinformation, "resembles Russian approaches around information laundering and narrative dissemination."[120]

114 HL Deb, 16 January 2025, col 1259

115 Ofcom, Enforcing the Online Safety Act: Platforms must start tackling illegal material from today, 17 March 2025

116 Ofcom (SMH0086)

117 Logically (SMH0049); Marc Owen Jones (SMH0071); The Northern Echo, Southport misinformation 'turbo charged' by foreign bots online, 20 November 2024

118 Andreu Casas, Georgia Dagher, and Ben O'Loughlin (SMH0030); Logically (SMH0049)

119 Clean Up The Internet (SMH0023)

120 Logically (SMH0049); For more information on 'Channel3Now', see Chapter 4, *Digital advertising market, Digital advertising and harm, Digital advertising and the 2024 unrest*

**57.** The Online Safety Act gives Ofcom responsibility for guiding platforms in managing and removing foreign interference.[121] We heard Ofcom had failed to give clear guidance to platforms on these offences.[122] Responsibility for monitoring mis/disinformation online, especially that resulting from foreign interference, appears to be spread across multiple departments: the Home Office, Cabinet Office, DSIT and Foreign Office. The Intelligence and Security Committee's (ISC) 2020 "Russia Report" stated that defence against Russian disinformation became "something of a 'hot potato'" with no organisation considering itself to be in the lead.[123] However, DSIT told us that "there were obviously cross-government structures in place [to monitor mis/disinformation online] during Southport."[124]

**58.** **RECOMMENDATION**
Foreign interference and disinformation campaigns, with use of technology such as bots and AI, put UK citizens at risk. The possibility that some of the divisive messages and deceptive content spread by users—and amplified by algorithms—last summer were part of such an influence operation is deeply concerning. In order to tackle amplified disinformation, identified by Principle 1, the government and Ofcom should collaborate with platforms to identify and track disinformation actors and the techniques and behaviours they use to spread adversarial and deceptive narratives online.

**59.** **RECOMMENDATION**
Responsibility for tracking foreign disinformation campaigns appears to be split between several departments, including DSIT. This suggests that the Intelligence and Security Committee's 2020 characterisation of countering Russian influence operations as a "hot potato", passed between different bodies, has not been addressed. To meet our Principle 1, the government should clarify which department has ownership over tracking and countering online narrative operations. It should consider consolidating responsibility within a single entity, for example the National Security Online Information Team, or establishing a clear chain of command, and in its response to this report the government should set out the actions it intends to take in this regard.

121 'Foreign intervention offences' as defined by Ss 13—16 National Security Act 2023; Schedule 7, Online Safety Act 2023; Ss 13—16
122 Logically (SMH0049)
123 Intelligence and Security Committee of Parliament, Russia, HC 632, 21 July 2020, paras 31—35
124 Q313 [Talitha Rowland]

## National Security Online Information Team

**60.** The government established the Counter Disinformation Unit (CDU) in 2019 to identify and counter false information, particularly that spread by foreign states.[125] In 2023 it was renamed the National Security Online Information Team (NSOIT). The team conducts targeted open-source monitoring and analysis to identify and assess potential narrative threats, and engages with platforms on measures to counter mis/disinformation.[126] Major platforms told us that they engaged with NSOIT during last year's summer unrest.[127]

**61.** We heard NSOIT plays a crucial role in addressing the spread of harmful and false content, but that there have been concerns over its monitoring of lawful political speech, and a lack of oversight.[128] The ISC and the Culture, Media and Sport Committee (CMS) have both raised concerns about oversight of the CDU/NSOIT, and the ISC called for the team to be brought within its remit.[129] The government rejected the CMS Committee's recommendation to lay an independent review of the team's activities before Parliament.[130] When we put these concerns to Baroness Jones, she said that ministers were answerable for the work of the NSOIT.[131]

**62.** **RECOMMENDATION**

The NSOIT is an important tool in protecting citizens from disinformation and needs appropriate scrutiny. Government should place NSOIT on a statutory footing and bring it under the remit of the Intelligence and Security Committee, to ensure that our Principle 1 is being effectively and safely pursued, in line with Principle 2.

125 Cabinet Office and Department for Science, Innovation and Technology, Fact Sheet on the CDU and RRU, 9 June 2023

126 UK Government (SMH0061)

127 Meta (SMH0037); X (SMH0064); Google (SMH0065); TikTok (SMH0068)

128 Glitch (SMH0028); Big Brother Watch (SMH0043); The Free Speech Union (SMH0059)

129 In December 2022, the ISC had complained of an "erosion of oversight" and claimed that the Government was "refusing" to expand the ISC's remit to include the CDU. In 2024 the former Culture, Media and Sport Committee stated concern over the "lack of transparency and accountability of the CDU and the "appropriateness of its reach", recommending that Government lay an independent review of its activities before Parliament. Government rejected this recommendation, citing the change from CDU to NSOIT as a change in remit. Intelligence and Security Committee of Parliament, Annual Report 2021—2022, HC 922, para 33, Annex E, para 8; Culture, Media and Sport Committee, Sixth Report of Session 2023–24, Trusted voices, HC 175, para 46

130 Culture, Media and Sport Committee, Second Special Report of Session 2024–25, Trusted voices: Government response, HC 292, p 7

131 Q341

# 3 Generative AI

## Harms from generative AI

**63.** Companies such as Google, Meta, X and Microsoft have all integrated generative AI, including in the form of LLMs and chatbots, into their services. Reportedly, ChatGPT is now the 5th most visited website in the world.[132] Generative AI creates new risks in terms of spreading misleading content online, particularly given its low cost and accessibility.[133]

### Inadvertent creation of harmful and misleading content

**64.** We received evidence warning that AI "hallucinations"—where the model produces false information—have a major impact on information integrity online.[134] Estimations of the frequency of hallucinations in generative AI outputs across different engines for different tasks range from 0.7% to 79%.[135] Google's 'AI Overview', introduced in May 2024 to summarise search results, has produced false information, such as inaccurate IQ numbers

132 Similarweb, Top Websites Ranking (accessed June 2025); Information is correct as of June 2025

133 The Alan Turing Institute (CETaS) (SMH0007); Molly Rose Foundation (SMH0016); 5Rights Foundation (SMH0024); Logically (SMH0049)

134 Examples of hallucinations include BBC finding "significant issues" in AI summaries of news stories, and the finding of hateful or false content in 78–80% of 100 prompts related to sensitive topics (narratives surrounding climate change, vaccines, Covid-19, anti-LGBTQ+ hate, sexism, antisemitism, racism, Ukraine and school shootings). Center for Countering Digital Hate (SMH0009); Global Witness (SMH0048); Professor Nishanth Sastry, Professor Alice Hutchings, Dr Diptesh Kanojia and Professor Gareth Tyson (SMH0055); BBC News, AI chatbots unable to accurately summarise news, 11 February 2025; NewsGuard, The Next Great Misinformation Superspreader: How ChatGPT Could Spread Toxic Misinformation At Unprecedented Scale, January 2023

135 Missioncloud, Who's the Most Delusional? The AI Hallucination Leaderboard Is Here, 20 February 2025; Visual Capitalist, Ranked: AI Models With the Lowest Hallucination Rates, 10 January 2025; Forbes, Why AI 'Hallucinations' Are Worse Than Ever, 9 May 2025

for different nationalities and ethnicities, reflecting racist ideas.[136] Google's representative told us that this violated Google's policies, and that AI Overviews were a new and "experimental" feature.[137]

## Deliberate creation of harmful and misleading content

**65.** Generative AI has made it much easier and cheaper to create realistic content that is hateful, harmful or deceptive.[138] This can have damaging effects, such as creating fake news reports, threatening political integrity, reinforcing prejudices, as well as harassing and scamming individuals.[139] Given the viral spread of misinformation in summer 2024, and the low cost, sophistication and popularity of generative AI, we are concerned about the impact it could have in future, similar crises.

**66.** Generative AI can cheaply and easily be used in online influence operations.[140] AI-powered sentiment analysis can analyse user sentiment, mimic social media profiles, and tailor narratives to specific audiences.[141] We heard such a system can be built for as little as $400, and a "very basic script" could use generative AI to produce 1,000 iterations of a false message, draw on engagement data and create a "perpetually improving disinformation machine".[142]

**67.** We heard that AI-generated disinformation is particularly effective when combined with exploitation of engagement-based social media recommendation algorithms.[143] BBC's disinformation and social media correspondent told us deepfakes "were only effective because this kind of

136 Google, Generative AI in Search: Let Google do the searching for you, 14 May 2024; CNBC, Google criticized as AI Overview makes obvious errors, such as saying former President Obama is Muslim, 24 May 2024; Wired, Google, Microsoft, and Perplexity Are Promoting Scientific Racism in Search Results, 24 October 2024

137 Qq43–4 [Amanda Storey]; Media reports from May 2025 suggest that inaccuracies and hallucinations remain a problem with this feature. Tech Radar, Google's AI Overviews are often so confidently wrong that I've lost all trust in them, 17 May 2025; Wired, Google AI Overviews Says It's Still 2024, 29 May 2025; OS/2 Museum, AI Responses May Include Mistakes, 20 May 2025

138 The Alan Turing Institute (CETaS) (SMH0007); Molly Rose Foundation (SMH0016); Hossein Dabbagh (PhD) (SMH0017); 5Rights Foundation (SMH0024); Q24 [Imran Ahmed, Dr Whittaker]

139 Sky News, Deepfake audio of Sir Keir Starmer released on first day of Labour conference, 9 October 2023; Institute for Strategic Dialogue, Misleading and manipulated content goes viral on X in Middle East conflict, 14 April 2024; Antisemitism Policy Trust (SMH0005); Center for Countering Digital Hate (SMH0009); The Electoral Commission (SMH0021); Glitch (SMH0028); Marc Owen Jones (SMH0071); Ofcom (SMH0078); Financial Times, The rise of deepfake scams—and how not to fall for one, 2 May 2025

140 Logically (SMH0049); Free Speech Union (SMH0059)

141 Logically (SMH0049)

142 Logically (SMH0049); Q24 [Imran Ahmed]

143 REPHRAIN (SMH0033); Logically (SMH0049); Marc Owen-Jones (SMH0071)

content was recommended by the algorithms to a widespread audience."[144] The fundamental lack of transparency around generative AI platforms and systems and their internal data makes this difficult to tackle.[145]

## Generative AI and the summer unrest

**68.** We received evidence that AI-generated content bolstered misinformation and harmful messaging during the 2024 unrest.[146] Meta's Oversight Board overturned the decision to keep up two "likely AI-generated" hateful images depicting Muslims in that period.[147] One analyst said that the website Channel3Now's combination of authentic police statements with the false name of the Southport suspect showed the site was likely using generative AI to harvest data from social and traditional media, making it vulnerable to "bad data."[148]

**69.** We heard warnings of the potential for generative AI to exacerbate information crises in the future.[149] Its low cost, wide availability and rapid advances means that large volumes of convincing deceptive content can increasingly be created at scale, eroding public trust, increasing division and reinforcing hate.[150] Bad actors can create convincing 'deepfakes', which could stoke divisions and incite violence.[151]

144 Q24 [Marianna Spring]

145 This includes the complex processes used to generate content, the data and information used for training, how this data is assessed for credibility, and the internal safety or moderation mechanisms to block harmful outputs. Antisemitism Policy Trust (SMH0005); OpenMined Foundation (SMH0046); Global Witness (SMH0048); Q24 [Imran Ahmed]; IBM Watson, Lack of training data transparency risk for AI (accessed June 2025); VKTR, The AI Transparency Gap: What Users Don't Know Can Hurt You, 4 April 2025;

146 Dr Mihaela Popa-Wyatt (SMH0045); Free Speech Union (SMH0059); Marc Owen Jones (SMH0071)

147 Meta's Oversight Board opened an investigation into Meta's decision to leave up three posts, two of which were AI-generated. One depicted a giant man wearing a union jack T-shirt chasing Muslim men, another depicted Muslim men chasing a blonde toddler in a union jack t shirt. Meta determined at the time that neither violated Violence and Incitement or Hate Speech policies. Meta's Oversight Board cited "strong concerns about Meta's ability to accurately moderate hateful and violent imagery" in its Case Decision. Oversight Board, New Cases Involve Posts Shared in Support of the UK Riots, 3 December 2024; Meta Oversight Board, Posts supporting UK riots, 23 April 2025

148 Stephanie Lamy, disinformation strategies analyst; The Guardian, How false online claims about Southport knife attack spread so rapidly, 31 July 2024; Channel3Now is discussed further in Chapter 4, *Digital Advertising, Digital advertising and the summer unrest.*

149 Dr Hossein Dabbagh (PhD) (SMH0017)

150 Dr Hossein Dabbagh (PhD) (SMH0017); Free Speech Union (SMH0059)

151 Antisemitism Policy Trust (SMH0005); Dr Áine MacDermott (SMH0010); The Electoral Commission (SMH0021)

## Generative AI and the Online Safety Act

**70.** The Online Safety Act fails to effectively address the risks of generative AI, as it regulates at a technology and content level, rather than based on principles or outcomes.[152] AI platforms are not explicitly covered by the categories the OSA identifies as high risk.[153] Ofcom told us that if generative AI uses features or functionalities that coincide with these categories they will fall within them, but that chatbots are an example of "areas of technology where the legal position is not entirely clear or it is complex".[154]

**71.** The Act contains no measures to identify AI-generated content, or any specifically relating to "deepfakes".[155] Ofcom stated that the Act would regulate illegal deepfakes, but that not all deepfakes are harmful.[156] Ofcom encouraged tech firms not regulated under the Act, such as "AI Model developers and hosts", to make their models safer, including by embedding watermarks for AI-generated content.[157] Despite this, Baroness Jones told us that "the Online Safety Act does cover generative AI", as it is "designed to be future-facing [and] technology-neutral."[158]

---

152 Jon Pearce, former MP for Weston-super-Mare, called for online safety regimes to regulate based on outcomes, rather than prescriptive methods: "This arms race means that regulators and policymakers should be sceptical of arguments from social media platforms that measures to combat misinformation and disinformation are too difficult. The frontier of what is possible and affordable is advancing rapidly and continuously, so providing regulators and policymakers focus on specifying the outcomes which must be achieved, but leave it to the platforms to use the latest and most efficient or effective technologies to deliver them" Jon Pearce (SMH0002)

153 These categories being: Category 1 (user to user/social media), Category 2A (search), and Category 2B (messaging services). UK Government (SMH0061); Ofcom (SMH0078); The Online Safety Act 2023 (Category 1, Category 2A and Category 2B Threshold Conditions) Regulations 2025, SI 2025/226

154 If a generative AI platform allows users to share content or data with other users, it will be covered as a user-to-user service (Category 1). If it incorporates a search service of more than one website, it will be covered as search service (Category 2A). Similarly, if a generative AI platform has pornographic outputs, it will be covered by the pornography regulations of the Online Safety Act, including "highly effective age assurance." The Online Safety Act 2023 (Category 1, Category 2A and Category 2B Threshold Conditions) Regulations 2025, SI 2025/226; Qq285, 316 [Mark Bunting; Talitha Rowland]; Ofcom, Open letter to UK online service providers regarding Generative AI and chatbots, 8 November 2024

155 Dr Elena Abrusci (SMH0050)

156 Ofcom (SMH0078); Clause 138 of the Data (Use and Access) Act amends the Sexual Offences Act 2003 to cover creating, or requesting the creation of, a "purported intimate image" of an adult; Data (Use and Access) Act 2025

157 Ofcom, A deep dive into deepfakes that demean, defraud and disinform, 23 July 2024

158 Q315

## Addressing harms caused by generative AI

**72.** We received evidence calling for the introduction of a universal system to automatically label synthetically generated content, to help address these risks.[159] Many platforms have features that allow or require users to label content that is AI-generated, and Google told us that it was committed to finding ways to make sure "every image generated" by Google has metadata labelling and embedded watermarking.[160] However, we heard that platforms often fail to label AI-generated content, contravening their own policies.[161]

**73.** We received evidence calling for mandated transparency and data access in generative AI platforms and systems, to allow researchers to view and audit these systems and inform policymakers.[162] The Data (Use and Access) Act amends the Online Safety Act to "require providers of regulated services to provide information [to researchers] for purposes related to the carrying out of independent research into online safety matters". This will only cover generative AI providers and services if they have features that coincide with user-to-user, search, or pornographic services.[163]

**74.** **CONCLUSION**

The Online Safety Act does not protect users from the commodification of synthetic mis/disinformation, or provide effective transparency for the systems that produce them. It fails to address the issue of tech companies rolling out experimental features that can feed false or harmful information to their enormous audiences, further threatening information integrity online. This has damaging effects on all users, undermining the reputation of the companies that introduce them.

159 Antisemitism Policy Trust (SMH0005); The Electoral Commission (SMH0021); Institute for Strategic Dialogue (SMH0062)

160 Meta (SMH0037); Google (SMH0065); TikTok (SMH0068)

161 Institute for Strategic Dialogue (SMH0062)

162 Antisemitism Policy Trust (SMH0005); OpenMined Foundation (SMH0046); Global Witness (SMH0048) Q24 [Imran Ahmed, Dr Whittaker]

163 Ofcom, Call for evidence: Researchers' access to information from regulated online services, 28 October 2024 (accessed June 2025); Data (Use and Access) Act 2025; S 154A, Online Safety Act 2023; Ofcom, Open letter to UK online service providers regarding Generative AI and chatbots, 8 November 2024

75. **CONCLUSION**

We are concerned at what appears to be contradiction and confusion between regulators and government over the capabilities, limitations and principles behind the Online Safety Act. We expect senior Ofcom officials and ministers to be fully aligned in their understanding of the Act—particularly in relation to critical issues such as misinformation and generative AI. Ofcom at times appeared complacent in its approach to public safety online, failing to live up to its role as the UK's online safety regulator and slipping into the role of mediator between the industry and the consumer.

76. **RECOMMENDATION**

To protect citizens from the AI-exacerbated spread of misinformation and harm, the government should pass legislation that covers generative AI platforms, bringing them in line with other online services that pose a high risk of producing or spreading illegal or harmful content. Following the Principles identified by this report, this legislation should require generative AI platforms to: provide risk assessments to Ofcom on the risks associated with different prompts and outputs, including how far they can create or spread illegal, harmful or misleading content; explain to Ofcom how the model curates content, responds to sensitive topics and what guardrails are in place to prevent content that is illegal or harmful to children; implement user safeguards such as feedback, complaints and output flagging; and prevent children from accessing inappropriate or harmful outputs.

77. **RECOMMENDATION**

Principle 5 is crucial for addressing potential harms from generative AI, as there is currently a serious shortfall in transparency and oversight of the platforms and systems that allow users to create AI-generated content. The government should require providers of generative AI services to provide information to those carrying out independent research into online safety. This should include data such as platforms' internal decision-making processes, training datasets, optimisation objectives, safety mechanisms and guardrails on outputs.

78. **RECOMMENDATION**

To effectively tackle amplified misinformation as per Principle 1, the government should work with relevant experts and platforms to develop technology that automatically detects AI-generated media, meeting mis/disinformation at its source. It should mandate all generative AI platforms, and platforms that employ generative AI technologies, to automatically label AI-generated media with metadata and visible watermarks that cannot be removed.

# 4 Digital advertising market

**79.** The algorithmic spread of false and harmful content is closely linked to the digital advertising market, which was estimated at $790 billion worldwide in 2024.[164] UK Stop Ad Funded Crime (UKSAFC), a coalition of advertising organisations and experts, told us that "any serious attempt to tackle misinformation must consider the role of online advertising in incentivising harmful content."[165]

**80.** Digital advertising involves several key players: advertisers (companies promoting products), publishers (websites and apps that host ads), and intermediaries such as ad exchanges and data brokers.[166] When a user visits a website, publishers often collect and share their browsing behaviour and demographic data. This data feeds into a largely automated system known as 'programmatic advertising', which enables real-time bidding for ad space based on the value of an individual ad impression.[167]

**81.** The digital advertising market is dominated by Google, which was reported to own 90% of the market share of the sell side, 40–80% of the buy side, and roughly 50% of the exchange that connects the two.[168] Almost 78% of Alphabet's 2024 overall revenue came from digital advertising, and it holds more than 89% of global search engine traffic.[169] In April 2025, a US district court ruled that Google had "harmed Google's publishing customers, the competitive process, and, ultimately, consumers of information on the open web" by monopolising key digital advertising technologies.[170]

164 UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Dr Karen Middleton (SMH0036); Global Witness (SMH0048); Dr Karen Middleton (SMH0077); DataReportal, Digital 2025: global advertising trends, 5 February 2025

165 UK Stop Ad Funded Crime (UKSAFC) (SMH0004)

166 Competition and Markets Authority, Online platforms and digital advertising, 1 July 2020

167 Dr Karen Middleton (SMH0036)

168 Checkmyads, Google, explained: Here's how it captures the online ad industry, 23 August 2024

169 Alphabet is Google's parent company; Reuters, Explainer: What does ruling on Google's illegal ad tech monopoly mean?, 17 April 2025; Statcounter, Search Engine Market Share Worldwide (accessed June 2025)

170 US Department of Justice Office of Public Affairs, Department of Justice Prevails in Landmark Antitrust Case Against Google, 17 April 2025

# Social media advertising and harm

**82.** Advertising is the lifeblood of most social media companies. Ad revenue makes up approximately 98% of Meta's revenue, 77% for TikTok and 68% for X.[171] Most platforms utilise 'behavioural advertising', where data on the demographic, characteristics, behaviour and interests of users are tracked, allowing advertisers to target specific groups for their products and services.[172] Meta's platforms dominate social media advertising—holding over 63% of total global social media ad spend in 2024.[173]

**83.** We received evidence that social media recommendation algorithms will therefore prioritise engaging content—regardless of its authenticity or safety—to increase time spent on platform and ad views.[174] Dr Karen Middleton, an academic, told us that algorithms are designed to prioritise content that is often sensationalist and emotional, and as a result, can, "by definition", amplify harmful material.[175] We heard that this has an impact across the entire internet, as sites are incentivised to design and promote content that will perform well according to how social media algorithms rank it, to gain traffic and increase advertising revenue.[176]

**84.** However, Meta, X, TikTok and Google told us that harmful content is counter to their business interests, as advertisers do not want to associate themselves with it; and that their algorithms are designed to reduce exposure to harmful content.[177] Ofcom told us that prioritisation of user engagement can "negatively affect a service's revenue in the long run, creating some countervailing financial incentives."[178]

---

171 Meta Investor Relations, Meta Reports Fourth Quarter and Full Year 2024 Results, 29 January 2025; (accessed June 2025); Business of Apps, TikTok Revenue and Usage Statistics (2025), 25 February 2025; Business of Apps, Twitter Revenue and Usage Statistics (2025), 26 February 2025

172 Dr Philip Seargeant (SMH0006); Dr Kimberley Hardcastle (SMH0026); Beatriz Kira, Zoe Asser, Phoebe Li and Julie Weeds (SMH0056)

173 Visual Capitalist, Visualizing the Social Media Giants Dominating Ad Spend, 27 November 2024

174 Faculty of Public Health (SMH0011); Swansea University Cyber Threats Research Centre (SMH0018); Clean Up The Internet (SMH0023); Digital Mental Health Programme at the University of Cambridge (SMH0027); Andreu Casas, Georgia Dagher, and Ben O'Loughlin (SMH0030); Atlantic Council's Democracy + Tech Initiative (SMH0034); Minderoo Centre for Technology and Democracy, University of Cambridge (SMH0051); Qq12, 22 [Marianna Spring, Imran Ahmed, Dr Whittaker]

175 Q180

176 Global Witness (SMH0048)

177 Meta (SMH0037); X (SMH0064); Google (SMH0065); TikTok (SMH0068); Q37 [Amanda Storey]; Qq 90, 142 [Chris Yiu, Wifredo Fernandez, Ali Law]

178 Ofcom (SMH0078)

85. **CONCLUSION**
Advertising is crucial to major social media companies, which depend on recommending engaging content to increase time spent on their platforms and draw attention to adverts. Their recommendation algorithms do not effectively differentiate between harmless and harmful engaging content, which can result in promotion of misleading, damaging, or hateful material. The effects spread through the online ecosystem, helping to incentivise the production and spread of harmful content.

## Digital advertising and harm

86. Evidence to this inquiry argued that the digital advertising supply chain is excessively complex.[179] The Incorporated Society of British Advertisers (ISBA) told us programmatic advertising is "the most opaque segment" in the market and that on average there are 9,000 websites involved in a single campaign.[180] As a result, brands have limited ability to track where their ads and money ultimately go. UKSAFC told us that "people literally do not know who is being paid."[181] The 2024 UN Global Principles for Information Integrity stated:[182]

> The technology sector has designed digital advertising processes to be complex and opaque with minimal human oversight [... This] can lead to advertising budgets inadvertently funding individuals, entities or ideas that advertisers might not have intended to support.[183]

179 UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Incorporated Society of British Advertisers (ISBA) (SMH0075)

180 It is estimated that more than four out of every five pounds invested in digital advertising in the UK are transacted automatically, with the share expected to grow further. Dr Karen Middleton (SMH0036); Q187 [Phil Smith]; Incorporated Society of British Advertisers (ISBA) (SMH0075)

181 UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Q187 [Phil Smith]; Incorporated Society of British Advertisers (ISBA) (SMH0075)

182 The UN Global Principles for Information Integrity included a number of recommendations to improve transparency and accountability in the digital advertising market. This includes: ensuring advertising upholds human rights; making use of industry standards to develop clear policies to minimise risks; working with industry and civil society to share best practice and mitigate risks to information integrity; carrying out thorough audits of advertising campaigns and maintaining detailed records; requiring digital advertising companies to disclose full ad campaign data, including placement and blocking details, for end-to-end validation; and to vet ad exchange supply partners. UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Dr Karen Middleton (SMH0036); United Nations, United Nations Global Principles For Information Integrity: Recommendations for Multi-stakeholder Action, July 2024

183 UN, United Nations Global Principles For Information Integrity, pp 10–11

Dr Middleton told us the measures brands currently employ to prevent ads from appearing near harmful content—such as brand verification technology, proactive vetting of sites, and block lists—are "highly defective."[184]

**87.** We heard calls for the digital advertising market to be regulated similarly to other large-scale financial exchanges, including 'Know Your Customer' (KYC) checks on participants in the programmatic advertising supply chain, like those that exist in finance and legal industries.[185] The ISBA told us that these checks could help to combat fraud and misinformation, but would introduce friction and would need careful evaluation.[186] A relevant real-world example is the UAE's recent development of a homegrown digital advertising exchange with built-in KYC checks.[187]

**88.** We heard that the digital advertising market is "largely unregulated".[188] Advertising is regulated in the UK through the Advertising Standards Authority (ASA), an independent, non-governmental, industry-funded body. Ofcom oversees regulation of video-sharing platforms such as YouTube and TikTok, including their advertising standards.[189] Under the Online Safety Act, Ofcom places duties on certain services to tackle fraudulent adverts.[190]

**89.** We heard that industry self-regulation has been inadequate, and has failed to tackle the monetisation of harmful content.[191] Dr Middleton argued that the ASA's remit should not be limited to advertising content; that authorities such as the Information Commissioner, Ofcom and possibly the Financial Conduct Authority should consider the monetisation of harmful content;

---

184 Dr Middleton described brand verification technology as "poorly implemented." She stated that companies fail to proactively vet sites they monetise, and advertisers lack the time or capacity to scrutinise the complex and opaque supply chains. "Block lists"—tools to avoid ads being displayed near harmful content by blocking specific key words—have been found to harm journalism by blocking words such as "gun" and media for diverse communities by blocking words such as "gay" or "black." Dr Karen Middleton (SMH0036); Dr Karen Middleton (SMH0077). Advertising Week, The Impact of Keyword Blocking: How Advertisers Missed Out During the Paris Olympics (accessed June 2025)

185 UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Qq186, 201 [Dr Middleton]; BBC Future, How big tech's ad systems helped fund child abuse online, 8 February 2025

186 Incorporated Society of British Advertisers (ISBA) (SMH0075)

187 Khaleej Times, UAE gets first homegrown ad exchange platform to combat advertising fraud, 2 June 2025

188 UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Dr Karen Middleton (SMH0077)

189 Part 4, The Audiovisual Media Services Regulations 2020; Ofcom, Video-sharing platform (VSP) regulation (accessed June 2025)

190 Chapter 5, Online Safety Act 2023; A 'fraudulent advert' must be paid-for, not generated by a user, and come under provisions included in the Financial Services and Markets Act 2000, the Fraud Act 2006 or the Financial Services Act 2012

191 Qq186, 209 [Dr Middleton]; Dr Karen Middleton (SMH0077)

and that Ofcom should be able to administer fines for bad practice in the space. She told us that digital advertising should be "within the remit of a regulatory body such as the Advertising Standards Authority."[192]

**90.** Interventions into the digital advertising market have focused on harmful and illegal advertisement content, rather than on market processes that allow the monetisation of harmful content.[193] These have usually been industry-led.[194] For example, the government's 'Online Advertising Taskforce', established in July 2023, focuses on adverts that are illegal or harmful to children, and operates through "industry-led working groups."[195]

192 Qq201, 209

193 ISBA told us that "both the [Online Advertising] Taskforce and the ASA IPP work are focused on the content of advertising rather than its process", and that the ASA's remit "is not enforcement against illegal activity." Q213 [Phil Smith]; Incorporated Society of British Advertisers (ISBA) (SMH0075)

194 In March 2020 then-Department for Digital, Culture, Media and Sport initiated a call for evidence into benefits, harms and regulatory effectiveness in online advertising. This fed into the Online Advertising Programme (OAT) consultation in 2022. Government response concluded that a new and targeted regulatory framework was needed to tackle illegal advertising and protect under-18s. The Online Advertising Taskforce was established in July 2023, operating through industry-led working groups to improve transparency and accountability in the online advertising supply chain and address illegal advertising and protect under-18s. DSIT's evidence stated that "its current remit does not extend to looking at the commercial practices of online services and platforms, or questions of where advertising appears (unless targeting of children is in question)." The ASA launched the Intermediary and Platform Principles (IPP) Pilot from June 2022–23 to promote ASA rules and cooperation. Its final report summarised that the IPPs promoted self-regulation, and that participating companies (Adform, Amazon Ads, Google, Index Exchange, Meta, TikTok, Twitter, Yahoo, Snap Inc and Magnite) would adhere to the Principles relevant to their businesses. ISBA's Phil Smith told us the IPPs remit is different to the focus of our inquiry (how social media business models monetise harmful content). In 2019 the World Federation of Advertisers set up the Global Alliance for Responsible Media (GARM), aiming to define illegal and harmful content and its monetisation. Major brands signed up and underwent compliance checks via transparency reports, and GARM claimed that ads inadvertently supporting harmful and illegal content decreased from 6.1% in 2020 to 1.7% in 2023. Department for Digital, Culture, Media and Sport, Online advertising—call for evidence, 18 March 2020; DCMS, Online Advertising Programme consultation, 25 July 2023; DCMS, Government response to Online Advertising Programme consultation, 25 July 2023; Government, Online Advertising Taskforce (accessed June 2025); Baroness Jones (SMH0084); ASA, Intermediary and Platform Principles (accessed June 2025); ASA, Intermediary and Platform Principles Pilot—Final Report, 5 October 2023, pp 3–4; Q204 [Phil Smith]; WFA, Statement on the Global Alliance for Responsible Media (GARM), 9 August 2024

195 Government, Online Advertising Taskforce (accessed June 2025); Baroness Jones (SMH0084)

The Global Alliance for Responsible Media, an industry-led intervention focused on the monetisation of harmful content, saw some success but ceased to exist following a legal challenge by X owner Elon Musk.[196]

**91.** We heard that most interventions and regulations have been "inadequate" to address harms.[197] Monetisation of harmful content can occur without knowledge or consent of the brands.[198] For example, in February 2025 companies such as Google, Amazon and Microsoft reportedly inadvertently facilitated advertisement-driven monetisation of websites hosting child sexual abuse material.[199]

## Digital advertising and the 2024 unrest

**92.** The proliferation of false and harmful content during the 2024 summer unrest is linked to the advertising-driven business models of social media companies—as set out above, we received evidence suggesting social media companies may have profited from increased engagement at that time.[200]

**93.** There is evidence that the digital advertising industry helped incentivise creation of false and harmful content after the Southport attack. On the day of the attack, a website purporting to be a news channel, Channel3Now, published an article combining the false name of the attacker with snippets

196 GARM was shut down after X-owner Elon Musk initiated a legal challenge, following advertisers pausing or stopping advertising due to a rise in harmful content on the platform. WFA, Statement on the Global Alliance for Responsible Media (GARM), 9 August 2024; WFA, BBC News, Elon Musk sues Unilever and Mars over x 'boycott', 7 August 2024; UC Berkeley News, Study finds persistent spike in hate speech on X, 13 February 2025; Campaign, X's ad revenue continues to fall after Musk takeover (accessed June 2025); Q20 [Imran Ahmed]; Q187 [Phil Smith]; Incorporated Society of British Advertisers (ISBA) (SMH0075)

197 Dr Karen Middleton (SMH0077)

198 ISBA's evidence detailed "Made for Advertising (MFA) sites", created for the singular purpose of buying and selling ad inventory whilst typically using "sensational headlines, clickbait and provocative content to attract visits. 2024 research from Stanford and Carnegie Mellon found that 74.5% of websites known for publishing disinformation were monetised by advertising. UK Stop Ad Funded Crime (UKSAFC) (SMH0004); Dr Karen Middleton (SMH0036); Incorporated Society of British Advertisers (ISBA) (SMH0075); Dr Karen Middleton (SMH0077)

199 BBC Future, How big tech's ad systems helped fund child abuse online, 8 February 2025

200 See Chapter 2, *Misleading and harmful content on social media, Online activity and the 2024 unrest, Advertising and the profit incentive*

of authentic police statements.[201] This boosted the misinformation—within an hour, 270 X accounts had referenced the name, with nearly 3 million impressions.[202]

**94.** An investigation by investigative group CheckMyAds found it was "likely" Google facilitated monetisation of Channel3Now's false information about Southport.[203] Google told us it demonetised the site on 31 July—two days after the misinformation was posted—but did not answer our questions on how much revenue either Google or Channel3Now earned from misinformation during the period.[204] According to CheckMyAds, Google failed to respond to their requests for comment.[205] Google told us that creating a "safe, high quality [advertising] ecosystem is absolutely critical" to its business, and that it has "robust Ads policies" that protect users and keep ads platforms safe.[206]

**95.** **CONCLUSION**
The global digital advertising market is overcomplicated, opaque and under-regulated, operating through an enormous, automated and inaccessible supply chain. This directly leads to the production, viral spread and monetisation of harmful and deceptive content, often without advertisers' knowledge. Platforms and advertisers appear to be either unable or unwilling to address this problem. We heard evidence that platforms may have profited from misinformation and hateful content after the Southport attack.

**96.** **CONCLUSION**
In particular, we were concerned by evidence that Google may have helped to monetise misinformation relating to the attacks, contributing to the violence. This is unacceptable, and is just one example of a much wider problem with the digital advertising industry. We are concerned that Google was seemingly unaware of the chain of events when we asked them about it; failed to tell us how much revenue was earned from this; and failed to reassure us that the company would prevent this from happening again.

201 Logically (SMH0049); Marc Owen-Jones (SMH0071); BBC News, The real story of the news website accused of fuelling riots, 8 August 2024
202 Sky News, Southport attack misinformation fuels far-right discourse on social media, 31 July 2024
203 Checkmyads, Digital Advertising and Its Role in the 2024 Southport Riots, March 2025
204 Letter from the Chair of the Science, Innovation and Technology Committee regarding Follow-ups from 25 February oral evidence session, 20 March 2025; Google (SMH0079)
205 Checkmyads, Digital Advertising and Its Role in the 2024 Southport Riots, March 2025
206 Q37 [Amanda Storey]; Google (SMH0079)

97. **CONCLUSION**

There is a regulatory gap around digital advertising, as much of the regulation and interventions have been industry-led and focused on tackling harmful advertising content, as opposed to the monetisation of harmful content through advertising. We are not convinced that the digital advertising industry is able, or willing, to effectively self-regulate. The government's reliance on industry-led, content-focused solutions, is insufficient to meet the current scale of harm. One industry-led intervention that saw some success in increasing transparency and reducing monetisation of harmful content, the Global Alliance for Responsible Media, ended following legal challenge.

98. **RECOMMENDATION**

Tackling online harm means addressing the principles that incentivise and monetise its spread. In line with Principle 3, responsibility, the government should create a new arms-length body—not funded by industry—to regulate and scrutinise the process of digital advertising, covering the complex and opaque automated supply chain that allows for the monetisation of harmful and misleading content. Or, at the least, the government should extend Ofcom's powers to explicitly cover this form of harm, and regulate based on the principle of preventing the spread of harmful or misleading content through any digital means, rather than limiting itself to specific technologies or sectors.

99. **RECOMMENDATION**

To tackle the incentive behind amplified misinformation—namely, the monetisation of harmful content—there should be clear and enforceable standards for digital advertising market processes, as well as advertising content. Following our Principles 1, 3 and 5, government should ask the Advertising Standards Authority to establish comprehensive guidelines for all actors within the digital advertising ecosystem and supply chain. These should be informed by the UN's 2024 Guiding Principles for Information Integrity and developed in consultation with civil society, academics, experts, industry and policymakers. It should be designed to remove incentives for algorithmic acceleration of harmful or misleading content whilst upholding freedom of expression; ensure advertisers can avoid harmful content; and ensure transparency in technologies with public safety implications, such as digital advertising.

100. **RECOMMENDATION**
The internet, and social media, could not operate without digital advertising. Given its implications for public safety, as per Principle 5, there needs to be heightened transparency in the market processes of online advertising. Government should mandate 'Know Your Customer' checks for participants in the programmatic advertising supply chain, as exists in other large markets. The government should also ensure that platforms disclose full ad campaign data, and allow independent third-party audits and vetting of ad exchange supply partners.

101. **RECOMMENDATION**
There are insufficient disincentives for bad practice in the digital advertising market. Bad actors can exploit the ecosystem, monetising harmful content through major platforms. Following Principle 3, Ofcom should be empowered to give penalty notices to platforms when they allow harmful content to be monetised through their services. These penalties should be based on a formula that considers: the severity of harm, the amount of revenue the publisher received, the amount of revenue the platform received, and the number of individuals that encountered the harmful content. The revenue generated from these penalties should be used to support victims of online harms.

# Annex: Legal definitions

## Freedom of expression

Prior to the Human Rights Act 1998, there was no single legal statute that defined "free speech" or "freedom of expression" for the purposes of UK law. The Human Rights Act 1998 incorporated into UK law Article 10 of the European Convention on Human Rights (ECHR), which states:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.[207]

The caselaw of the European Court of Human Rights sets out that Article 10 "is applicable not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population."[208]

Freedom of expression is a qualified right that can be restricted under certain conditions. Article 10(2) of the ECHR states that:

> The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[209]

There is no legal definition of "hate speech" in UK criminal law. "Hate crime" can cover any crime where the offender has "demonstrated [or been motivated by] hostility based on race, religion, disability, sexual orientation

207 S 12, Human Rights Act 1998; Article 10, European Convention on Human Rights
208 Handyside v United Kingdom (1976) 1 EHRR 737
209 Article 10(2), European Convention on Human Rights

or transgender identity."[210] The Public Order Act 1986 makes it an offence to stir up racial hatred, with later amendments adding religious hatred, and hatred on the grounds of sexual orientation.[211]

The Online Safety Act 2023 introduced the 'False Communications' offence. A person commits an offence if:

**a.** the person sends a message

**b.** the message conveys information that the person knows to be false,

**c.** at the time of sending it, the person intended the message, or the information in it, to cause non-trivial psychological or physical harm to a likely audience, and

**d.** the person has no reasonable excuse for sending the message.[212]

### Online services

Section 230 of the US Communications Act of 1934, enacted as part of the Communications Decency Act of 1996, states the following:

> No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.[213]

In the US, this "provides immunity to online platforms from civil liability based on third-party content and for the removal of content in certain circumstances."[214] In 2020, the US Department of Justice announced a review of Section 230, stating that:

> The Department of Justice has concluded that the time is ripe to realign the scope of Section 230 with the realities of the modern internet."[215]

210 CPS, Hate crime (accessed June 2025)
211 Part 3, Public Order Act 1986
212 S 179, Online Safety Act 2023
213 47 U.S. Code § 230
214 Department of Justice, Department of Justice's Review of Section 230 of the Communications Decent Act of 1996, June 2020
215 Department of Justice, Department of Justice's Review of Section 230 of the Communications Decent Act of 1996, June 2020

# Conclusions and recommendations

## Introduction

1. In the course of this inquiry, we identified five key principles that we believe are crucial for regulation of social media and related technologies:

   1) Public safety: Algorithmically accelerated misinformation is a danger that companies and government need to address—the government and platform companies should work together to protect the public from it.

   2) Free and safe expression: Neither government nor private companies should be arbiters of truth. Steps to tackle amplified misinformation should be in line with the fundamental right of free expression, with restrictions where proportionate and necessary to protect national security, public safety, health, or to prevent disorder and crime.

   3) Responsibility: Users should be held liable for what they post online, but the platforms they post on are also responsible, especially with regard to the systems used to moderate, circulate or amplify content.

   4) Control: Users should have control over both their personal data and what they see online. This includes the right to delete the data stored by platforms and services which is used to drive content and advertisement recommendation algorithms.

   5) Transparency: The technology used by platform companies, including social media algorithms, has huge public safety implications, and should be transparent and accessible to public authorities. (Conclusion, Paragraph 6)

## Misleading and harmful content on social media

2. We launched this inquiry in the wake of the riots that followed the horrific attack in Southport in 2024. We received overwhelming evidence that online activity, including social media recommendation algorithms amplifying harmful and misleading content, played a key part in driving the unrest and violence. Social media companies' responses were inconsistent and

inadequate, often enabling, if not encouraging, this viral spread, with evidence that they may have profited due to the heightened engagement. The evidence supports the conclusion that social media business models incentivise the spread of content that is damaging and dangerous, and did so in a manner that endangered public safety in the hours and days following the Southport murders. (Conclusion, Paragraph 14)

3. The Online Safety Act was not designed to tackle misinformation—we heard that even if it had been fully implemented, it would have made little difference to the spread of misleading content that drove violence and hate in summer 2024. Therefore, the Act fails to keep UK citizens safe from a core and pervasive online harm. (Conclusion, Paragraph 18)

4. We welcome Ofcom's consultation on a 'crisis response protocol' for companies to follow in response to events such as the 2024 unrest. The protocol should directly address misinformation by including all online services at risk of contributing to the spread of false or harmful information, including large online social media, search and messaging services; those with smaller user numbers but high-risk profiles; and others, such as generative AI platforms. In establishing the mechanism, Ofcom should acknowledge the different ways in which different services operate. Following our Principle 2, it should hold platforms responsible for: decelerating the spread of harmful misinformation without censoring lawful speech; ensuring substantial and continuous engagement with law enforcement and government bodies; giving users control over the content they see; and providing transparency around their actions. (Recommendation, Paragraph 19)

5. Social media and other online platforms have huge power and reach into our lives, with positive and negative impacts. They can democratise knowledge and access to the public sphere, and help to build social connections and global communities. Generative AI provides further opportunities in terms of productivity, creativity and content moderation. For these reasons, it is imperative that we regulate and legislate these technologies based on the principles set out in this report, harnessing the digital world in a way that protects and empowers citizens. (Conclusion, Paragraph 25)

6. Internet users are exposed to large volumes of harmful and misleading content which can deceive, damage mental health, normalise extremist views, undermine democracy, and fuel violence. We are concerned by the evidence that recommendation algorithms—integral to the advertisement- and engagement-driven business models of social media companies—play a role in this. Young people are particularly vulnerable to these harms, and

those born today will never have known a world without AI—we plan to explore in detail the impact the online world has on their developing brains in our future work. (Conclusion, Paragraph 26)

7. The technology used by social media companies should be transparent, explainable and accessible to public authorities, as stated in our Principle 5. This is currently not the case: when we asked, major platforms did not give us detailed, transparent up-to-date representations of their recommendation algorithms. (Conclusion, Paragraph 27)

8. Social media companies have often argued that they are not publishers but platforms, abdicating responsibility for the content they put online. We believe that these services, with sophisticated recommendation algorithms that directly amplify and push content to users, are not merely platforms but curators of content. As we have seen, the amplification and spread of this content can have serious, large-scale impacts. We recognise that this is a complex area of law and that defining social media companies as publishers would have major consequences, but the current situation is deeply unsatisfactory. We call on the government to set out its position on this question in its response to this report. (Conclusion, Paragraph 28)

9. There is a shortfall in data needed to accurately analyse the scale of the problem and identify policy solutions. In line with our Principle 4, the government should commission a large-scale research project into how far social media recommendation systems spread, amplify or prioritise harmful content. This should be undertaken by a group of credible independent researchers, bringing diverse perspectives, with full access to the inner functions of the systems that major platforms use to algorithmically recommend content, including the private, external, and third party data used to train their systems; the user, content and engagement attributes the algorithms rely on and how these are weighted, and the objectives the algorithms are optimised for; where user interactions reinforce future recommendations; and any curation rules or interventions that influence promotion or suppression of content. We expect full cooperation from all major services that employ recommendation algorithms. (Recommendation, Paragraph 29)

10. Based on the research described above, the government should publish conclusions on the level and nature of harm that these platforms promote through their recommendation systems. Following our Principle 3, if significant harm is found, the responsible online services should publish the actions they will take to address these harms. Ofcom should be given the power to serve penalty notices to services that fail to comply, either 10% of the company's worldwide revenue, or £18 million, whichever is higher. (Recommendation, Paragraph 30)

11. Following our Principles 2 and 3, the government should compel social media platforms to embed tools within their systems that identify and algorithmically deprioritise fact-checked misleading content, or content that cites unreliable sources, where it has the potential to cause significant harm. It is vital that these measures do not censor legal free expression, but apply justified and proportionate restrictions to the spread of information to protect national security, public safety or health, or prevent disorder or crime. (Recommendation, Paragraph 31)

12. As per Principle 4, users should have more control over the content that is pushed to them online. Government should mandate all online services with a content recommendation algorithm to give the user a 'right to reset', which would delete all data stored by their recommendation algorithm, in the manner that users can clear their cookie history. This option should be displayed prominently on the platform's main feed or homepage. (Recommendation, Paragraph 32)

13. The UK government—like its counterparts around the world—is facing the challenge of attempting to regulate hugely powerful technology companies that operate across the world, providing technologies that transform societies, with bigger budgets than many countries. It is essential that their impact on our society be understood, effectively scrutinised and, where necessary, regulated in the public interest by Parliament. The committee has experienced some of the challenges of this in its engagement with these companies. We were reassured by the statements from Google, Meta, TikTok and X in our evidence session that they accepted their responsibility to be accountable to Parliament, and we hope that this will be put into practice. (Conclusion, Paragraph 39)

14. We are concerned by disjointed approaches from platforms to false and harmful content; in particular by recent moves from X and Meta to water down their Terms of Service and approach to content moderation. While there are merits to crowd-sourcing models of context provision and fact-checking—as part of a wider policy on misleading and harmful content—these platforms seem to be prioritising this method over third-party fact checking without clear evidence on whether it will adequately protect users from algorithmically amplified harm and misinformation. (Conclusion, Paragraph 40)

15. In line with our Principle 1 of tackling amplified misinformation, the government should compel platforms to put in place minimum standards for addressing the spread of misleading content online. More information is needed on the merits of different approaches to this. The government should commission research into the relative benefits of independent third-party fact-checkers, crowd-sourced context provision, and AI driven detection of misinformation, using researchers who are independent, bring

diverse perspectives on the issue, and have full access to the data of these systems. The research should enable Ofcom to offer guidance on the most effective method, or combination of methods, to address misinformation. (Recommendation, Paragraph 41)

16. The Online Safety Act will lead to some improvements, but is designed only to protect users from harm that is illegal or affects children. The decision not to include measures related to the algorithmic amplification of "legal but harmful" content, such as misinformation, means that full enforcement of the Act would have made little difference to the online environment that helped to incite the violence of last summer. (Conclusion, Paragraph 45)

17. It is vital that platforms are held responsible for the algorithmic spread of misleading or deceptive content that can radicalise and harm users. The few measures in the Act that address misinformation fall short. The False Communications offence is vaguely worded and will be difficult to implement; the advisory committee on mis/disinformation has been renamed, suggesting a change of emphasis; and the media literacy strategy does not set ambitious goals. (Conclusion, Paragraph 46)

18. The broad scale—and serious impact—of misinformation online requires greater transparency and accountability from the government. In line with our Principle 1, the government should submit an annual report to Parliament on the state of misinformation online, tracking trends and issues from the year, and setting out successes and failures in addressing them. (Recommendation, Paragraph 47)

19. In line with Principle 5, transparency, the government should introduce duties for platforms to undertake risk assessments and reporting requirements on legal but harmful content, such as potentially harmful misinformation, with a focus on the role of recommendation algorithms in its spread. (Recommendation, Paragraph 48)

20. To ensure true responsibility from platform companies, as per Principle 3, Ofcom and DSIT should confirm that services are required to act on all risks identified in risk assessments, regardless of whether they are included in Ofcom's Codes of Practice. (Recommendation, Paragraph 49)

21. The Online Safety Act does not do enough to address the risks posed by small platforms due to its exclusive focus on size. Ofcom should create an additional category to cover 'small but risky' platforms, based on analysis of the role that harmful smaller platforms can play in the online ecosystem, interacting with the recommendation algorithms of large platforms to spread harms such as misinformation, and disinformation campaigns. This regulation of small platforms should be in line with our Principles 1, 3 and 5. (Recommendation, Paragraph 55)

22. Foreign interference and disinformation campaigns, with use of technology such as bots and AI, put UK citizens at risk. The possibility that some of the divisive messages and deceptive content spread by users—and amplified by algorithms—last summer were part of such an influence operation is deeply concerning. In order to tackle amplified disinformation, identified by Principle 1, the government and Ofcom should collaborate with platforms to identify and track disinformation actors and the techniques and behaviours they use to spread adversarial and deceptive narratives online. (Recommendation, Paragraph 58)

23. Responsibility for tracking foreign disinformation campaigns appears to be split between several departments, including DSIT. This suggests that the Intelligence and Security Committee's 2020 characterisation of countering Russian influence operations as a "hot potato", passed between different bodies, has not been addressed. To meet our Principle 1, the government should clarify which department has ownership over tracking and countering online narrative operations. It should consider consolidating responsibility within a single entity, for example the National Security Online Information Team, or establishing a clear chain of command, and in its response to this report the government should set out the actions it intends to take in this regard. (Recommendation, Paragraph 59)

24. The NSOIT is an important tool in protecting citizens from disinformation and needs appropriate scrutiny. Government should place NSOIT on a statutory footing and bring it under the remit of the Intelligence and Security Committee, to ensure that our Principle 1 is being effectively and safely pursued, in line with Principle 2. (Recommendation, Paragraph 62)

## Generative AI

25. The Online Safety Act does not protect users from the commodification of synthetic mis/disinformation, or provide effective transparency for the systems that produce them. It fails to address the issue of tech companies rolling out experimental features that can feed false or harmful information to their enormous audiences, further threatening information integrity online. This has damaging effects on all users, undermining the reputation of the companies that introduce them. (Conclusion, Paragraph 74)

26. We are concerned at what appears to be contradiction and confusion between regulators and government over the capabilities, limitations and principles behind the Online Safety Act. We expect senior Ofcom officials and ministers to be fully aligned in their understanding of the Act—particularly in relation to critical issues such as misinformation and generative AI. Ofcom at times appeared complacent in its approach to

public safety online, failing to live up to its role as the UK's online safety regulator and slipping into the role of mediator between the industry and the consumer. (Conclusion, Paragraph 75)

**27.** To protect citizens from the AI-exacerbated spread of misinformation and harm, the government should pass legislation that covers generative AI platforms, bringing them in line with other online services that pose a high risk of producing or spreading illegal or harmful content. Following the Principles identified by this report, this legislation should require generative AI platforms to: provide risk assessments to Ofcom on the risks associated with different prompts and outputs, including how far they can create or spread illegal, harmful or misleading content; explain to Ofcom how the model curates content, responds to sensitive topics and what guardrails are in place to prevent content that is illegal or harmful to children; implement user safeguards such as feedback, complaints and output flagging; and prevent children from accessing inappropriate or harmful outputs. (Recommendation, Paragraph 76)

**28.** Principle 5 is crucial for addressing potential harms from generative AI, as there is currently a serious shortfall in transparency and oversight of the platforms and systems that allow users to create AI-generated content. The government should require providers of generative AI services to provide information to those carrying out independent research into online safety. This should include data such as platforms' internal decision-making processes, training datasets, optimisation objectives, safety mechanisms and guardrails on outputs. (Recommendation, Paragraph 77)

**29.** To effectively tackle amplified misinformation as per Principle 1, the government should work with relevant experts and platforms to develop technology that automatically detects AI-generated media, meeting mis/disinformation at its source. It should mandate all generative AI platforms, and platforms that employ generative AI technologies, to automatically label AI-generated media with metadata and visible watermarks that cannot be removed. (Recommendation, Paragraph 78)

## Digital advertising market

**30.** Advertising is crucial to major social media companies, which depend on recommending engaging content to increase time spent on their platforms and draw attention to adverts. Their recommendation algorithms do not effectively differentiate between harmless and harmful engaging content, which can result in promotion of misleading, damaging, or hateful material. The effects spread through the online ecosystem, helping to incentivise the production and spread of harmful content. (Conclusion, Paragraph 85)

**31.** The global digital advertising market is overcomplicated, opaque and under-regulated, operating through an enormous, automated and inaccessible supply chain. This directly leads to the production, viral spread and monetisation of harmful and deceptive content, often without advertisers' knowledge. Platforms and advertisers appear to be either unable or unwilling to address this problem. We heard evidence that platforms may have profited from misinformation and hateful content after the Southport attack. (Conclusion, Paragraph 95)

**32.** In particular, we were concerned by evidence that Google may have helped to monetise misinformation relating to the attacks, contributing to the violence. This is unacceptable, and is just one example of a much wider problem with the digital advertising industry. We are concerned that Google was seemingly unaware of the chain of events when we asked them about it; failed to tell us how much revenue was earned from this; and failed to reassure us that the company would prevent this from happening again. (Conclusion, Paragraph 96)

**33.** There is a regulatory gap around digital advertising, as much of the regulation and interventions have been industry-led and focused on tackling harmful advertising content, as opposed to the monetisation of harmful content through advertising. We are not convinced that the digital advertising industry is able, or willing, to effectively self-regulate. The government's reliance on industry-led, content-focused solutions, is insufficient to meet the current scale of harm. One industry-led intervention that saw some success in increasing transparency and reducing monetisation of harmful content, the Global Alliance for Responsible Media, ended following legal challenge. (Conclusion, Paragraph 97)

**34.** Tackling online harm means addressing the principles that incentivise and monetise its spread. In line with Principle 3, responsibility, the government should create a new arms-length body—not funded by industry—to regulate and scrutinise the process of digital advertising, covering the complex and opaque automated supply chain that allows for the monetisation of harmful and misleading content. Or, at the least, the government should extend Ofcom's powers to explicitly cover this form of harm, and regulate based on the principle of preventing the spread of harmful or misleading content through any digital means, rather than limiting itself to specific technologies or sectors. (Recommendation, Paragraph 98)

**35.** To tackle the incentive behind amplified misinformation—namely, the monetisation of harmful content—there should be clear and enforceable standards for digital advertising market processes, as well as advertising content. Following our Principles 1, 3 and 5, government should ask the Advertising Standards Authority to establish comprehensive guidelines for

all actors within the digital advertising ecosystem and supply chain. These should be informed by the UN's 2024 Guiding Principles for Information Integrity and developed in consultation with civil society, academics, experts, industry and policymakers. It should be designed to remove incentives for algorithmic acceleration of harmful or misleading content whilst upholding freedom of expression; ensure advertisers can avoid harmful content; and ensure transparency in technologies with public safety implications, such as digital advertising. (Recommendation, Paragraph 99)

**36.** The internet, and social media, could not operate without digital advertising. Given its implications for public safety, as per Principle 5, there needs to be heightened transparency in the market processes of online advertising. Government should mandate 'Know Your Customer' checks for participants in the programmatic advertising supply chain, as exists in other large markets. The government should also ensure that platforms disclose full ad campaign data, and allow independent third-party audits and vetting of ad exchange supply partners. (Recommendation, Paragraph 100)

**37.** There are insufficient disincentives for bad practice in the digital advertising market. Bad actors can exploit the ecosystem, monetising harmful content through major platforms. Following Principle 3, Ofcom should be empowered to give penalty notices to platforms when they allow harmful content to be monetised through their services. These penalties should be based on a formula that considers: the severity of harm, the amount of revenue the publisher received, the amount of revenue the platform received, and the number of individuals that encountered the harmful content. The revenue generated from these penalties should be used to support victims of online harms. (Recommendation, Paragraph 101)

# Formal Minutes

## Wednesday 25 June 2025

### Members present

Dame Chi Onwurah (in the Chair)

Dr Allison Gardner

Kit Malthouse

Dr Lauren Sullivan

Adam Thompson

## Social media, misinformation and harmful algorithms

Draft Report (*Social media, misinformation and harmful algorithms*), proposed by the Chair, brought up and read.

*Ordered*, That the draft Report be read a second time, paragraph by paragraph.

Paragraphs 1 to 101 read and agreed to.

Annex agreed to.

Summary agreed to.

*Resolved*, That the Report be the Second Report of the Committee to the House.

*Ordered*, That the Chair make the Report to the House.

*Ordered*, That embargoed copies of the Report be made available (Standing Order No. 134).

## Adjournment

Adjourned till Tuesday 1 July 2025 at 9 am

# Witnesses

The following witnesses gave evidence. Transcripts can be viewed on the inquiry publications page of the Committee's website.

## Tuesday 21 January 2025

**Zara Mohammed**, Secretary General, Muslim Council of Britain; **Ravishaan Muthiah**, Director of Communications, Joint Council for the Welfare of Immigrants; **Kelly Chequer**, Councillor, Sunderland City Council Q1–11

**Marianna Spring**, Disinformation and social media correspondent, BBC; **Mr Imran Ahmed**, CEO, Center for Countering Digital Hate; **Dr Joe Whittaker**, Lecturer, School of Social Sciences, Cyber Threats Research Centre, Swansea University Q12–27

## Tuesday 25 February 2025

**Amanda Storey**, Managing Director, Trust & Safety at Google EMEA, Google Q28–87

## Tuesday 25 February 2025

**Chris Yiu**, Director of Public Policy for Northern Europe, Meta; **Ali Law**, Director of Public Policy and Government Affairs, UK and Ireland, TikTok; **Wifredo Fernandez**, Senior Director for Government Affairs, X (formerly known as Twitter) Q88–179

## Tuesday 18 March 2025

**Dr Karen Middleton**, Senior Lecturer in Marketing, University of Portsmouth and Advisor to the Conscious Advertising Network; **Phil Smith**, Director General, Incorporated Society of British Advertisers (ISBA) Q180–215

**Dr Eirliani Abdul Rahman**, Online Safety Advocate and Former Trust and Council Member, Twitter; **Lyric Jain**, CEO, Logically Q216–248

## Tuesday 29 April 2025

**Mark Bunting**, Director, Online Safety Strategy Delivery, Ofcom; **John Edwards**, Information Commissioner, Information Commissioner's Office
Q249–229

**The Baroness Jones of Whitchurch,** Minister for the Future Digital Economy and Online Safety, Department for Science, Innovation and Technology; **Talitha Rowland**, Director for Security and Online Harm, Department for Science, Innovation and Technology
Q300–342

# Published written evidence

The following written evidence was received and can be viewed on the inquiry publications page of the Committee's website.

SMH numbers are generated by the evidence processing system and so may not be complete.

| | | |
|---|---|---|
| 1 | 5Rights Foundation | SMH0024 |
| 2 | ACM Europe Technology Policy Committee | SMH0035 |
| 3 | Abdul Rahman, Dr Eirliani | SMH0074 |
| 4 | Abrusci, Dr Elena (Senior Lecturer in Law, Brunel, University of London) | SMH0050 |
| 5 | Amnesty International | SMH0083 |
| 6 | Antisemitism Policy Trust | SMH0005 |
| 7 | Atlantic Council's Democracy + Tech Initiative | SMH0034 |
| 8 | Big Brother Watch | SMH0043 |
| 9 | Casas Salleras, Dr. Andreu (Lecturer in Political Communication, Department of Politics and International Relations, Royal Holloway University of London); Georgia Dagher (Research Assistant, Department of Politics and International Relations, Royal Holloway University of London); and Prof. Ben O'Loughlin (Professor of International Relations, Department of Politics and International Relations, Royal Holloway University of London) | SMH0030 |
| 10 | Center for Countering Digital Hate (CCDH) | SMH0009 |
| 11 | Children's Commissioner for England's Office | SMH0014 |
| 12 | Clean up the Internet | SMH0023 |
| 13 | Computer and Communications Industry Association | SMH0029 |
| 14 | Dabbagh, Dr Hossein (Assistant Professor in Philosophy, Northeastern University London) | SMH0017 |
| 15 | Digital Mental Health Programme, University of Cambridge | SMH0027 |
| 16 | Edwards, John (UK Information Commissioner, Information Commissioner's Office) | SMH0085 |
| 17 | Faculty of Public Health | SMH0011 |

| | | |
|---|---|---|
| 18 | Foxglove | SMH0066 |
| 19 | Full Fact | SMH0070 |
| 20 | Full Fact | SMH0047 |
| 21 | Gentile, Dr Giulia (Lecturer in Law, Essex Law School); and Professor Lorna Woods (Professor of Internet Law, Essex Law School) | SMH0038 |
| 22 | Glitch | SMH0028 |
| 23 | Global Witness | SMH0048 |
| 24 | Google | SMH0079 |
| 25 | Google | SMH0065 |
| 26 | Hardcastle, Dr Kimberley (Assistant Professor in Business and Marketing, Northumbria University) | SMH0026 |
| 27 | Hilbert, Professor Martin (Professor, University of California, Davis) | SMH0008 |
| 28 | Howard, Professor Jeffrey (Professor of Political Philosophy and Public Policy, University College London); and Dr Maxime Lepoutre (Lecturer in Politics and International Relations, University of Reading) | SMH0013 |
| 29 | Incorporated Society of British Advertisers (ISBA) | SMH0075 |
| 30 | Institute for Strategic Dialogue | SMH0062 |
| 31 | Integrity Institute | SMH0054 |
| 32 | Internet User Behaviour Lab | SMH0058 |
| 33 | Joint Council for the Welfare of Immigrants | SMH0067 |
| 34 | Jones, Marc Owen (Associate Professor, Northwestern University in Qatar) | SMH0071 |
| 35 | Kira, Dr Beatriz (Assistant Professor in Law, University of Sussex); Professor Julie Weeds (Professor in Artificial Intelligence, University of Sussex); Professor Phoebe Li (Professor of Law and Technology, University of Sussex); and Zoe Asser (Research Assistant, University of Sussex) | SMH0056 |
| 36 | Kumar, Dr Akshi (Senior Lecturer, Department of Computing, Goldsmiths, University of London) | SMH0040 |
| 37 | Logically | SMH0076 |
| 38 | Logically | SMH0049 |
| 39 | MacDermott, Dr Aine (Senior Lecturer, Liverpool John Moores University) | SMH0010 |

| | | |
|---|---|---|
| 40 | Marlow, Josephine | SMH0012 |
| 41 | McDonald, Professor Kevin (Professor of Sociology, Middlesex University London) | SMH0020 |
| 42 | Meta | SMH0080 |
| 43 | Meta | SMH0037 |
| 44 | Middleton, Dr Karen (Senior Lecturer in Marketing and Advisor, University of Portsmouth and the Conscious Advertising Network) | SMH0077 |
| 45 | Middleton, Dr Karen (Senior Lecturer in Marketing and Advisor, University of Portsmouth and the Conscious Advertising Network) | SMH0036 |
| 46 | Minderoo Centre for Technology and Democracy, University of Cambridge | SMH0051 |
| 47 | Molly Rose Foundation | SMH0016 |
| 48 | NSPCC | SMH0032 |
| 49 | Ofcom | SMH0086 |
| 50 | Ofcom | SMH0078 |
| 51 | Office of the Kent Police and Crime Commissioner | SMH0019 |
| 52 | Online Safety Act Network | SMH0031 |
| 53 | OpenMined Foundation | SMH0046 |
| 54 | Oxford Internet Institute, University of Oxford | SMH0057 |
| 55 | Penrose, John (Founder & Director, Centre for Small State Conservatives) | SMH0002 |
| 56 | Penrose, John (Founder & Director, Centre for Small State Conservatives) | SMH0003 |
| 57 | Policy Connect | SMH0063 |
| 58 | Popa-Wyatt, Dr Mihaela (Lecturer in Philosophy, The University of Manchester) | SMH0045 |
| 59 | REPHRAIN | SMH0033 |
| 60 | Sastry, Professor Nishanth (Principal Investigator of the AP4L project, Associate Head of School for Research and Innovation at the School of Computer Science and Electronic Engineering, and co-lead of the Social Data Science Special Interest Group at the Alan Turing Institute, University of Surrey); Dr Diptesh Kanojia (Lecturer in Artificial Intelligence for Natural Language Processing, University of Surrey); Professor Alice Hutchings (Professor | |

| | | |
|---|---|---|
| | of Emergent Harms, University of Cambridge); and Professor Gareth Tyson (Professor of Computer Science, Queen Mary University of London) | SMH0055 |
| 61 | Seargeant, Dr Philip (Senior Lecturer in Applied Linguistics, Open University) | SMH0006 |
| 62 | Sense about Science | SMH0041 |
| 63 | Shout Out UK | SMH0072 |
| 64 | Spring, Marianna (Disinformation and social media correspondent, BBC) | SMH0069 |
| 65 | Sutherland, Dr Jessica (Research Fellow, University of Warwick); and Professor Keith Hyams (Professor of Political Theory and Ethics, University of Warwick) | SMH0015 |
| 66 | The Alan Turing Institute (CETaS) | SMH0007 |
| 67 | The Electoral Commission | SMH0021 |
| 68 | The Free Speech Union | SMH0059 |
| 69 | TikTok | SMH0081 |
| 70 | TikTok | SMH0068 |
| 71 | Tong, Dr Jingrong (Senior Lecturer in Media and Information Studies, University of Sheffield) | SMH0025 |
| 72 | UK Government | SMH0061 |
| 73 | UK Safer Internet Centre | SMH0044 |
| 74 | UK Stop Ad Funded Crime (UKSAFC) | SMH0004 |
| 75 | UKCVFamily | SMH0022 |
| 76 | University of Manchester | SMH0053 |
| 77 | van der Linden, Professor Sander (Professor of Social Psychology in Society, University of Cambridge); Dr Jon Roozenbeek (Lecturer in Psychology and Security, King's College London); and Professor Stephan Lewandowsky (Chair in Cognitive Psychology, University of Bristol) | SMH0052 |
| 78 | Whitchurch, Baroness Jones of (Minister for the Future Digital Economy and Online Safety, Department for Science, Innovation and Technology) | SMH0084 |
| 79 | Whittaker, Dr Joe (Senior Lecturer, Swansea University); Miss Ellie Rogers (Doctoral Researcher, Swansea University); Dr Nicholas Micallef (Lecturer, Swansea University); and Dr Sara Correia (Lecturer, Swansea University) | SMH0018 |

| | | |
|---|---|---|
| 80 | X (formerly known as Twitter) | SMH0064 |
| 81 | X (formerly known as Twitter) | SMH0082 |
| 82 | Yoti | SMH0039 |

# List of Reports from the Committee during the current Parliament

All publications from the Committee are available on the publications page of the Committee's website.

## Session 2024–25

| Number | Title | Reference |
|---|---|---|
| 1st | Pre-appointment hearing for the Executive Chair of Innovate UK | HC 834 |
| 2nd Special | Insect decline and UK food security: Government Response | HC 717 |
| 1st Special | Governance of artificial intelligence (AI): Government Response | HC 591 |