

Digital & Media Interventions

Policy Proposals to Protect British Information Integrity in a New Era

Council for Countering Online Disinformation

About us

The CCOD is a mission-led organisation fighting against online disinformation. We have two primary goals that define who we are, what we do, and what we stand for:

1) We aim to influence the policy and regulatory environment to ensure that the UK government effectively tackles the growing issue of online misinformation and disinformation. CCOD has helped convene a cross-party group of parliamentarians and works directly with policymakers to improve understanding of disinformation threats. We educate policymakers on disinformation challenges and work with them on workable policy, regulatory and legislative initiatives to protect online information integrity in the UK.

2) Our objective is to bring together the key actors working across the information integrity ecosystem, ranging from technology developers and researchers to civil society organisations and media specialists, and ensure that those at the forefront of developing innovative solutions to online disinformation are given a prominent voice in shaping public policy outcomes. In parallel, we support organisations in protecting themselves and their workforces from disinformation risks with our advisory services, helping to safeguard markets, economic stability, and public trust.

Our collaborators and partners have included RUSI, the Aspen Institute, the Alan Turing Institute, Adobe, Egregious, Refute, Schillings, Ghost Shift, and academics from the University of Cambridge, UCL, and LSE. Our work has been featured on Times Radio, Politico and PoliticsHome.

Our research and policy thinking is supported by a world-class advisory board, which includes, among others, John Penrose (former MP and Government Minister); Professor Sander van der Linden (Professor of Social Psychology at the University of Cambridge); Matthew Wicker (Assistant Professor at Imperial College and Research Associate at the Alan Turing Institute); Joseph Brandifino (Leads Policy Strategy at Scale AI); Naomi Krieger Carmy (VP Portfolio Management at the DeepTech VC fund Earth & Beyond Ventures); Clare Rewcastle Brown (award-winning investigative journalist); and a former Chair of Ofcom.

Table of Contents

About us 1
Problem Statement 3
Factual Accuracy 6
Duty of Impartiality 8
Increasing the Prominence of PSB News 10
Media Literacy 11
Covering Micro, Local, Regional, and National News 12
Independent Disinformation Ombudsman 13
Tackling Fake Accounts 14
AI Watermarking 15
Summary 17
Bibliography 18

Problem Statement

Online disinformation is a growing threat to the United Kingdom's society, politics, and economy. Driven by generative artificial intelligence (AI), malicious actors can propagate false content in increasing volume and with ever-greater sophistication and complexity, blurring the lines between fiction and reality. These hostile entities include outfits both foreign and domestic, as well as state-backed and criminal.

The size of the danger has been repeatedly highlighted. In the last two years, both the Home Office and Ofcom presented evidence to the Joint Committee on National Security Strategy describing disinformation as an active threat to British democracy.[1][2] Across the Atlantic Ocean, the Office of the Director of National Intelligence's 2025 annual threat assessment labelled disinformation as a key tactic of Russia, China, and Iran in their malign influence operations.[3] The World Economic Forum, meanwhile, named misinformation and disinformation among its top global risks in 2026, alongside state-based conflict and environmental hazards.[4]

Similarly, the pervasiveness of the threat has also been widely reported. The Reuters Institute for the Study of Journalism, for example, reported that nearly six in ten are concerned globally about discerning between genuine and fabricated news stories online.[5] Closer to home, Ofcom reported that more than four in ten people it surveyed were exposed to misinformation and disinformation in 2024, with seven in ten stating they encountered it online.[6] Misinformation and disinformation can also take a multitude of forms, ranging from,

[1] Joint Committee on the National Security Strategy. (2025). Oral evidence: Defending democracy. [Online]. UK Parliament. Last Updated: 17 March 2025. Available at: https://committees.parliament.uk/oralevidence/15590/pdf/ [Accessed 27 March 2026].
[2] Joint Committee on the National Security Strategy. (2024). Oral evidence: Defending democracy. [Online]. UK Parliament. Last Updated: 18 March 2024. Available at: https://committees.parliament.uk/oralevidence/14513/pdf/ [Accessed 27 March 2026].
[3] Office of the Director of National Intelligence. (2025). Annual Threat Assessment of the U.S. Intelligence Community. [Online]. Office of the Director of National Intelligence. Last Updated: 25 March 2025. Available at: https://www.dni.gov/files/ODNI/documents/assessments/ATA-2025-Unclassified-Report.pdf [Accessed 27 March 2026].
[4] World Economic Forum. (2026). Global Risks Report 2026. [Online]. World Economic Forum. Last Updated: 4 January 2026. Available at: https://www.weforum.org/publications/global-risks-report-2026/digest/ [Accessed 27 March 2026].
[5] The Reuters Institute for the Study of Journalism. (2025). Overview and key findings of the 2025 Digital News Report. [Online]. Reuters Institute. Last Updated: 17 June 2025. Available at: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary [Accessed 27 March 2026].
[6] Ofcom. (2024). Four in 10 UK adults encounter misinformation. [Online]. Ofcom. Last Updated: 27 November 2024. Available at: https://www.ofcom.org.uk/media-use-and-attitudes/attitudes-to-news/four-in-10-uk-adults-encounter-mi [Accessed 27 March 2026].

but not limited to, deepfakes, AI-generated audio and video, unsubstantiated claims, to misleading images or social media posts.

Despite this wealth of evidence, disinformation has become increasingly politicised, especially globally. In the United States, the Trump administration has largely dismissed the threat of disinformation, notably abolishing the Counter-Foreign Information Manipulation and Interference Office, formerly known as the Global Engagement Center.[7] In September, the administration purportedly went further, allegedly terminating memorandums of understanding with European partners to counter disinformation campaigns waged by international adversaries.[8]

The political stance of the US government on this issue is apparent in Executive Order 14149—Restoring Freedom of Speech and Ending Federal Censorship. This order accused the Biden administration of violating the First Amendment rights of US citizens by censoring their speech on online platforms "under the guise of combating 'misinformation,' 'disinformation,' and 'malinformation'". The directive then prevented any federal agency from using federal resources in a manner that would "unconstitutionally abridge" free speech.[9]

While seemingly distant, this presents a serious problem to the UK, given that the Trump administration's hostility toward digital regulation extends beyond America's borders. President Trump himself expressed his ire when he vowed to "stand up" to foreign digital taxes, digital services legislation, and digital markets regulations he claimed discriminate against American firms.[10] The State Department was equally critical concerning the UK's Online Safety Act (OSA), which it claimed in its 2024 annual assessment of countries' human rights practices, limited free speech and freedom of the press.[11]

---

[7] M. Rubio. (2025). Protecting and Championing Free Speech at the State Department. [Online]. US Department of State. Last Updated: 16 April 2025. Available at: https://www.state.gov/protecting-and-championing-free-speech-at-the-state-department [Accessed 27 March 2026].

[8] A. Croft. (2025). US 'terminates agreements with Europe' to help fight disinformation from hostile states. [Online]. Independent. Last Updated: 9 September 2025. Available at: https://www.independent.co.uk/news/world/americas/us-europe-russia-disinformation-memorandum-b282314 [Accessed 27 March 2026].

[9] The White House. (2025). Restoring Freedom of Speech and Ending Federal Censorship. [Online]. The White House. Last Updated: 20 January 2025. Available at: https://www.whitehouse.gov/presidential-actions/2025/01/restoring-freedom-of-speech-and-ending-feder [Accessed 27 March 2026].

[10] D. Trump. (2025). "As the President of the United States, I will stand up to Countries that attack our incredible American Tech Companies". [Online]. Truth Social. Last Updated: 26 August 2025. Available at: https://truthsocial.com/@realDonaldTrump/posts/115092243259973570 [Accessed 27 March 2026].

[11] US Department of State. (2025). 2024 Country Reports on Human Rights Practices: United Kingdom. [Online]. US Department of State. Last Updated: 12 August 2025. Available at: https://www.state.gov/reports/2024-country-reports-on-human-rights-practices/united-kingdom [Accessed 27 March 2026].

The US has coupled this animosity with pressure to deregulate the UK's digital space, with the Trump administration suggesting that the UK could gain improved trading terms with the US if it relaxed its current digital regulations. This offer was explicitly made to Westminster when US Commerce Secretary Howard Lutnick hinted that the Technology Prosperity Deal could have attracted investment ten times greater – equivalent to roughly £310 billion – had the UK abolished the OSA and the Digital Services Tax.[12]

Despite this, the government has demonstrated that it will regulate US tech companies when necessary, as evident in the UK government's response to the controversy surrounding Grok, after the AI assistant began generating sexual child abuse content and intimate non-consensual images of individuals, primarily women, which were shared on X. When Ofcom announced that it had launched an investigation into the social media company, the government gave the watchdog its full backing, with Secretary of State for Science, Innovation and Technology Liz Kendall stating to Parliament that the government would support Ofcom "to use the full powers which Parliament has given them", up to and including a court order to stop UK users from accessing the site.[13] Minister for Safeguarding and Violence Against Women and Girls Jess Phillips, meanwhile, told The Times that the UK would make it illegal to create, possess, or distribute tools capable of generating child sexual abuse content, while other AI systems designed to generate non-consensual images would also be banned.[14]

Clearly, the government possesses the will to combat online harms. The challenge now is to encourage the authorities to channel this same energy toward tackling online disinformation, despite potential objections from the US. To assist the government in this effort, the CCOD is proposing six policy solutions: a factual accuracy amendment, a duty of impartiality amendment, a public service broadcaster (PSB) news prominence strategy, the establishment of a new independent ombudsman, action to tackle fake accounts, and AI watermarking.

[12] J. Nevett. (2025). Trump suggests Starmer could use military to control UK borders. [Online]. BBC News. Last Updated: 18 September 2025. Available at: https://www.bbc.co.uk/news/articles/cpd91wjypj9o [Accessed 27 March 2026].
[13] L. Kendall. (2026). Secretary of State statement to the House of Commons: 12 January 2026. [Online]. GOV.uk. Last Updated: 12 January 2026. Available at: https://www.gov.uk/government/speeches/secretary-of-state-statement-to-the-house-of-commons-12-janua [Accessed 27 March 2026].
[14] E. Yeomans | C. Alt. (2026). Elon Musk's AI undressing tool on Grok could be banned. [Online]. The Times. Last Updated: 7 January 2026. Available at: https://www.thetimes.com/uk/technology-uk/article/grok-ai-images-elon-musk-news-p6v2kfmfv [Accessed 27 March 2026].

Factual Accuracy

The effects of mis- and disinformation can only percolate through society and online spaces if there is a lack of factually accurate content to counter them. Studies have demonstrated that factually accurate content acts as an effective counter. A study by the European Commission's Joint Research Centre into "debunking", for example, found that individuals were less likely to fall for false claims and share them if they were provided with factually accurate information highlighting what was true and what was untrue.[15] A study published in the Cognitive Research: Principles and Implications, meanwhile, found that regardless of the format, corrective, factual information "substantially decreased belief in misinformation".[16]

In recent years, however, social media companies have taken steps that increased the reach of mis- and disinformation, but not that of factually accurate content, as illustrated by the cuts to content moderation, trust, and safety teams. In January 2025, Meta announced that it would discontinue its third-party fact-checking program.[17] A more striking example is X, which slashed its in-house trust and safety team from roughly 230 people to around 20 in the first year of Elon Musk's ownership.[18] This policy change has had a direct effect in enabling the spread of mis- and disinformation. A 2023 study by TrustLab of 6,000 social media posts found that disinformation was most prevalent on X across the six largest social networks, including Facebook, Instagram, LinkedIn, TikTok, and YouTube.[19]

It is essential, therefore, to push for policies that would correct this imbalance. As the evidence has shown, increasing the reach of factually accurate content is essential in the fight against mis- and disinformation, since truthful information is a direct antidote to the effects of maliciously created falsehoods. The CCOD therefore proposes the following amendment to the OSA to ensure factually accurate content.

---

[15] Joint Research Centre. (2024). Misinformation and disinformation: both prebunking and debunking work for fighting it. [Online]. European Commission. Last Updated: 25 October 2024. Available at: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/misinformation-and-disinformation-bo [Accessed 27 March 2026].
[16] B. Swire-Thompson, et al. (2021). Correction format has a limited role when debunking misinformation. [Online]. PubMed Central. Last Updated: 29 December 2021. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8715407/ [Accessed 27 March 2026].
[17] J. Kaplan. (2025). More Speech and Fewer Mistakes. [Online]. Meta. Last Updated: 7 January 2025. Available at: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ [Accessed 27 March 2026].
[18] K. Hays. (2023). New layoffs at Twitter hit trust and safety workers even as advertisers worry about toxic speech. [Online]. Business Insider. Last Updated: 19 September 2023. Available at: https://www.businessinsider.com/new-twitter-layoffs-hit-trust-and-safety-workers-again-2023-9 [Accessed 27 March 2026].
[19] C. Vallance. (2023). Disinformation most active on X, formerly known as Twitter, EU says. [Online]. BBC News. Last Updated: 26 September 2023. Available at: https://www.bbc.co.uk/news/technology-66926080 [Accessed 27 March 2026].

"Duty to empower adult users

(1) The Online Safety Act 2023 is amended as follows

(2) In section 15, subsection 3b, after the words 'alert the user', insert the words "including in a real-time, dynamic manner,"

(3) In section 15, after subsection (9), insert the words "or content which does not meet independent journalistic standards on factual accuracy

(3a) The Secretary of State may, by regulations, subject to approval by the super-affirmative procedure in Parliament, specify which independent journalistic standards on factual accuracy are valid for the purposes of section 15. The initial list shall include the standards in the IPSO Editors' Code of Practice and the IFJ Global Charter of Ethics for Journalists."

Member's explanatory statement

This amendment would place a duty on social media platforms to offer adult users simple, user-friendly controls to adjust how factually inaccurate content appears in their online social media feeds, so they can either exclude it completely or allow it to appear with clear, real-time labelling so they can make a fully-informed judgement about what it says. This amendment also clarifies how factual accuracy should not be defined by social media platforms, politicians, tech executives, or anyone with a vested interest: there are well-established, independent, international journalistic standards for what counts as factual accuracy that must be used instead.

This amendment calls on social media platforms and tech companies to develop real-time, dynamic fact-checking systems that provide users with toggles or filters, which are switched on by default. Additionally, content and news articles curated under these systems should not be altered by hosting providers.

Social media platforms and tech companies will need time to develop the necessary algorithmic and technical solutions required to make such a system possible. Therefore, the CCOD recommends that the government provide these entities with a short period of time

— six months to a year — to use the considerable resources at their disposal to develop their own factual-accuracy filters. Significantly, this amendment allows social media platforms and tech companies to develop these systems how they see fit, so long as their product meets the amendment's requirements.

However, it is also important to note that there already exists precedent for the sort of controls that the amendment calls for, as evidenced by Instagram's new feature, "Your Algorithm". As the company states, Your Algorithm allows users, through the power of AI, to view and personalise the content recommended to them by Instagram Reels, including whether they want to see more or less of a certain topic or subject matter.[20] While this level of user control is welcome, this policy could be taken one step further. Under the proposed amendment, a model such as Instagram's could be expanded to allow users to adjust their algorithm based on factual accuracy, so they can decide how much of the content recommended to them is factual.

Duty of Impartiality

Rather than creating a comprehensive code of conduct for digital platforms and social media companies from scratch, which may be difficult to enforce and likely to face American backlash, the government should instead consider implementing rules that require social media platforms to abide by the same duty-of-impartiality standards as traditional broadcasters.

This would help tackle disinformation by countering the effects of filter bubbles, referring to "a state of intellectual or ideological isolation that may result from algorithms feeding us information we agree with, based on our past behaviour and search history".[21] Filter bubbles exacerbate the threat of disinformation by trapping individuals in an informational echo chamber that constantly reinforces their beliefs. Within these limited informational spaces, evidence suggests that individuals are more likely to be exposed to mis- and disinformation and, if it aligns with their worldview, become more susceptible to it.

So long as algorithms are allowed to show users only forms of hyper-personalised content that align with their own beliefs in pursuit of higher engagement, then the problem and all its

[20] Instagram. (2025). Introducing a new way to see and control your algorithm. [Online]. Instagram. Last Updated: 10 December 2025. Available at: https://about.instagram.com/blog/announcements/reels-algorithm-control [Accessed 27 March 2026].
[21] Dr. R. Fletcher. (2020). The truth behind filter bubbles: Bursting some myths. [Online]. Reuters Institute. Last Updated: 24 January 2020. Available at: https://reutersinstitute.politics.ox.ac.uk/news/truth-behind-filter-bubbles-bursting-some-myths [Accessed 27 March 2026].

symptoms, growing polarisation, extremism, etc., will continue unabated. As a result, the CCOD is proposing the following amendment to the Communications Act 2003.

“Duty of Impartiality

Special applicability of OFCOM’s standards code to digital platforms

(1) The Communications Act 2003 is amended as follows.

(2) In section 319 (OFCOM’s standards code), after subsection (1), insert—

“(1A) Special content standards shall also apply to a “user-to-user-service”, as defined in section 3(1) of the Online Safety Act, when such services are designated by OFCOM.

(1B) OFCOM may only designate the services in para (1A) if they select or prioritise particular -

(a) User-generated content
(b) News-publisher content, or
(c) Comments and reviews on provider content

(1C) “Special content standards” means only the Duties of Due Impartiality and of Undue Prominence of Views and Opinions as per the OFCOM Broadcasting code.”

Member's explanatory statement:

This amendment guarantees freedom of speech and thought by ensuring citizens can continue to follow whoever they want and read whatever they choose, while preventing social media algorithms from only adding more extreme and one-sided versions of the same content to their feeds to create radicalisation echo-chambers and rabbit-holes. Instead, the additional content that platforms select would have to show users both sides of an argument, as the long-established 'duty of impartiality' has successfully done for broadcasters over many years. The additional content would have to comply with whatever factual accuracy settings each user had already selected and could include material from independent citizen-journalists or mainstream media, too.

As with the CCOD's factual accuracy proposal, social media platforms and tech companies will need time to develop the necessary technical and algorithmic solutions to make the proposed impartiality systems possible. The government, therefore, should allow these entities a short period of time — six months to a year — to create algorithmic impartiality systems of their own design, so long as they meet the balancing requirements of the Ofcom Broadcasting Code.

Increasing the Prominence of PSB News

Alongside proposals on factual accuracy and impartiality, the government should play to its strengths and utilise the PSBs at its disposal.

Unlike other media outlets, the PSBs — which include the BBC, ITV, Channel 4, and Channel 5 — are rigorously regulated by Ofcom under the Ofcom Broadcasting Code, which, as previously mentioned, imposes strict requirements for impartiality and factual accuracy on their content. Promoting the digital news content produced by the PSBs, therefore, is an effective way to tackle mis- and disinformation by increasing the reach of coverage that is legally required to be factually accurate and to offer balanced perspectives on news events.

As a result, the CCOD recommends that the government push social media platforms and tech companies to develop an additional toggle or filter system that would give prominence to sources governed by the Ofcom Broadcasting Code. If enabled by default, this setting would automatically prioritise PSBs within users' algorithms and social media feeds. As a result, PBS content would reach the majority of UK users, helping to break down filter bubbles and expose audiences to more reliable, factually accurate information — particularly given that most users do not actively opt out of default features.

Crucially, the PSB filter/toggle would respect the right of UK citizens to engage with the PSB content or not, by allowing them to disable the feature whenever they so choose, much like the CCOD's proposed factual accuracy filter.

In addition, the government should also adopt Ofcom's recommendation to increase the "prominence and discoverability" of PSB content on YouTube and potentially other third-party platforms. Video sharing platforms (VSPs), and principally YouTube, are increasingly the medium through which media (including news content) is consumed, especially by younger audiences.

As Ofcom's evidence shows, a British child aged between 4 and 15, for example, will spend significantly more time watching YouTube than all the PSBs combined.[22] It is crucial, therefore, that the government supports PSBs in expanding their presence on these platforms, enabling them to reach younger demographics better. If they fail in this, today's children may not turn to PSB content as they grow older, putting the future of public service media itself in jeopardy.

Media Literacy

To support the stated endeavours of the CCOD's proposed factual accuracy and duty of impartiality amendments, the government should also consider launching a public awareness campaign to inform the public about changes to social media platforms and their algorithms, as well as how the legislation empowers them.

This strategy is significant given that disinformation is becoming increasingly difficult to detect by the naked eye and ear. The government should emphasise that the new amendments would allow the public to make fully formed choices about the media and news content they consume, using the filters and algorithmic tools the amendments require.

While the public as a whole must be considered, the campaign should target younger people in particular, who are more likely to use non-traditional media for news. Ofcom's 2025 report into news consumption in the UK found that three-quarters of young people aged 16-24 used social media as the main way to access the news.[23] Being more likely to use alternative media, this demographic is also more likely to be exposed to mis- and disinformation, which means it is the age bracket most in need of convincing to utilise the tools and adjustments that the government would require social media platforms to offer.

Media literacy could be helpful in this regard. Along with raising awareness of the dangers of disinformation and teaching how to distinguish true from false content, effective media literacy interventions also encourage recipients to use and access trusted sources of

[22] Ofcom. (2025). Transmission Critical: The future of Public Service Media [Online]. Ofcom. Last Updated: 21 July 2025. Available at: https://www.ofcom.org.uk/siteassets/resources/documents/public-service-broadcasting/public-service-media-review/transmission-critical-the-future-of-public-service-media.pdf?v=400631 [Accessed 27 March 2026].

[23] Ofcom. (2025). News consumption in the UK: 2025. [Online]. Ofcom. Last Updated: 21 July 2025. Available at: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/adult-and-teen-news-consumption-survey/news-consumption-in-the-uk-2025-research-findings.pdf?v=400636 [Accessed 27 March 2026].

information. In this way, a public awareness campaign, as described in this position paper, could also be integrated into a broader media literacy campaign. Information about the amendments and the public's digital rights under them, therefore, could be used to complement the recommendations of the Department of Education's 'Curriculum and Assessment Review,' which calls for the role of media literacy to be strengthened in English and Citizenship classes, by offering resources teachers could refer students to to teach them to how to use the toggles, filters, and other potential systems to curate their news feed as they so choose on social media, as well as how to engage with different narratives in an increasingly polarised world.

Covering Micro, Local, Regional, and National News

In any discussion of reliable, factual media, it is essential to acknowledge the importance of funding and the challenges news providers face. Those reliant on advertising revenue have been hit particularly hard in recent years, with AA/WARC estimating that national newsbrand print revenues fell by 4.3% in 2024 — a drop of £32.9 million. While digital advertising in the sector grew by 2.5%, this was insufficient to cover the print losses, with the growth generating only £8.4 million in comparison. AA/WARC observed a similar trend for regional newsbrands, with print advertising revenues falling by £16.05 million while digital revenues grew by only £8.97 million.[24]

If these patterns continue, the country's media landscape, at the local, regional, and national levels, will face increasing financial strain, threatening the plurality of reliable news sources available in the future, which are essential to tackling the spread of hyper-localised forms of mis- and disinformation. A sustainable funding model, therefore, may have to be considered moving forward to prevent the spread of 'news desertification,' with the Public Interest News Foundation estimating that 4.4 million UK residents live in a "local news desert" where there is "no dedicated local news outlet" as of December 2025.[25]

It is vital, therefore, that any future review of the BBC license fee considers the importance and future circulation of micro, local, regional, and national news (together with commercial PSBs for national news). Given that the BBC already provides these services to a certain extent, the Department for Culture, Media and Sport should consider expanding the BBC's

[24] D. Ponsford. (2025). Google and Meta's vast estimated UK earnings revealed. [Online]. Press Gazette. Last Updated: 1 May 2025. Available at: https://pressgazette.co.uk/marketing/google-meta-advertising-market-share/ [Accessed 27 March 2026].
[25] Public Interest News Foundation. (2025). Local news launches, closures, and the state of play in 2025. [Online]. Public Interest News Foundation. Last Updated: 11 December 2025. Available at: https://www.publicinterestnews.org.uk/blog/local-news-launches-closures-and-the- [Accessed 27 March 2026].

responsibilities for covering micro, local, regional, and national news to counter the rise of mis- and disinformation.

Independent Disinformation Ombudsman

Most mis-and disinformation today takes place in digital spaces, which Ofcom regulates through tools such as the OSA. However, Ofcom is mandated by the OSA to monitor this threat only through the lens of systems and processes.

This is evident in the context of so-called "super-complaints". Ofcom plans to publish its guidance on super-complaints in February 2026, following the conclusion of its consultation. According to the watchdog's draft guidance published in September 2025, however, a super-complaint is a complaint made about an eligible organisation concerning systemic issues in regulated services that could cause "significant harm" to UK users and/or "significantly adversely" affect their right to freedom of expression. For a super-complaint to be eligible, it has to be made by an entity that represents users and/or members of the public, can act independently of regulated services, has contributed "significantly" to public discussion about online safety, and has due regard to Ofcom's super-complaints guidance. If a complaint is found to be admissible, Ofcom can take regulatory action, opening a policy project to consider new measures, or create another enforcement body to tackle the issue.

While welcome, this system does not account for individual complaints. Indeed, as Ofcom explicitly states, "we are not empowered to act on or investigate individual complaints or to instruct services to remove or reinstate specific content", which includes even if multiple individual complaints about a particular online service are lumped together. Instead, current guidance recommends that users raise individual complaints in the first instance with online services, which are "best placed to respond to individual complaints, and if appropriate, take action".

If that fails, Ofcom's current guidance recommends that they report it to the online complaints portal. However, when users access that portal, they are again explicitly told that the watchdog "can't respond to your complaint". This effectively leaves individuals with no recourse if the online services fail to act. This is a significant problem if those individuals are reporting examples of mis- and disinformation, and especially so if the falsehood specifically targets them.

Accordingly, there must be a tool to engage with individual complaints. An independent ombudsman centred around mis- and disinformation would facilitate this. The service

should enable UK users and victims to file complaints that the ombudsman can act on if deemed admissible. This would require a thorough eligibility and admissibility criteria, as well as a list of potential actions the ombudsman could take. These could be created through a rigorous consultation process, not dissimilar to the one that Ofcom used for super-complaints.

Tackling Fake Accounts

Fake accounts and bot farms play an outsized role in the dissemination of online mis- and disinformation. They are also a tool increasingly used by hostile states to sway elections and undermine cohesion in democratic countries. The Russian disinformation campaign Doppelgänger, which was reportedly established by two Russian companies, Strucura National Technologies and Social Design Agency, according to Meta, utilised tens of thousands of fake accounts to amplify spoofed websites masquerading as real publications (such as the German newspaper Der Spiegel), which were themselves producing content designed to reduce support for Ukraine, further incite Western political and social divisions, and erode trust in Western media and public institutions.[26]

Outside of state-run disinformation operations, bots account for an alarming proportion of online activity, with bots generating over half of the world's internet traffic in 2024, according to the 2025 Imperva Bad Bot Report. Thirty-seven per cent of internet traffic, meanwhile, was generated by so-called "bad bots" specifically designed for malicious activities and to avoid detection.[27] The problem is likewise pervasive on social media platforms, where automated accounts are estimated to be roughly 15% of total membership.[28]

Consequently, reducing the footprint of fake and bot accounts on social media is essential for mitigating the spread of online mis- and disinformation. Social media platforms and search engines, for example, could be encouraged to introduce a filter or toggle that allows users to block content from bot accounts. The filter could be powered through the same bot management systems currently used to identify bot-driven traffic and material, such as Cloudflare Bot Management and Imperva Bot Management. While false positives are a

[26] Polis Analysis. (2025). Doppelgänger: An analysis of Russia's most infamous disinformation campaign. [Online]. Polis Analysis. Last Updated: July 2025. Available at: https://www.polisanalysis.com/_files/ugd/394c34_fb1f8173396b4c819b61892b7e86ac19.pdf [Accessed 27 March 2026].
[27] Imperva. (2025). 2025 Bad Bot Report. [Online]. Imperva. Last Updated: 15 April 2025. Available at: https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/ [Accessed 27 March 2026].
[28] Cloudflare. (n.d.). What is a social media bot? | Social media bot definition. [Online]. Cloudflare. Last Updated: n.d. Available at: https://www.cloudflare.com/en-gb/learning/bots/what-is-a-social-media-bot/ [Accessed 27 March 2026].

concern, these systems maintain high accuracy in bot detection, with Cloudflare's system reporting a false positive rate of less than 0.01%.

AI Watermarking

AI-generated content is becoming increasingly life-like and difficult to differentiate from reality. This can be seen with the release of Sora 2 in September, which saw an explosion of realistic AI-generated videos featuring deceased celebrities. Content depicting Martin Luther King Jr., for example, became so widespread that OpenAI was forced to ban users from creating content featuring the US civil rights leader using its AI video app.[29]

The threat that this technology, as well as other similar AI-video generative applications, poses to democratic institutions is immense. Through these tools, malicious users can easily create deceptive, life-like content about politicians or public figures to mislead audiences, sow discord, and defame individuals. In the UK, this occurred even before the advent of advanced generative AI videos. London Mayor Sadiq Khan was the victim of AI disinformation in 2023 when fake audio clips of Khan supposedly dismissing Armistice Day were circulated online and among far-right groups.[30] That same year, Prime Minister Keir Starmer was also targeted by AI disinformation during the 2023 Labour Party conference, with deepfake audio clips posted online of the then leader of the opposition abusing staffers and criticising the city of Liverpool.[31]

AI-generated videos have compounded the problem further by introducing a visual element. Notably, in October 2025, Conservative MP George Freeman was forced to report a deepfake video to the police. The footage depicted Freeman announcing that he was switching party allegiance from the Conservatives to Reform UK. The realism of the video meant that the MP had to issue a public statement clarifying that he had not, and had no intention, of joining Reform UK.[32]

Similar examples of political disinformation, driven by generative AI, are likely to become more prevalent, especially given the "significant threat" the UK faces from espionage and

[29] OpenAI. (2025). Statement from OpenAI and King Estate, Inc.. [Online]. X. Last Updated: 17 October 2025. Available at: https://x.com/OpenAINewsroom/status/1979005850166648933 [Accessed 27 March 2026].
[30] M. Spring. (2024). Sadiq Khan says fake AI audio of him nearly led to serious disorder. [Online]. BBC News. Last Updated: 13 February 2024. Available at: https://www.bbc.co.uk/news/uk-68146053 [Accessed 27 March 2026].
[31] Sky News. (2023). Deepfake audio of Sir Keir Starmer released on first day of Labour conference. [Online]. Sky News. Last Updated: 9 October 2023. Available at: https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir- [Accessed 27 March 2026].
[32] O. Sennitt. (2025). Tory MP reports deepfake defection video to police. [Online]. BBC News. Last Updated: 18 October 2025. Available at: https://www.bbc.co.uk/news/articles/c62e7xz02dpo [Accessed 27 March 2026].

foreign interference campaigns, many of which deploy disinformation as a tool, as highlighted by the National Protective Security Authority in its guidance.[33] It is critical, therefore, to promote a solution that helps ordinary people distinguish between what is real and what is fabricated. The most effective way to do so quickly would be to make pre-existing industry best practices legal requirements by requiring AI watermarking for AI companies. For responsible AI entities, this task would be no different from what they are already doing. Videos created by Sora 2, for example, already carry a clearly identifiable AI watermark and embedded C2PA metadata.

The government must also be willing to crack down on the readily available free online tools that remove AI watermarks. A simple Google search for an "AI watermark remover" will yield dozens of results from Unwatermark AI to ChatGPT Watermark Remover, which allow users to remove AI watermarks — both visible and invisible — from their creations. These tools pose a clear challenge to information integrity if users wish to avoid their AI-generated content carrying identifiers that reveal its origin, making it appear, for all intents and purposes, legitimate to the public.

A stopgap measure in the meantime could include age-restricting the use of LLMs, such as ChatGPT, Gemini, or Gork. However, similar actions under the OSA regarding pornography websites have been circumvented by users through Virtual Private Networks (VPNs).

[33] National Protective Security Authority. Protecting our Democratic Institutions: Countering Espionage and Foreign Interference. [Online]. National Protective Security Authority. Last Updated: n.d. Available at: https://www.npsa.gov.uk/national-security-act/defending-democracy/countering-espionage-and-foreign-i [Accessed 27 March 2026].

Summary

Together, these policies would help protect the UK from AI-powered online threats to information integrity and serve as the basis of a multifaceted approach encompassing media, education, legislation, and regulatory policies.

Policies compelling social media platforms and tech companies to develop toggles and filter systems for factual accuracy and PSB content will empower users, not politicians or tech CEOs, to make informed choices about the content they wish to consume, while encouraging them to use more reliable sources of information.

Similarly, legislation requiring these same organisations to create the technical solutions and algorithmic systems to ensure impartiality will help counter the harms of filter bubbles, which leave users more susceptible to the influence of mis- and disinformation, by introducing balance in algorithms and social media feeds.

Elsewhere, an independent ombudsman for mis- and disinformation would enable UK users to seek redress for potentially harmful instances of such misinformation. This could prove particularly helpful in countering financial disinformation, where false information ranging from fake reviews to deceptive social media reporting causes $78 billion in economic damage every year, harming companies and destabilising markets.

Tackling fake accounts and bots is also vital, given their outsized role in spreading mis- and disinformation, as well as their use in state-run disinformation campaigns. A toggle feature would empower users to decide whether to engage with bot-driven content, without restricting their choice.

Meanwhile, stricter AI watermarking and action against free online tools that allow users to circumvent them will help address the mass production of realistic AI-generated videos and deepfakes by clearly identifying them as fabricated. Politically, this is crucial, especially in light of the deepfake made about George Freeman, to prevent the proliferation of life-like AI content that could influence the next election.

Bibliography

A. Croft. (2025). US ‘terminates agreements with Europe’ to help fight disinformation from hostile states. [Online]. Independent. Last Updated: 9 September 2025. Available at: https://www.independent.co.uk/news/world/americas/us-europe-russia-disinformation-memorandum-b282314 [Accessed 27 March 2026].

B. Swire-Thompson, et al. (2021). Correction format has a limited role when debunking misinformation. [Online]. PubMed Central. Last Updated: 29 December 2021. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8715407/ [Accessed 27 March 2026].

Cloudflare. (n.d.). What is a social media bot? | Social media bot definition. [Online]. Cloudflare. Last Updated: n.d. Available at: https://www.cloudflare.com/en-gb/learning/bots/what-is-a-social-media-bot/ [Accessed 27 March 2026].

C. Vallance. (2023). Disinformation most active on X, formerly known as Twitter, EU says. [Online]. BBC News. Last Updated: 26 September 2023. Available at: https://www.bbc.co.uk/news/technology-66926080 [Accessed 27 March 2026].

D. Ponsford. (2025). Google and Meta’s vast estimated UK earnings revealed. [Online]. Press Gazette. Last Updated: 1 May 2025. Available at: https://pressgazette.co.uk/marketing/google-meta-advertising-market-share/ [Accessed 27 March 2026].

D. Trump. (2025). "As the President of the United States, I will stand up to Countries that attack our incredible American Tech Companies". [Online]. Truth Social. Last Updated: 26 August 2025. Available at: https://truthsocial.com/@realDonaldTrump/posts/115092243259973570 [Accessed 27 March 2026].

E. Yeomans and C. Alt. (2026). Elon Musk’s AI undressing tool on Grok could be banned. [Online]. The Times. Last Updated: 7 January 2026. Available at: https://www.thetimes.com/uk/technology-uk/article/grok-ai-images-elon-musk-news-p6v2kfmfv [Accessed 27 March 2026].

Imperva. (2025). 2025 Bad Bot Report. [Online]. Imperva. Last Updated: 15 April 2025. Available at: https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/ [Accessed 27 March 2026].

Instagram. (2025). Introducing a new way to see and control your algorithm. [Online]. Instagram. Last Updated: 10 December 2025. Available at: https://about.instagram.com/blog/announcements/reels-algorithm-control [Accessed 27 March 2026].

Joint Research Centre. (2024). Misinformation and disinformation: both prebunking and debunking work for fighting it. [Online]. European Commission. Last Updated: 25 October 2024. Available at: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/misinformation-and-disinformation-bo [Accessed 27 March 2026].

Joint Committee on the National Security Strategy. (2025). Oral evidence: Defending democracy. [Online]. UK Parliament. Last Updated: 17 March 2025. Available at: https://committees.parliament.uk/oralevidence/15590/pdf/ [Accessed 27 March 2026].

Joint Committee on the National Security Strategy. (2024). Oral evidence: Defending democracy. [Online]. UK Parliament. Last Updated: 18 March 2024. Available at: https://committees.parliament.uk/oralevidence/14513/pdf/ [Accessed 27 March 2026].

J. Kaplan. (2025). More Speech and Fewer Mistakes. [Online]. Meta. Last Updated: 7 January 2025. Available at: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ [Accessed 27 March 2026].

J. Nevett. (2025). Trump suggests Starmer could use military to control UK borders. [Online]. BBC News. Last Updated: 18 September 2025. Available at: https://www.bbc.co.uk/news/articles/cpd91wjypj9o [Accessed 27 March 2026].

K. Hays. (2023). New layoffs at Twitter hit trust and safety workers even as advertisers worry about toxic speech. [Online]. Business Insider. Last Updated: 19 September 2023. Available at: https://www.businessinsider.com/new-twitter-layoffs-hit-trust-and-safety-workers-again-2023-9 [Accessed 27 March 2026].

L. Kendall. (2026). Secretary of State statement to the House of Commons: 12 January 2026. [Online]. GOV.uk. Last Updated: 12 January 2026. Available at:

https://www.gov.uk/government/speeches/secretary-of-state-statement-to-the-house-of-commons-12-janua [Accessed 27 March 2026].

M. Rubio. (2025). Protecting and Championing Free Speech at the State Department. [Online]. US Department of State. Last Updated: 16 April 2025. Available at: https://www.state.gov/protecting-and-championing-free-speech-at-the-state-department [Accessed 27 March 2026].

M. Spring. (2024). Sadiq Khan says fake AI audio of him nearly led to serious disorder. [Online]. BBC News. Last Updated: 13 February 2024. Available at: https://www.bbc.co.uk/news/uk-68146053 [Accessed 27 March 2026].

National Protective Security Authority. Protecting our Democratic Institutions: Countering Espionage and Foreign Interference. [Online]. National Protective Security Authority. Last Updated: n.d. Available at: https://www.npsa.gov.uk/national-security-act/defending-democracy/countering-espionage-and-foreign-i [Accessed 27 March 2026].

Ofcom. (2024). Four in 10 UK adults encounter misinformation. [Online]. Ofcom. Last Updated: 27 November 2024. Available at: https://www.ofcom.org.uk/media-use-and-attitudes/attitudes-to-news/four-in-10-uk-adults-encounter-mi [Accessed 27 March 2026].

Ofcom. (2025). Transmission Critical: The future of Public Service Media [Online]. Ofcom. Last Updated: 21 July 2025. Available at: https://www.ofcom.org.uk/siteassets/resources/documents/public-service-broadcasting/public-service-media-review/transmission-critical-the-future-of-public-service-media.pdf?v=400631 [Accessed 27 March 2026].

Ofcom. (2025). News consumption in the UK: 2025. [Online]. Ofcom. Last Updated: 21 July 2025. Available at: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/adult-and-teen-news-consumption-survey/news-consumption-in-the-uk-2025-research-findings.pdf?v=400636 [Accessed 27 March 2026].

Office of the Director of National Intelligence. (2025). Annual Threat Assessment of the U.S. Intelligence Community. [Online]. Office of the Director of National Intelligence. Last Updated: 25 March 2025. Available at: https://www.dni.gov/files/ODNI/documents/assessments/ATA-2025-Unclassified-Report.pdf [Accessed 27 March 2026].

OpenAI. (2025). Statement from OpenAI and King Estate, Inc.. [Online]. X. Last Updated: 17 October 2025. Available at: https://x.com/OpenAINewsroom/status/1979005850166648933 [Accessed 27 March 2026].

O. Sennitt. (2025). Tory MP reports deepfake defection video to police. [Online]. BBC News. Last Updated: 18 October 2025. Available at: https://www.bbc.co.uk/news/articles/c62e7xz02dpo [Accessed 27 March 2026].

Polis Analysis. (2025). Doppelgänger: An analysis of Russia's most infamous disinformation campaign. [Online]. Polis Analysis. Last Updated: July 2025. Available at: https://www.polisanalysis.com/_files/ugd/394c34_fb1f8173396b4c819b61892b7e86ac19.pdf [Accessed 27 March 2026].

Public Interest News Foundation. (2025). Local news launches, closures, and the state of play in 2025. [Online]. Public Interest News Foundation. Last Updated: 11 December 2025. Available at: https://www.publicinterestnews.org.uk/blog/local-news-launches-closures-and-the- [Accessed 27 March 2026].

R. Fletcher. (2020). The truth behind filter bubbles: Bursting some myths. [Online]. Reuters Institute. Last Updated: 24 January 2020. Available at: https://reutersinstitute.politics.ox.ac.uk/news/truth-behind-filter-bubbles-bursting-some-myths [Accessed 27 March 2026].

Sky News. (2023). Deepfake audio of Sir Keir Starmer released on first day of Labour conference. [Online]. Sky News. Last Updated: 9 October 2023. Available at: https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir- [Accessed 27 March 2026].

The Reuters Institute for the Study of Journalism. (2025). Overview and key findings of the 2025 Digital News Report. [Online]. Reuters Institute. Last Updated: 17 June 2025. Available at: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary [Accessed 27 March 2026].

The White House. (2025). Restoring Freedom of Speech and Ending Federal Censorship. [Online]. The White House. Last Updated: 20 January 2025. Available at: https://www.whitehouse.gov/presidential-actions/2025/01/restoring-freedom-of-speech-and-ending-feder [Accessed 27 March 2026].

US Department of State. (2025). 2024 Country Reports on Human Rights Practices: United Kingdom. [Online]. US Department of State. Last Updated: 12 August 2025. Available at: https://www.state.gov/reports/2024-country-reports-on-human-rights-practices/united-kingdom [Accessed 27 March 2026].

World Economic Forum. (2026). Global Risks Report 2026. [Online]. World Economic Forum. Last Updated: 4 January 2026. Available at: https://www.weforum.org/publications/global-risks-report-2026/digest/ [Accessed 27 March 2026].



Council for Countering Online Disinformation

Website: https://ccod.co.uk/