

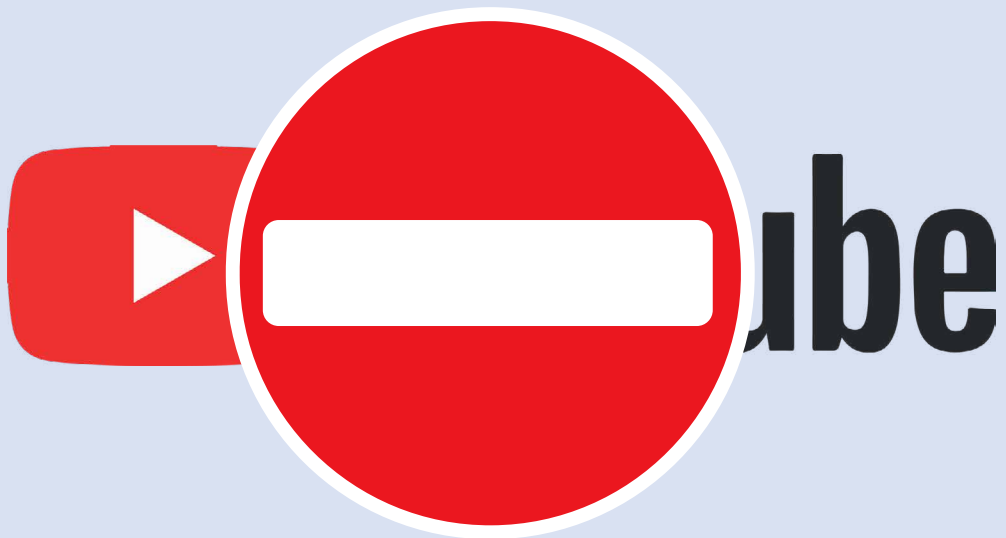
Free Speech Union briefing

Misinformation about Misinformation

Why the British Government's case for unprecedented
online censorship rests on sand

By David Rose

September 2026



Contents

Executive Summary	3
1. Introduction: towards a new machinery for restricting online speech	7
2. The Southport murders: a gift to the censorship lobby	11
3. Nationalising internet censorship	15
4. What is misinformation?	19
5. Policy-based evidence making: the GO-Science review	24
6. The missing counter-arguments	28
7. The algorithm panic	31
8. Exaggerating the impact of misinformation	34
9. Promoting “trustworthy” sources: the case of the BBC	38
10. Back to the future: restoring Charles I’s censorship regime	43

Executive Summary

The British Government is preparing an apparatus of internet censorship without precedent in the modern era. The DCMS Green Paper *Watch this Space* proposes that the law should force social media algorithms to give prominence to news sources the state deems “trustworthy”. Lucy Powell, Labour’s Deputy Leader, wants platforms placed under broadcast-style duties of “accuracy and balance” during election campaigns. The Government Office for Science’s review of misinformation recommends statutory “quality signals” in content ranking, a state-linked database of “false and misleading narratives”, and official “prebunking” and “debunking” across education and public communications. Ofcom holds a crisis protocol in readiness for the next emergency. And under the new Prime Minister, responsibility for “information resilience” and online harms has been concentrated in a renamed DCMS, whose Secretary of State has quit X because, she says, it “favours abuse and misinformation over meaningful debate”.

These proposals did not fall from the sky. They are the demands of a lobby: a small, interconnected ecosystem of campaign groups, consultancies, self-appointed fact-checkers, former broadcasters and parliamentarians, whose recommendations are always more censorship. Its ties to government are close. The Centre for Countering Digital Hate was co-founded by Morgan McSweeney, later Sir Keir Starmer’s Chief of Staff; after Southport, it chaired an “emergency” roundtable with officials from DSIT, the Home Office, Ofcom and the Metropolitan Police. The Global Disinformation Index has been funded by the Foreign Office. And when the Government commissioned a review to prove the problem was real, members of the same lobby served as its “expert” reviewers, marking their own homework.

The case for control rests on two claims, endlessly repeated. The first is that mis- and disinformation now saturate the internet and inflict grave social and political harm: riots, vaccine refusal, subverted elections – in the GO-Science review’s words, a “systemic national risk”. The second is that platform algorithms are the engine of the crisis, amplifying hatred and division because outrage maximises engagement, and engagement maximises revenue.

On close inspection, the evidential base for both claims fails. The GO-Science review – commissioned to supply the “full evidence base” ministers admitted they lacked – is itself a case study in misinformation. Its authors made no attempt to establish what is true or false. Instead they relied on AI-generated summaries of 228 papers they had not read, and on monthly questionnaires returned to the European Digital Media Observatory by fact-checkers: a digest of what fact-checkers chose to check, which says nothing about reach, impact or belief. Most of the GO-Science authors and reviewers lacked any scientific grounding in the fields the report adjudicates – no climate scientist, no economist, no clinician, no epidemiologist. And it ignores a substantial body of peer-reviewed research, published in the world’s leading journals, that contradicts its every premise. The FSU can reveal the review cost taxpayers almost half a million pounds. Among this briefing’s findings:

- The review’s sweeping claim that platform algorithms are “exclusively profit- and engagement-driven” rests on a single citation: a paper from Universitas Muhammadiyah Buton, a small private institution in the province of Southeast Sulawesi ranked 433rd in Indonesia and 10,370th in the world, which was not original research but an AI-assisted synthesis of earlier work.
- A 2024 study in *Science* found that misinformation flagged by fact-checkers attracted just 0.3% of 2.7 billion vaccine-related views on Facebook during the rollout. Unflagged “vaccine-skeptical” content containing no falsehoods had 46 times the aggregate impact on vaccination intentions. The most influential single item was an accurate *Chicago Tribune* article – ‘A healthy doctor died two weeks after getting a COVID vaccine; CDC is investigating why’ – seen by more than 50 million people, over six times the views of all flagged misinformation combined. In other words, the content that caused genuine vaccine hesitancy came from exactly the kind of “trustworthy” mainstream source the Green Paper would force into every feed. A 2024 synthesis in *Nature* concluded that exposure to misinformation is far smaller than supposed, and concentrated in a narrow fringe of users who actively seek it out. Visits to 490 “untrustworthy” websites made up 5.9% of Americans’ clicks on news sites; “fake news” accounts for roughly 0.15% of the average American’s daily media diet.

- Twitter’s own researchers, studying millions of accounts, found no evidence that its algorithm favoured the far left or the far right over moderate voices. A landmark Facebook experiment found the algorithm served less untrustworthy content than a simple chronological feed.
- A study of 46 conspiracist beliefs, tracked in national surveys since 1966, found no evidence that belief in conspiracy theories is rising: significant declines outnumbered increases.
- The Global Disinformation Index pronounced “clearly untrue” a vaccine statistic drawn from Public Health England’s own data, and set about persuading advertisers to defund the publisher. Its chief executive has explained that content can be “factually accurate but still extremely harmful”. That is not fact-checking. It is the policing of unwelcome truths.

Ofcom’s Media Nations 2026 report found that only 33% of adults associate the BBC – the state’s exemplar of trustworthiness – with impartiality, and 45% with trustworthiness. Recent events, such as the doctoring of a Donald Trump speech by *Panorama* and Ofcom’s critical ruling on a Gaza documentary, suggest the public may have a point.

Three conclusions follow.

First, Britain has seen nothing like this since Charles I. In July 1637 the Star Chamber decreed that nothing be published unless licensed by state-approved authorities, with enforcement delegated to a private company. The proposals now on the table – exalted legal status for approved publishers, emergency powers over content, official programmes to teach children and adults what to believe – are that regime’s closest peacetime echo in nearly four centuries. The last vestiges of press licensing lapsed in 1695. It was against that system that Milton wrote *Areopagitica*, and his answer still stands: let truth and falsehood grapple, for truth’s confuting is the best and surest suppressing.

Second, “misinformation” is being redefined as dissent. For example, for the CCDH, the GDI and the GO-Science review, it is “misinformation” to “decouple” net zero from climate science – that is, to treat the pace and cost of decarbonisation as matters of economic and political judgement rather than physics. It is “misinformation” to question the efficiency of heat pumps or the range and resale value of electric vehicles. A label that purports to denote falsehood is being stretched to cover accurate statements and legitimate opinions that officialdom finds unhelpful. Whoever controls the label controls the debate.

Third, this briefing has tested the lobby’s arguments – above all those of the GO-Science review – against primary sources and the peer-reviewed literature, and found many of them wanting: exaggerated, unevidenced, or flatly contradicted by research the review never mentions. Repetition has given these claims currency. It has not made them true, and they have become a classic example of a moral panic. The case for the most far-reaching apparatus of peacetime speech control in modern British history rests, on the evidence, on misinformation about misinformation.

1. Introduction: towards a new machinery for restricting online speech

The British Government appears to be moving rapidly towards imposing unprecedented state control of the content published on social media platforms. The direction of travel has been visible for at least a decade, but several developments in the summer of 2026 suggest that legislation may be imminent. The first and most significant was the Green Paper *Watch this Space*, published by the Department for Culture, Media and Sport in June, which suggests that the algorithms governing social media and video-sharing platforms should be forced by law to give prominence to public service outlets deemed “trustworthy” by the state, such as the BBC.¹ The second was a proposal by Lucy Powell, Labour’s Deputy Leader and Education Secretary, that the Representation of the People Bill now before Parliament be amended so that platforms come under “some of the same requirements” to observe accuracy and balance as licensed broadcasters during election periods.² Finally, shortly after Andy Burnham replaced Sir Keir Starmer as Prime Minister in July, the Department for Science, Innovation and Technology (DSIT) was abolished and its responsibilities for “information resilience” and online harms transferred to a renamed Department for Digital, Culture, Media and Sport (DCMS) under Secretary of State Lisa Nandy. This change, the Government said, will bring “coherence to the online safety agenda” by concentrating it under one Whitehall roof.³

1 Department for Culture, Media and Sport, *Watch this Space: A New Strategic Direction for UK Media*, Green Paper, June 2026, <https://www.gov.uk/government/consultations/watch-this-space-a-new-strategic-direction-for-uk-media-green-paper-and-public-consultation>

2 Lucy Powell MP, interviews on LBC (*Sunday with Lewis Goodall*) and Times Radio, 5 July 2026.

3 Written Ministerial Statement by Baroness Smith of Basildon on behalf of the Prime Minister, 21 July 2026, <https://questions-statements.parliament.uk/written-statements/detail/2026-07-21/hlws298>.

As well as regulating the internet, Nandy's department is also in charge of the national Media Literacy Action Plan and Ofcom, whose *Making Sense of Media* scheme is supposed to help schools teach their students how not to be misled. Meanwhile, to Labour's left, the Green Party leader Zack Polanski urged the Government to "sort out the press" by creating a mandatory, statutory regulator. Only this could stem the "constant misinformation" the print media pumps out to protect the fossil fuel firms that were "destroying lives and our planet".⁴ Two incidents triggered Polanski's demands: an innocuous column in the *Mail* that urged Britons to enjoy the warm summer, despite the heatwaves and drought; and the death of the former Cambridge professor Jason Arday, who resigned after journalists accused him both of plagiarism and a series of fabrications about his own life.⁵

"Misinformation" has been advanced many times as a reason for new forms of censorship. Their advocates assert that the internet is awash with both mis- and disinformation. By the first, they mean false information shared without an intent to deceive; by the second, falsehoods disseminated through a malicious desire to deceive or manipulate. They insist that the prevalence of both is a product of content-hungry algorithms, which drive up platforms' revenue by engaging their users' emotions and fostering hatred and social division. In the words of the DCMS Green Paper, "To safeguard an informed democracy and cohesive society, we must tackle these threats with interventions to support a media ecosystem that is independent, accountable, plural and, crucially, trustworthy."

Underpinning this agenda is what purports to be a fresh and objective evidence base: a 136-page review of "false and misleading information" commissioned by the Government Office for Science, which was published in May 2026. This document, whose methods and conclusions are examined in this briefing, claims the problem of fake news is now so serious that it constitutes "a systemic national risk".⁶ Powell echoed its arguments when she issued her call to regulate platforms at election time, saying that the biggest influence over what voters see is not TV news but

4 <https://www.gbnews.com/politics/jason-arday-zack-polanski-media-culture-cambridge-professor-death>

5 <https://www.theguardian.com/uk-news/2026/aug/06/rise-and-fall-of-jason-arday>

6 Baird, H. M., Zhang, H., Uttley, L., Bontcheva, K., Webb, T. L., Wright, J. & Bauld, L., 'False and Misleading Information: An Evidence Summary and Map for Policy and Practice' (Government Office for Science, May 2026), <https://usher.ed.ac.uk/sites/default/files/2026-05/False%20and%20Misleading%20Information%3A%20An%20Evidence%20Summary%20and%20Map%20for%20Policy%20and%20Practice%20pdf>

“social media feeds decided by opaque algorithms”, through which, she claimed, deepfake images and deliberately coordinated disinformation are being spread at alarming speed.⁷ That is why platforms’ content has to be “checked for its accuracy and balance”.⁸ She did not say who would do the checking, or whether there would be any right of appeal if the checkers recommend content removal. In any event, the effect would be that citizens could not post what they might say over a drink with friends on Facebook, TikTok or X.

Meanwhile Nandy, the minister whose department holds the ring, has made her own views very public. On 2 July she announced that she and all her staff were quitting X, because it “favours abuse and misinformation over meaningful debate” and was “not healthy” for democracy.⁹

Worryingly enough for free speech campaigners such as the FSU, her approach seems to be shared, at least in part, by the Conservative Party. It was the Conservatives who were in office when Parliament passed the Online Safety Act 2023 (OSA), arguably the most far-reaching speech law in modern British history, which grants the government regulator Ofcom sweeping powers to fine platforms up to 10% of their annual worldwide turnover for failing to prevent the dissemination of “unlawful” content. (The Act does contain measures, which are still not in force, to protect journalistic content and speech of democratic importance. But it is a blunt and imprecise instrument, obliging platforms to deny access to “harmful” content to those under the age of 18. Numerous instances have come to light of posts being blocked for people who have not verified their age on issues such as the wars in Gaza and Ukraine¹⁰ – even as the Government prepares to pass a new Representation of the People Bill that will reduce the voting age to 16.)

For many, however, the OSA is not sufficiently draconian. In July 2025, the Commons Science, Innovation and Technology Select Committee issued a blistering attack, saying in a report that the OSA “fails to keep UK citizens safe from a core and pervasive online harm”. Its provisions to deal with misinformation “fall short”, it went on, for the Act “does not protect users from the commodification of synthetic mis/disinformation”,

⁷ Lucy Powell MP, LBC, 5 July 2026.

⁸ Lucy Powell MP, Times Radio, 5 July 2026.

⁹ <https://www.bbc.com/news/articles/clyrx1ee2r4o>

¹⁰ <https://www.newarab.com/news/gaza-posts-blocked-social-media-under-uk-online-safety-act>

and thanks to the advent of generative AI, was “already out of date”. The committee also demanded that action be taken not only against content that was *per se* unlawful, such as hate speech or incitement to commit acts of violence, but posts that were “legal but harmful”, that is to say, speech that Parliament has declined to prohibit.¹¹ Platforms, it said, must be “held responsible for the algorithmic spread of misleading or deceptive content” and “amplifying” harmful content. Throughout the report, misinformation is described as being as part of the same “harmful” category as “content that is hateful, extreme or dangerous”.

Rhetoric of this kind is not new. It entered British public life with the EU referendum in 2016, with the claim – later refuted by both academic studies and the Information Commissioner – that Facebook users’ data harvested by the firm Cambridge Analytica had been used to distribute Brexit propaganda and so influenced the referendum result.¹² In November 2017, the then-Prime Minister Theresa May accused Russia of systematically planting lies on the internet to “weaponise information” and sow discord,¹³ and in February 2019 another select committee found British elections were being menaced by malign forces, pumping out disinformation “to distort, to disrupt, and to destabilise”.¹⁴

Thus the pattern was set. This briefing will argue that much of the evidence adduced to support such calls for restrictions is questionable, and some has been refuted by academic research: an alarmingly weak foundation on which to build what would arguably be the strictest regime of censorship seen in peacetime Britain since the 17th century. More than 50 years ago, the sociologist Stanley Cohen identified the elements of what he called “moral panics”: unfounded claims by politicians, the media and other influential figures about perceived threats to society and its values that feed on each other, leading to disproportionate measures that cause far more damage than good. The current fever over misinformation is a classic example of the genre.¹⁵

11 Science, Innovation and Technology Committee, *Social media, misinformation and harmful algorithms*, HC 441 (11 July 2025), <https://publications.parliament.uk/pa/cm5901/cmsselect/cmsctech/441/report.html>

12 <https://www.bbc.com/news/uk-politics-54457407>

13 Digital, Culture, Media and Sport Committee, *Disinformation and ‘fake news’: Final Report*, HC 1791 (February 2019), para 240; for the report itself see note 14 below.

14 Digital, Culture, Media and Sport Committee, *Disinformation and ‘fake news’: Final Report*, HC 1791 (February 2019), <https://publications.parliament.uk/pa/cm201719/cmsselect/cmcmums/1791/1791.pdf>

15 Cohen, S: *Folk Devils and Moral Panics*, 1st edn. MacGibbon & Kee 1972, updated 3rd edn 2011, <https://www.taylorfrancis.com/books/mono/10.4324/9780203828250/folk-devils-moral-panics-stanley-cohen>

2. The Southport murders: a gift to the censorship lobby

The proposals described above did not fall from the sky. They emerged from a lobby: a small, interconnected ecosystem of campaign groups, consultancies, self-appointed fact-checkers, former broadcasters and parliamentarians. Its *raison d'être* is the identification of false information and online harms, and its recommendations are always more censorship.

For the lobby, the disturbances that followed the murders of three little girls at a summer holiday dance class in Southport in 2024 were a perfect gift, and helped create the climate we face now. Within hours, the killer, who turned out to be the pathologically violent but UK-born teenager Axel Rudakubana, was being wrongly described online as a recent Muslim migrant. The claim prompted Home Secretary Yvette Cooper to declare that social media platforms had put “rocket boosters” under falsehoods. The then-Greater Manchester mayor, the new Prime Minister Andy Burnham, made similar statements, warning against “mob rule” fed by “half-truths” on social media.¹⁶ Ofcom wrote publicly to the platform owners, later observing that the riots showed what “algorithmic recommendations” could do in a crisis.¹⁷ The Science, Innovation and Technology Select Committee later said misleading claims about Southport had 155 million impressions on X, and had a “potential reach of 1.7 billion people”.¹⁸

A few weeks later, the Centre for Countering Digital Hate (CCDH), a campaign group co-founded by Imran Ahmed, a former adviser to Labour shadow ministers, and Morgan McSweeney, Sir Keir Starmer’s former Chief of Staff, chaired an “emergency” roundtable attended by officials from DSIT, the Home Office, Ofcom, the Metropolitan Police counter-

¹⁶ <https://www.bbc.com/news/articles/c3gvq7vyx9do>

¹⁷ <https://techcrunch.com/2024/08/07/uks-internet-regulator-warns-social-media-platforms-over-risks-of-content-inciting-violence>

¹⁸ Science, Innovation and Technology Committee, *ibid* note 11.

terrorism internet unit and current and former MPs. Ahmed had been the first witness to give evidence when a parliamentary committee began to scrutinise what would become the OSA. Now the roundtable he convened recommended it be amended to grant the state emergency powers to combat the spread of misinformation if it was threatening national security or public health and safety.¹⁹ His close friend Josh Simons, a newly-elected but highly influential Labour MP, who before the 2024 general election had been running the Starmerite think tank Labour Together in which more than half the new Cabinet had senior roles, agreed. According to Simons, tech bosses could not be allowed to “wield power far exceeding that of any newspaper editor” without bearing commensurate responsibilities.²⁰ Simons, it will be recalled, was eventually appointed a minister, but in 2026 resigned, first from his Cabinet Office post and then from Parliament, after it emerged that he had commissioned an investigation that falsely accused journalists who revealed undeclared payments to Labour Together of being Russian agents – and then passed on these false allegations to the security services.²¹

Others who joined the calls for tougher restrictions included the campaign group Hope not Hate. It said the Southport riots showed the OSA was “not fit for purpose” when it came to clamping down on “lawful but harmful” misinformation.²² Happily for Hope not Hate, it has just acquired a hotline to Burnham’s Downing Street, where its former campaigns chief Matthew McGregor has been appointed Director of Political Strategy.

Hard cases make bad law, but in the first half of 2026 the censorship lobby began to get its way. In March, the Government tabled amendments to pending legislation to award itself new Henry VIII powers to censor internet content without needing to obtain consent from Parliament. The first, an amendment to the now-enacted Crime and Policing Act 2026, amends the OSA to allow the blocking of AI-generated content in the name of “minimising or mitigating the risks of harm to individuals”; the second, an amendment to what has become the Children’s Wellbeing and Schools Act, went even further, granting ministers the ability to alter any piece of legislation to protect children and restrict their access to

¹⁹ <https://freespeechunion.org/archive/ofcom-needs-emergency-powers-to-tackle-misinformation-claims-well-connected-left-wing-lobby-group>

²⁰ <https://www.malaymail.com/news/world/2024/08/07/uk-mps-call-for-social-media-platforms-to-match-newspaper-and-tv-standards-after-disinformation-fuels-worst-riots-in-13-years/146243>

²¹ <https://freespeechunion.org/publications/starmer-misinformation-mafia>

²² <https://fullfact.org/technology/government-learned-lessons-from-southport-riots/>

“certain internet services”.²³ Then, in June, Ofcom announced it was requiring platforms to draw up new protocols to reduce online “illegal activity”, such as “inciting racial or religious hatred, making threats, or inciting violence”²⁴ during a time of crisis, defined as an “extraordinary situation in which there is a serious threat to public safety in the United Kingdom”. Afterwards, the CCDH claimed credit for all these changes, saying they were the fruit of its “years of evidence and advocacy” and its 2024 roundtable.²⁵

Later that month, rioting broke out in Belfast when a Sudanese migrant was filmed allegedly attempting to murder a local man by knifing him repeatedly. By no means could misinformation be blamed for causing the subsequent disturbances: they were triggered by a phone camera video of the incident, which went viral for entirely understandable reasons. But Michelle O’Neill, the Sinn Féin First Minister of Northern Ireland, held “the Elon Musks of this world” responsible, saying that while Belfast burned, they were “sitting right comfy in their homes, orchestrating hate and tension”. Musk, the owner of X, pushed back, saying that an apparent attempt to behead a man in the street was “what’s making people angry, not social media”. For his part, Starmer did not criticise Musk by name, although he did say there were people “trying to interfere in our democracy and seeking to stir up division on our streets”. But in case anyone had doubts as to who those people might be, his spokesperson briefed reporters that the Government “will not tolerate platforms being used to spread violence”.²⁶

Meanwhile, Hope not Hate and the CCDH had acquired competition. The misinformation lobby’s newest member, the Council for Countering Online Disinformation (CCOD), founded in 2025, states its “mission” is to “protect the OSA”, and to ensure that platforms are punished with colossal fines if they “fail to take down or filter out illegal and harmful content”. Its board includes the former Tory minister John Penrose,²⁷ and one of its staff,

23 <https://www.politico.eu/article/uk-eyes-sweeping-powers-to-regulate-tech-without-parliamentary-scrutiny/>

24 <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-crisis-response-protocol>

25 Centre for Countering Digital Hate, “UK government strengthens Online Safety Act after years of CCDH evidence on public safety crises”. <https://counterhate.com/blog/uk-government-strengthens-online-safety-act-after-years-of-ccdh-evidence-on-public-safety-crises/>

26 <https://www.nbcnews.com/world/united-kingdom/belfast-riots-elon-musk-anti-immigrant-violence-stabbing-rcna349384>

27 <https://ccod.co.uk/about>

Joshua Tyler, told me that he and his colleagues prided themselves on their close contacts with MPs.

Other members of the censorship lobby are likely to play important roles as restrictive measures evolve. One is Rebecca Skippage, who founded the BBC's "disinformation team" in 2018 and served as its grandly-titled Global Disinformation Editor until 2025. Now a private consultant, she says she has advised the government, the EU, the Police Foundation and the Army, and moderates NATO StratCom sessions on the "cognitive battlefield". She has publicly called for a counter-disinformation body "at the centre of government", whose assessments should flow simultaneously to the Home Office, Ofcom and the Department for Education.²⁸

Skippage was recently interviewed by Amil Khan of Valent Projects for his Substack blog.²⁹ Khan, a former journalist, claims Ariadne, his firm's AI-powered software, can detect "false narratives" online,³⁰ predict which ones will become threats, and generate reports to support takedown requests. According to Valent's website, Ariadne gives private and public bodies "the intelligence to separate manufactured narratives from real public concern", and hence "the clarity to respond". Its disclosed clients include the Foreign Office and the Mayor of London, and in the interview Khan mentions Ariadne at the very moment Skippage tells him why a product of this kind is so necessary. The CCOD appears to agree: its published "manifesto" says one of its goals is to "raise the profile" of companies offering "AI-led tools for content moderation" – specifically, Tyler told me, Valent.

Yet despite its scope, the emerging policy on online speech still rests on a small set of factual claims, often repeated but seldom challenged. The first is that misinformation on social media is driving support for extreme positions, destroying social cohesion and damaging democracy. Second, that this derives from platforms' secret algorithms, which promote inflammatory falsehoods because they maximise engagement and, therefore, revenue. Third, that the danger all this poses is so great that the state must step in, even at an acknowledged cost to freedom of expression. All require close scrutiny.

28 <https://www.linkedin.com/in/rebeccaskippage/>

29 <https://substack.com/@amilkhan/p-198575442>

30 <https://www.valent.co.uk/ariadne>

3. Nationalising internet censorship

An unspoken premise of the third of these arguments is that online speech is at present totally ungoverned, so that the choice is between state regulation and feral chaos. In a Substack essay published in August 2026,³¹ the pseudonymous writer Arctotherium pointed out that this is far from true. In fact, the era of lawless internet freedom ended many years ago, and social media are already governed by rules that no legislature wrote, no court has reviewed, and no citizen may appeal. In other words, Britain's proposed new measures would not introduce restrictions for the first time, but nationalise and extend ones that already exist.

In July 2012, the UN Human Rights Council declared that the rights people had offline must also be protected online, including that to freedom of expression.³² Then, the major social media platforms – which already had some 1.4 billion users – had little content policy beyond the prohibition of plainly criminal activity. Within a decade, all had built extensive mechanisms to identify and remove lawful speech.

Take YouTube. In March 2017, the *Times* reported that advertising, including that placed by the British government, was appearing on the platform alongside extremist videos. The then-Conservative government responded by declaring that this was “totally unacceptable”, pulling all its ads, and summoning senior figures from YouTube's owner Google to the Cabinet Office. Taking the government's lead, major advertisers also issued protests.³³ Within days, Google's Chief Business Officer announced that the company was “taking a tougher stance on hateful, offensive and derogatory content”, and would reconsider what content was permitted on the platform. By July, YouTube was restricting the distribution of videos identified by campaign groups such as the Institute for Strategic Dialogue and, in the United States, the Anti-Defamation League, and demonetising

31 <https://substack.com/home/post/p-209852382>

32 United Nations Human Rights Council, resolution A/HRC/RES/20/8, “The promotion, protection and enjoyment of human rights on the Internet”, adopted without a vote, 5 July 2012, <https://undocs.org/A/HRC/RES/20/8>.

33 <https://fortune.com/2017/03/21/google-youtube-britain-advertising>

them by removing ads. In 2019 it began to use algorithms not to *amplify* hateful or divisive videos, but to *suppress* what it called “borderline content” that it considered inflammatory, if not unlawful.

This template was adopted elsewhere. Until 2016, Twitter’s rules stated that it would not censor user content, but the publication of the “Twitter files” after Musk bought the platform in 2022 (he renamed it X the following year) revealed it had been operating blacklists, applying “do not amplify” labels to posts and their authors and “visibility filtering”,³⁴ colloquially known as “shadow banning”. In December 2016, Facebook ceased attempting to adjudicate truth for itself and contracted this job to third-party fact-checkers, who acquired the power to label, demote and “flag” material for removal. Here too, alleged misinformation was being stifled, rather than boosted. In some jurisdictions, such as Germany, this policy was supported by a statute, which said that although a platform could be sanctioned for the under-removal of unlawful content, it would face no penalty for over-removal, i.e. taking down content that was harmless, so creating a structural incentive for platform-led censorship.³⁵ Among other examples cited by Arctotherium is Reddit, which removed 222,000 items for “policy breaches” and banned 21,900 forums in 2019 alone. Unpaid volunteer Reddit moderators removed 84.1 million items in the same period.³⁶

Musk’s acquisition of Twitter put a stop to some of these policies, and since the second Donald Trump administration took office in 2025, Facebook too has scaled back its fact-checking programme. But if the pro-censorship lobby within Burnham’s Government prevails, the state will take the reins. Ministers may claim that to compel platforms to give prominence to accredited, “trustworthy” sources does not amount to silencing those deemed less reliable. But this is a zero-sum game. What is not promoted will be demoted, hard to spot, and less likely to reach a monetisable audience – thereby depriving content creators deemed to be “untrustworthy” of their income and driving them out of business.

Meanwhile the Government already has an organisation, now supervised by the DCMS, whose role might easily be expanded – the National Security

³⁴ <https://x.com/bariweiss/status/1601007575633305600>.

³⁵ See <https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/> and, on NetzDG, <https://cdt.org/insights/overview-of-the-netzdg-network-enforcement-law/>.

³⁶ Reddit transparency report for 2019. <https://www.socialmediatoday.com/news/reddit-releases-2019-transparency-report/572899/>.

Online Information Team (NSOIT), formerly known as the Counter Disinformation Unit (CDU). This began to operate in March 2020, just as the first wave of the Covid-19 pandemic struck.

Its baleful record was laid bare in *Ministry of Truth*, a report issued in 2023 by Big Brother Watch,³⁷ which revealed that, initially, misinformation relating to Covid was its core mission. The OSA had not yet come into existence, and lacking the power to fine platforms that failed to take down unlawful content, the CDU mainly “flagged” content it deemed to be false, and tried to persuade platforms that they should remove it for violating their terms of service. *Ministry of Truth* contained extensive evidence that its targets included numerous MPs, such as the former Conservative minister Sir David Davis, who had questioned the wisdom of lockdowns and the accuracy of Covid modelling; articles published on *Lockdown Sceptics* (now renamed the *Daily Sceptic*), the website edited by the FSU’s Director Lord Young; and critiques of the government’s Covid policies from the writer and broadcaster Julia Hartley-Brewer. According to the Big Brother Watch report, the unit “strayed significantly from its ‘disinformation’ remit to monitor and delegitimise domestic political dissent”.

Its rebranding in late 2023 as the NSOIT was supposed to silence such criticism: henceforth, it would only address false information deliberately spread by foreign states – or so it was claimed. That claim did not survive contact with the unit’s own paperwork. In April 2024, after filing freedom of information requests and an intervention by the Information Commissioner’s Office, Big Brother Watch obtained NSOIT’s internal policy documents, which contained almost no reference to national security, describing the NSOIT’s function as being “to build a picture of mis/disinformation threats”, and extending its remit to matters of public safety.³⁸ A year after the Southport murders, Jim Jordan, the chair of the US House Judiciary Committee, forced the disclosure of emails from the unit to social media platforms, warning them against continuing to host “significant volumes of anti-immigrant content” and “concerning narratives about the police and a ‘two-tier’ system that we are seeing across the online environment”. None of the content the NSOIT flagged was unlawful: one such post consisted of nothing more than an accurate statement that a

37 Big Brother Watch, *Ministry of Truth: the secretive government units spying on your speech*, January 2023, pp. 6–7, <https://bigbrotherwatch.org.uk/wp-content/uploads/2023/01/Ministry-of-Truth-Big-Brother-Watch-290123.pdf>

38 Big Brother Watch, ‘Government’s secret counter-disinformation policies revealed’, June 2024, <https://bigbrotherwatch.org.uk/blog/governments-secret-counter-disinformation-policies-revealed/>.

request made under the Freedom of Information Act about the location of hotels used by asylum seekers had been refused. Another was a video posted on X showing men waving Pakistani flags with a caption reading “Looks like Islamabad but it’s Manchester.” According to the NSOIT, this was an example of “misleading or dangerous narratives”, which had been “shared in order to incite fear of the Muslim community”.³⁹

³⁹ <https://freespeechunion.org/archive/britains-expanding-censorship-machine>

4. What is misinformation?

Despite the claims made by actors such as the NSOIT, the CCDH, Zack Polanski, the CCOD and Valent Projects, there is no settled definition of mis- or disinformation, and no tried and tested method by which a government regulator, social media platform or independent researcher can distinguish accurate reports from mis- or disinformation. This is not a semantic quibble, and the academic literature on the subject is deeply divided. By adopting what it calls a “reasonably inclusive definition”, one side of this debate claims misinformation is so prevalent and so damaging that it must be suppressed by passing new censorship laws.⁴⁰ The other disagrees, warning that an over-expansive definition will lead to measures that jeopardise free speech.⁴¹ Yet as Britain’s new internet policy takes shape, the pro-freedom side of this debate is barely being heard.

At a granular level, the labelling of misinformation turns out to be a slippery business. A statement may be denounced as false when first published, but later endorsed by authoritative opinion. Early on in the Covid-19 pandemic, for instance, those who suggested that the virus might have leaked from a lab at the Wuhan Institute of Virology were widely denounced as peddlers of mis- or disinformation. Slowly but surely, this erstwhile “conspiracy theory” has come to be accepted both by numerous medical experts and western intelligence agencies as a plausible or likely explanation.⁴²

Moreover, some censorship lobbyists’ definitions of mis- and disinformation are surprisingly elastic, and amount to an attempt to redefine dissenting but legitimate opinion or analysis as falsehood. For example, funded by an earmarked \$3 million grant from a US foundation, the CCDH currently

40 For a pessimistic summary see Ecker, U. K. H., Tay, L. Q., Roozenbeek, J., van der Linden, S., Cook, J., Oreskes, N. & Lewandowsky, S., ‘Why misinformation must not be ignored’ (2025) 80 *American Psychologist* 867–878, psycnet.apa.org/fulltext/2025-57011-001.pdf

41 Ó Fathaigh, R., Helberger, N. & Appelman, N., ‘The perils of legally defining disinformation’ (2021) 10(4) *Internet Policy Review*, <https://policyreview.info/articles/analysis/perils-legally-defining-disinformation>

42 <https://www.bbc.com/news/articles/cd9qjj4zy5o>

campaigns against what it terms “the new climate denial”. This, it said in a report⁴³ published in 2024, has largely superseded traditional denialists’ claim that the climate is not currently in “crisis” as a result of greenhouse gases emitted by the burning of fossil fuels. It consists instead of misinformation claiming that “climate solutions”, such as the use of biofuels and renewable energy, “won’t work” – that is to say, objections to policies such as the British Government’s commitment to achieve Net Zero energy by 2050. Anyone doubting that the CCDH has few scruples about restricting free speech need look no further than its early success in defunding *The Canary*, a left-wing website that supported Jeremy Corbyn when he was Labour leader. Determined to boost the right-wing, Starmerite faction of the party, it set up Stop Funding Fake News (SFFN), an “Astroturf” campaign group: an organisation that falsely claimed to be a network of grassroots activists when it was in fact the creation of the high-level apparatchiks Ahmed and McSweeney. This succeeded in persuading advertisers to withdraw support from *The Canary* on the bogus grounds that it was a cesspit of “hatred” and misinformation,⁴⁴ and it was forced to shed all but one of its 29 staff. SFFN also targeted right-wing websites, such as *Westmonster* and *Breitbart*.

The record of the various independent and semi-official fact-checking organisations is no more encouraging. In Britain, the most influential has been the Global Disinformation Index (GDI), founded in 2018 with funding from the Foreign Office. During the Covid pandemic, it was also funded by the UK government’s CDU. Like the CCDH, its stated aim is to defund publications it considers undesirable by pressurising advertisers.

In an essay published by *UnHerd* in 2024, its editor-in-chief, Freddie Sayers, described how the GDI weaponises its “dynamic exclusion list” of publications that “supposedly promote ‘disinformation’ and should therefore be boycotted by all advertisers”. That January, having been repeatedly asked whether *UnHerd* was on this list, the GDI finally admitted that it was, and provided examples of the “offending” content behind their determination: “Kathleen Stock [the gender critical philosopher hounded from her post at Sussex University by trans activists], whose columns are up for a National Press Award this week, Julie Bindel, a lifelong campaigner against violence against women, and Debbie Hayton, who is transgender. Apparently the GDI equates ‘gender-critical’ beliefs, or maintaining that biological sex differences exist, with ‘disinformation’, despite the fact that

43 <https://counterhate.com/research/new-climate-denial/#about>

44 <https://orbooks.com/catalog/the-fraud/>; for details see also <https://freespeechunion.org/publications/starmer-misinformation-mafia>

those beliefs are specifically protected in British law and held by the majority of the population.”⁴⁵

Sayers went on to outline the consequences: “The verdicts of ‘ratings agencies’ such as the GDI, within the complex machinery that serves online ads, are a little-understood mechanism for controlling the media conversation. In *UnHerd*’s case, the GDI verdict means that we only received between 2% and 6% of the ad revenue normally expected for an audience of our size. Meanwhile, neatly demonstrating the arbitrariness and subjectivity of these judgements, NewsGuard, a rival ratings agency, gives *UnHerd* a 92.5% trust rating, just ahead of the *New York Times* at 87.5%.”

Initially, the GDI purported to expose only untruth. But as its chief executive Clare Melford explained in an interview in 2021, it has undergone a little mission creep: now, she said, she thought it more useful to adopt a broader definition that encompassed not only inaccuracies but anything that fostered an “adversarial narrative” that could be “harmful or divisive”. According to Melford, “a lot of disinformation is not just whether something is true or false – it escapes from the limits of fact-checking. Something can be factually accurate but still extremely harmful.”⁴⁶ Earlier, Melford adopted a neologism: what she calls “malinformation”, another harmful content category she has defined as “correct information distorted or misrepresented to cause harm”.⁴⁷ She also revealed in this interview that many of the GDI’s judgments as to what *was* disinformation were not made by human beings, but AI: “Each adversarial narrative is given its own machine-learning classifier, which then allows us to search for content that matches that narrative at scale... misogyny, Islamophobia, anti-Semitism, anti-black content, climate change denial, etc.”

She went on to give what turned out to be an unfortunate example: a Spanish report claiming that a third of UK deaths from the Covid Delta variant in 2021 had been among people who had been vaccinated. This, Melford said, was “clearly untrue”, and the GDI duly tried to persuade the Chipotle restaurant chain, which had placed an online ad next to this “highly dangerous disinformation”, to withhold future support. In fact, a June 2021 report by Public Health England found that 29% of patients who died from Delta had been fully vaccinated, and 17% partly vaccinated

⁴⁵ <https://unherd.com/2024/04/inside-the-disinformation-industry/>

⁴⁶ <https://www.youtube.com/watch?v=W-q1I5QA2Fs>

⁴⁷ <https://committees.parliament.uk/writtenevidence/88260/html/>

– in other words, the Spanish claim was not untrue at all in the way Melford claimed, but an underestimate.⁴⁸ Earlier, in 2020, the GDI had published a blog about those peddling the Wuhan lab leak “conspiracy theory”, arguing that “cutting off ads to these fringe sites and their outer networks is the first action needed”.⁴⁹

In 2022, the GDI produced a list of US news website rankings, saying which were the least and which the most liable to publish disinformation. All but one of the ten lowest risk sites were liberal or left leaning, while all of the ten judged most risky were libertarian or conservative, including the *New York Post*, *RealClearPolitics*, *Reason*, the *Federalist* and the *American Spectator*. That may partly explain why the Trump administration announced last year it intended to deny US entry to Melford and Imran Ahmed of the CCDH. (Ahmed, who is fighting his corner in a federal court, currently resides in Washington.) According to Secretary of State Marco Rubio, they had led “organized efforts to coerce American platforms to censor, demonetize, and suppress American viewpoints they oppose”.⁵⁰ Apparently without irony, a GDI spokesperson promptly accused Rubio of mounting “an authoritarian attack on free speech and an egregious act of government censorship”.⁵¹ In Britain, the FSU and the Conservative MP Philip Davies worked together to elicit the information in 2023 that in the period 2019–22, the GDI received some £2.6 million from taxpayers, channelled via the Foreign Office.⁵² This funding has now ceased.

Any regime that purports to deal with the threat posed by misinformation must be able to answer crucial questions. Who will determine what is misinformation? What standard will they use? Will they give reasons for their determinations, and can they be appealed? Without good answers to those questions, any mechanism established by the state will pose a mortal threat to lawful free expression.

A strong indication of the Government’s answers is due by 1 November. By that date, under section 152 of the OSA, the six-person Online Information Advisory Committee, chaired by Lord Allan of Hallam, a former

48 Public Health England, *SARS-CoV-2 variants of concern and variants under investigation in England*, Technical Briefing 16, 18 June 2021

49 Sayers, *ibid* note 45.

50 <https://edition.cnn.com/2025/12/23/politics/sanctions-censorship-state-rubio>

51 <https://www.theguardian.com/technology/2025/dec/24/undermines-free-speech-labour-mp-hits-back-at-us-government-over-visa-related-sanctions>

52 <https://europeanconservative.com/articles/commentary/governments-are-beginning-to-resist-disinformation-index/>

Ofcom non-executive director, is due to deliver its first statutory report.⁵³ This, says the Act, will “advise Ofcom on how providers of regulated services should deal with disinformation and misinformation”, and how officialdom should counter it. The committee is supposed to post online minutes of its monthly meetings, but the most recent currently available date from October 2025. These suggest the committee shares Powell’s attitude to unregulated speech on social media during election campaigns. The minutes state that as of then, the committee was planning to meet the Electoral Commission to discuss a joint approach, and that both last October and at previous meetings it had discussed “concerns” over “misinformation and disinformation on the eve of an election” and “efforts to challenge or undermine elections based on false information”.⁵⁴ Tackling this, the minutes say, could extend the “online safety regime”.

The October 2025 minutes also record the committee’s distress at “the apparent scaling back of some fact-checking by some online platforms”, as well as “the benefits of having independent fact-checking organisations” such as the GDI – although they would benefit from greater “agility” and faster “fact-checking response times”.

⁵³ <https://www.ofcom.org.uk/about-ofcom/structure-and-leadership/advisory-committee-on-disinformation-and-misinformation>

⁵⁴ <https://www.ofcom.org.uk/siteassets/resources/documents/about-ofcom/structure-and-leadership/online-information-advisory-committee/meeting-minutes/minutes-of-the-fifth-meeting-of-the-online-information-advisory-committee-held-on-3-october-2025.pdf?v=416442>

5. Policy-based evidence making: the GO-Science review.

May 2026 saw the publication of a crucial official document: a review of online misinformation commissioned by the Government Office for Science from a seven-person team based at Sheffield University. It appears to stem from the Government's response to the Science, Innovation and Technology Select Committee report on the aftermath of Southport referred to above, which admitted it was "difficult to assess the true extent of algorithmic amplification of harmful content", and that further research was needed to establish the scale of misinformation.⁵⁵ In the words of the information minister Baroness Jones, "policymaking in this space has lacked a full evidence base". The review was supposed to supply it.

Its executive summary set out the censorship lobby's usual shibboleths, claiming that false and misleading information "is now widely recognised as a systemic national risk", that "algorithms and business models contribute to [its] amplification and impact" and that "system-level interventions are needed, especially addressing algorithms and business models". Meanwhile, debates over how to define misinformation "should not become a distraction from action".⁵⁶

The report took just six months to compile from start to finish, and concluded that in areas including health, climate change, sex and gender, race, immigration and finance, "false and misleading narratives" had become "an embedded part of the information environment". Yet most of its seven co-authors and eleven "expert" reviewers lacked the scientific background that might have enabled them to assess the accuracy of research in these fields. The two groups contained no climate scientist, no economist and no business expert, although the report speaks of misinformation about "EU financial policy" and "digital currency surveillance"; no demographer, criminologist or scholar of migration, though it condemns "misleading" claims about "migrant entitlement" and "the demographic composition of

⁵⁵ Science, Innovation and Technology Committee, *ibid* note 11.

⁵⁶ Baird et al., *ibid* note 6.

migrant arrivals”; no clinician, immunologist, or epidemiologist, though it says misinformation on health is the “most persistent” category; and no one with experience of research on the hotly-contested subject of sex and gender, though the report attacks misinformed “anti-LGBTQ+ narratives” and the spread of “transgender athlete controversies”.

Then again, the report’s methodology meant this made little difference. For despite purporting to reveal hard evidence of the malignant growth of misinformation, the team made no attempt to decide for itself what was true or false, nor how many incorrect statements there might be on its chosen subject areas. Instead, by their own account, they relied on summaries produced by large language AI models of academic articles that the authors had not read, along with claims made by fact-checking organisations whose methods and reliability the report leaves opaque.

Moreover, many of those involved in this process were marking their own homework. Among the reviewers was none other than Melford of the GDI, a partisan combatant in the politicised contest of which the report was supposed to be an impartial judge. Joining her were Bob Ward, a communications officer at the London School of Economics, whose professional output largely consists of complaints against politicians and journalists who criticise aspects of climate science and energy policy; Chris Morris, Chief Executive of the UK fact-checking organisation Full Fact; and Paula Gori, Chief Executive of the European Digital Media Observatory (EDMO), an all-EU body with 16 national branches, which coordinates the work of fact-checkers across the continent. Self-evidently, they had a strong interest in ensuring the report concluded that misinformation is jeopardising social cohesion, while their own work is cited numerous times in its final text.

EDMO’s monthly briefings were the GO-Science authors’ principal source of data adduced to support their argument that the problem of misinformation is getting worse. However, they do not constitute empirical evaluations of what gets posted month by month on social media, and they say nothing about its impact. They merely record replies to questionnaires that EDMO sends to fact-checking organisations in different parts of Europe, asking what subjects they have dealt with. Hence, the review notes that misinformation about health was detected by EDMO in every month, and about climate change almost as often. But this constitutes merely a digest of what EDMO-affiliated fact-checkers choose to check in each given period. It says nothing about how many times misinformed posts were seen, the standards by which they were judged, or whether they were believed.

And like the CCDH and the GDI, the review blurs the line between posts about science and public policy. For example, its section on climate change says that “misinformation is increasingly focused on undermining strategies for mitigating climate change rather than denying climate change outright”, and that social media narratives “decouple policies (e.g. net zero) from climate science, creating public confusion”. Yet decoupling net zero policy from climate science is not confusing at all. It is merely a recognition that the desirability, pace, and costs of decarbonisation are matters of legitimate economic and political debate, not physics. The same section treats “targeted misinformation campaigns against technologies (e.g., EVs, heat pumps)” as an “emerging risk”. It is, apparently, misinformation to note that heat pumps are an expensive and inefficient way to keep people warm in winter, or that electric vehicles lack range and depreciate in value more quickly than petrol vehicles.

The GO-Science authors and experts may not have the skills to distinguish fact from falsehood in the fields they consider. However, many of them do have a background in the strange meta discipline that is the academic study of alleged misinformation, a discipline that depends for its future existence on governments and other funders continuing to believe that mis- and disinformation poses a deadly social threat. For example, co-author Kalina Bontcheva, a Sheffield University professor of natural language processing, is a former member of the European Commission’s High-Level Expert Group on fake news and disinformation and a member of the EDMO advisory council. Among the expert reviewers were Stephan Lewandowsky, a Bristol University psychology professor, and John Cook from the Melbourne Centre for Behaviour Change, Lewandowsky’s one-time doctoral student. Their work argues that it is futile to engage with climate sceptics’ arguments, because their criticism of mainstream climate science is best explained by their right-wing, free market politics and tendency to believe in conspiracy theories. Hence the title of a notorious 2013 paper of which Lewandowsky was a co-author, *NASA Faked the Moon Landing – Therefore, (Climate) Science Is a Hoax*.⁵⁷ Another paper covering similar ground, in which both men were involved, *Recursive Fury*, was eventually retracted by the journal *Frontiers in Psychology* following complaints that it failed to protect the rights of those it had studied.⁵⁸

57 Lewandowsky, S., Oberauer, K. & Gignac, G. E., ‘NASA Faked the Moon Landing – Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science’ (2013) 24(5) *Psychological Science* 622–633, <https://pubmed.ncbi.nlm.nih.gov/23531484/>

58 *Frontiers in Psychology* Editorial Office, ‘Retraction: Recursive fury’ (2014) 5 *Frontiers in Psychology* 293, retracting Lewandowsky, S., Cook, J., Oberauer, K. & Marriott, M., ‘Recursive fury: conspiracist ideation in the blogosphere in response to research on conspiracist ideation’ (2013) 4 *Frontiers in Psychology* 73, <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2014.00293/full>

The GO-Science review admitted that in most areas there is “no standardised framework” for measuring the harm misinformation causes. But in its view, the shortage of such evidence is no reason to delay action. It went on to recommend new, statutory means of imposing platform censorship, including the mandatory incorporation of “independent quality signals” into the ranking of content – effectively the proposal in the DCMS Green Paper – and a state-approved, monthly updated “database of false and misleading narratives spreading in the UK” cross-linked to EDMO.

The report’s conclusions were stark: “Voluntary platform self-regulation is insufficient. Statutory frameworks with independent audit and enforcement are needed... System-level interventions are needed, especially addressing algorithms and business models.” They also advocate both “prebunking” of misinformed messages across “education, crisis preparedness and public communications”, and “debunking” by officialdom after they have emerged: a radical blueprint for far-reaching state control.

Just as this briefing went to press, Sheffield University responded to my Freedom of Information Act inquiry as to how much the GO-Science review cost the taxpayers who funded it. The answer: almost half a million pounds, or to be exact, £496,355.

6. The missing counter-arguments

Anyone who uses a large language AI model will learn that the value of its output depends on the quality of the input prompt. This may explain the gaps and deficiencies in the GO-Science review, for its list of the 228 published papers that it asked AI to summarise makes no mention of significant studies published by some of the world's most prestigious journals that might have challenged its dystopian outlook. Take, to begin with, the core claim that misinformation is wrecking democracy and undermining social cohesion. It begs some basic questions. How much of it do people actually see? How does that compare with other sources of news? And is it influencing its recipients, filling their previously docile minds with hatred and conspiracy theories? When censorship advocates make their claims about misinformation's reach, they rarely address these studies.

Some of their assertions are essentially meaningless, such as the Science, Innovation and Technology Select Committee's claim that misinformation about the Southport child murders had a "potential reach" of 1.7 billion people. Presumably, accurate information had an even bigger "potential" audience, given that most members of the world's 8 billion population have access to radio or television, even if they don't have accounts on Instagram or X.

There are earlier examples. In February 2019, when the Commons Culture, Media and Sport Select Committee published its own report on disinformation, it looked at the claims that Russian bots spewing falsehoods had influenced the 2016 EU referendum. It largely relied on analysis by 89up, a communications agency, which claimed that anti-EU articles from Kremlin-owned outlets had achieved a "social reach" of 134 million "potential impressions" – a term used by the advertising industry to describe the theoretical ceiling on the audience that might, just conceivably, have encountered a given message. It does not measure the number of unique viewers, their verified exposure, possible shifts of belief or the message's

electoral effect. But it is the kind of figure a consultancy produces when a large number is wanted. (The Southport 1.7 billion figure also appears to belong in this category.)

The assertions made by 89up and the select committee garnered headlines. Less widely known is a rigorous study by the Oxford Internet Institute, which examined 5.8 million Brexit-related tweets from the referendum period against lists of accounts supplied both by Twitter and other sources that were suspected of links to the Russian state. The result: just 105 supposedly Russian accounts produced any Brexit content at all, generating a total of 15,964 tweets – less than 0.3% of the sampled traffic. There was, said the Institute, “little evidence” of successful interference. Links to known Russian outlets comprised just 511 of 87,169 shared links: less than 0.6%. The same went for other platforms, with only a “tiny amount” of Russia-linked content about Brexit on YouTube.⁵⁹ In the same period, it was widely claimed that Facebook posts backed by Russia might have reached 126 million Americans in the period leading up to Trump’s first presidential election victory in 2016. Less well known is the subsequent, empirical caveat published in the *Columbia Journalism Review*: that such content represented a paltry 0.004% of the content that US citizens saw in their Facebook news feeds in the relevant period, a total of just four posts in every ten thousand.⁶⁰

More recent attempts to measure the real impact of misinformation include an authoritative synthesis of empirical evidence published in 2024 in what is arguably the world’s leading peer-reviewed scientific journal, *Nature*. The work of a team led by Ceren Budak of the University of Michigan, it suggested that the problem is far less severe than the censorship lobby claims. Citing earlier research by Guess, Nyhan and Reifler, it noted that visits to 490 websites deemed “untrustworthy” on the basis of lists published by fact-checkers made up just 5.9% of Americans’ clicks on news sites. Moreover, a large majority of those who *did* access such content across the different platforms did not see it by accident, nor because it was fed to them by malign, revenue-maximising algorithms, but because they took

59 Narayanan, V., Howard, P. N., Kollanyi, B. & Elswah, M., ‘Russian Involvement and Junk News during Brexit’, COMPROP Data Memo 2017.10, Oxford Internet Institute, <https://demotech.oii.ox.ac.uk/research/posts/russian-involvement-and-junk-news-during-brexit/>

60 Watts, D. J. & Rothschild, D. M., ‘Don’t blame the election on fake news. Blame it on the media’, *Columbia Journalism Review* (2017), <https://www.cjr.org/analysis/fake-news-media-election-trump.php>.

active steps to seek it out, being already motivated to do so.⁶¹ In other words, it consolidated, rather than formed, their view of the world. The paper was not cited by the GO-Science review, and it was not among the 228 its authors had summarised by AI.

According to the paper's five authors, "exposure to misinformation is low as a percentage of people's information diets and concentrated among a small minority... exposure to misinformation and extremist content is not frequent; instead, it is relatively rare and highly concentrated among small groups of extremists". They did not deny that in dealing with the "tails", the relatively small segment of the population which consumed the most misinformation, there was a case for some kind of intervention. But this should be targeted and proportionate, not a wholesale transformation of the information environment of the kind the Green Paper proposes.

Another group of researchers added TV news to the mix, concluding that "fake news" and misinformation constituted just 0.15% of Americans' daily media diet.⁶² Yet another study found that while news consumption increased in 2020 during the Covid pandemic, it was "trustworthy" sites that benefited most, while claims about a supposed "infodemic" of misinformation decried by the censorship lobby had been overblown: "Despite widespread public concern about the reliability of online news and real problems with various kinds of misinformation, untrustworthy news outlets accounted for a small minority of web traffic to news sites in 2020." Overall, the study concluded that "untrustworthy" news sources accounted for between 0.1% and 4.4% of internet browsing across the US, UK, France and Germany over the period 2017–2021.⁶³ As Budak and her colleagues noted, "This reality is not reflected in public discourse about social media."⁶⁴

61 Budak, C., Nyhan, B., Rothschild, D. M., Thorson, E. & Watts, D. J., 'Misunderstanding the harms of online misinformation' (2024) 630 *Nature* 45–53. <https://doi.org/10.1038/s41586-024-07417-w>, citing Guess, A. M., Nyhan, B. & Reifler, J., 'Exposure to untrustworthy websites in the 2016 US election' (2020) 4 *Nature Human Behaviour* 472–480.

62 Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J., 'Evaluating the fake news problem at the scale of the information ecosystem' (2020) 6(14) *Science Advances* 3539, <https://www.science.org/doi/10.1126/sciadv.aay3539>

63 Altay, S., Nielsen, R. K. & Fletcher, R., 'Quantifying the "infodemic"' (2022) 2 *Journal of Quantitative Description: Digital Media* 1–30, <https://www.researchgate.net/publication/362977394>

64 Budak et al., *ibid* note 61.

7. The algorithm panic

In a speech he gave in Parliament⁶⁵ on the threats to social cohesion in March 2026, the then-Communities Secretary Steve Reed named a source of evil. Our enemies, he said, “feed off deliberate misinformation, hatred and extremism, carried across social media by algorithms, and funded by hostile states and rogue billionaires determined to pull our communities apart”. Echoing similar assertions made in other jurisdictions, he was rolling the pitch for the censorship proposals now in play. But does the evidence support these claims?

There is evidence that platform algorithms predict users’ engagement with content, and that this can amplify “emotionally charged” material. An experimental study published in 2025 found that X’s algorithm did exactly this, which could make users feel more hostile to their “political out group”. However, the same study also noted: “Users do not prefer the political tweets selected by the algorithm, suggesting that [it] underperforms in satisfying users’ stated preferences.”⁶⁶

The study was cited by the DCMS Green Paper. However, “emotionally charged” is not the same as “extreme”, and here the evidence does not support censorship. The biggest and most authoritative study to address this issue was published in the journal *PNAS* by a team of researchers at Twitter itself in 2022. Covering seven countries, millions of accounts and more than six million news items, this paper found that the algorithm tended to amplify posts from right-wing legislators and news sources more than those on the left. But there was no sign it was also fostering extremism: “We further looked at whether algorithms amplify far-left and far-right political groups more than moderate ones; contrary to prevailing public belief, we did not find evidence to support this hypothesis.” In countries where far-right and far-left parties were strong, such as France, Spain

65 <https://hansard.parliament.uk/commons/2026-03-09/debates/7675E154-5277-444E-B837-D34D7495F562/SocialCohesionActionPlan>

66 Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S. & Dragan, A. D., ‘Engagement, user satisfaction, and the amplification of divisive content on social media’ (2025) 4(3) *PNAS Nexus* pgaf062, academic.oup.com/pnasnexus/article/4/3/pgaf062/8052060

and Germany, “the amplification of these parties is generally lower than that of moderate/centrist parties in the same country”. Overall, the authors concluded, “our analysis of far-left and far-right parties in various countries does not support the hypothesis that algorithmic personalization amplifies extreme ideologies”.⁶⁷ This study is not mentioned in either the Green Paper or the GO-Science review.

Meanwhile, a study published in *Science* that examined the impact of Facebook and Instagram during the 2020 US election found that its algorithmic feed served users *less* content from “untrustworthy” sources than the alternative, a “chronological” feed that ranked posts simply according to when they were issued. Moreover, switching users off algorithmic ranking for three months produced no measurable change in their political knowledge or attitudes, and “did not significantly alter levels of issue polarization”.⁶⁸

As for YouTube, research published in 2023 based on the social media habits of some 1,200 survey participants found only a tiny minority of those viewing extremist videos had been directed to them by the algorithm. Instead, most were subscribers who were actively looking for material of this kind: “Exposure to alternative and extremist channel videos on YouTube is heavily concentrated among a small group of people with high prior levels of gender and racial resentment.”⁶⁹ A separate study published in 2021, derived from a vast dataset of more than 300,000 users, reached similar conclusions: “News consumption on YouTube is dominated by mainstream and largely centrist sources. Consumers of far-right content, while more engaged than average, represent a small and stable percentage... We find no evidence that engagement with far-right content is caused by YouTube recommendations systematically, nor do we find clear evidence that anti-woke channels serve as a gateway to the far right.”⁷⁰

67 Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A. & Hardt, M., ‘Algorithmic amplification of politics on Twitter’ (2022) 119(1) *PNAS* e2025334119, www.pnas.org/doi/10.1073/pnas.2025334119

68 Guess, A. M. et al., ‘How do social media feed algorithms affect attitudes and behavior in an election campaign?’ (2023) 381 *Science* 398–404. <https://www.science.org/doi/10.1126/science.abp9364>

69 Chen, A. Y., Nyhan, B., Reifler, J., Robertson, R. E. & Wilson, C., ‘Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels’ (2023) 9 *Science Advances* eadd8080, www.science.org/doi/10.1126/sciadv.add8080

70 Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M. & Watts, D. J., ‘Examining the consumption of radical content on YouTube’ (2021) 118 *PNAS* e2101967118, www.pnas.org/doi/10.1073/pnas.2101967118

Disregarding the research cited above, as many have done in the past, the GO-Science review claimed that platform algorithms promote inflammatory content because inflammatory content is profitable. On the basis of a single citation, a paper published by academics at the Universitas Muhammadiyah Buton in Bau-bau, Indonesia in 2025, it stated that “the governance of online platform algorithms is exclusively profit- and engagement-driven”. This, the review went on, meant they were incentivised to promote “false and misleading information”.⁷¹ The Universitas Muhammadiyah Buton is a small private institution on Buton island in the province of Southeast Sulawesi. According to the international higher education directory Unirank, it is the 433rd best university in Indonesia, and the 10,370th best in the world.⁷² In any event, the paper cited by the GO-Science review was not an original piece of empirical research, but a synthesis of earlier papers the researchers produced with help from AI.⁷³ Some might consider this a somewhat fragile basis for such a sweeping claim.

In fairness, it could be said that the allegations made in 2021 by the Meta whistleblower Frances Haugen, a former member of Facebook’s Civic Integrity Team, supported the assertions made by the GO-Science review, Steve Reed and others. According to Haugen, a deliberate policy to promote posts that drew angry reactions was leading to the spread of “hateful and harmful” material, and polarising political discourse.⁷⁴ But Meta pushed back against her claims as soon as she made them, with its spokesperson saying that to host such content was “bad for our community, bad for advertisers, and ultimately, bad for our business”. In February 2024, the beginning of a year of hotly-contested elections in both Britain and the United States, the company took steps to *reduce* the number of political posts users were likely to see. It changed its algorithm so that it started to remove such posts from feeds unless an account holder was already following their authors. Far from intensifying political engagement and divisions, this could only lessen their reach.⁷⁵

71 Baird et al. *ibid*.

72 <https://www.unirank.org/id/uni/universitas-muhammadiyah-buton/>

73 Hastuti, H., Maulana, H., Lawelai, H. & Anser, H., ‘Algorithmic influence and media legitimacy: a systematic review of social media’s impact on news production’ (2025) *Frontiers in Communication*, <https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1667471/full>

74 <https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-60-minutes-polarizing-divisive-content/>

75 <https://www.npr.org/2024/03/26/1240737627/meta-limit-political-content-instagram-facebook-opt-out>

8. Exaggerating the impact of misinformation

It is reasonable to assume that if misinformation really is a threat as great as ministers seem to suppose, there would be copious evidence of its impact, such as data showing that levels of belief in the spurious conspiracy theories that misinformation is supposed to encourage are rising. This question was addressed directly in the title of a study led by Joseph Uscinski of the University of Miami that was published in 2022: *Have beliefs in conspiracy theories increased over time?* This examined the results of national US opinion surveys relating to 46 separate conspiracist beliefs conducted repeatedly since 1966 using identical wording each time; support for six theories based on surveys in six European countries, including the UK; and generalised “conspiracy thinking”, measured eight successive times in America between 2012 and 2021. “In no instance”, the study states, “do we observe systematic evidence for an increase in conspiracism”.⁷⁶

In America, belief in many conspiracy theories had not risen but declined: of the 46 theories surveyed, declines outnumbered increases by 15 to six, and in Europe, just one conspiracy theory had acquired more support over time in just one of the six countries that were surveyed. Ironically, the only misinformed belief that was more widely held was that conspiracy theories had become a serious problem: according to the study, 73% of Americans had come to believe that conspiracy theories were “out of control”, and 59% agreed that people were more likely to believe conspiracy theories “compared to 25 years ago”. This, the study suggested, was a result of alarmist statements by politicians and the mainstream media. The authors even found that belief in QAnon, often cited as a particularly damaging and widespread phenomenon, declined between two surveys conducted in 2019 and 2021.

76 Uscinski, J., Enders, A., Klofstad, C., Seelig, M., Drochon, H., Premaratne, K. & Murthi, M., ‘Have beliefs in conspiracy theories increased over time?’ (2022) 17(7) *PLOS ONE*, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270429>

Overall, the study concluded: “We do not observe supporting evidence that beliefs in conspiracy theories or generalized conspiracy thinking have increased during the internet/social media era. Instead, our findings comport with arguments that the internet may be less hospitable to conspiracy theories than is often assumed.” The evidence also suggested that “people are unlikely to believe a conspiracy theory unless they are both 1) already disposed to believe conspiracy theories generally, and 2) inclined towards the content of that particular conspiracy theory or the source from which it emanates. In other words, the presence of online conspiracy theories might not *persuade* as much as *reinforce* existing views.”⁷⁷

The GO-Science review devotes considerable space to the subject of vaccines and Covid-19, inferring that alleged misinformation may have led to “societal harms” such as vaccine hesitancy. To be sure, one study did find that Covid-19 vaccine misinformation was likely to produce a decline in vaccine uptake, although this was based not on uptake data, but an opinion survey conducted before vaccines were available, in which respondents were asked whether misinformation might, hypothetically, make them reluctant to be inoculated.⁷⁸

However, another study, published in 2022, suggested that “exposure to false information both supporting and opposing Covid-19 vaccines actually *increased* vaccination intentions”, although the effect was small.⁷⁹ It might seem logical, the authors said, “to assume that the relationship between misinformation exposure and behavioural intentions is causal and mediated by the perceived accuracy of the false news encountered. In this paper, we have tried to provide empirical evidence for these effects, with no success... the relationship between exposure to fake news, perceived accuracy, beliefs, and behaviour is not as straightforward as is often assumed.”

Candour requires an acknowledgment: these findings concern stated intentions rather than observed behaviour, and the honest reading of the academic literature is that the causal evidence is thin and contested in

⁷⁷ Uscinski et al., *ibid*.

⁷⁸ Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K. & Larson, H. J., ‘Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA’ (2021) 5 *Nature Human Behaviour*, <https://doi.org/10.1038/s41562-021-01056-1>

⁷⁹ de Saint Laurent, C., Murphy, G., Hegarty, K. & Greene, C. M., ‘Measuring the effects of misinformation exposure and beliefs on behavioural intentions: a COVID-19 vaccination study’ (2022) 7 *Cognitive Research: Principles and Implications* 87 (<https://doi.org/10.1186/s41235-022-00437-y>)

both directions – which is itself an argument against building statutory speech controls based upon it. The most substantial recent study was published in *Science* in 2024. Combining survey experiments with Facebook exposure data covering some 13,000 vaccine-related links and 233 million American users, it estimated that content flagged as misinformation by fact-checkers had little aggregate effect on vaccination intentions for the simple reason that hardly anyone saw it: such material attracted just 0.3% of the 2.7 billion views that the vaccine stories in the dataset received. However, what the study claimed *did* depress intentions to be vaccinated was unflagged “vaccine-sceptical” content that contained no falsehoods at all, which had 46 times the aggregate impact on vaccination intentions of content that did. The most influential was a single accurate article in the *Chicago Tribune* headlined ‘A healthy doctor died two weeks after getting a COVID vaccine; CDC is investigating why’. This was seen by more than 50 million people: a total more than six times as high as the views of all flagged vaccine misinformation combined.⁸⁰

Advocates of the measures examined in this briefing cite this study as proof that content moderation must go further, but it proves the opposite. The material with the measurable effect was truthful reporting by an established metropolitan newspaper – precisely the “trustworthy” journalism the Green Paper proposes to amplify. A regime aimed at “misinformation” would have missed it; a regime capable of suppressing it would have crossed from the correction of falsehood into the suppression of accurate news, on the Melford principle that something can be “factually accurate but still extremely harmful”. That is the choice the censorship lobby prefers not to state.

Research on the effect of misinformation on politics has reached analogous conclusions, which provide a strong argument against Lucy Powell’s suggestion that social media should be bound by constraints similar to those imposed on broadcasters during election campaigns. The Ofcom Broadcasting Code’s rules on accuracy, balance and election period purdah derive from broadcasting’s nature and origins, in the licensed use of a finite spectrum of frequencies and the passive reception, by listeners and viewers, of content they cannot generally influence. Social media are obviously very different, and any attempt to introduce restrictions of the kind envisaged by Powell would be liable to legal challenge under the

80 Allen, J., Watts, D. J. & Rand, D. G., ‘Quantifying the impact of misinformation and vaccine-skeptical content on Facebook’ (2024) 384 *Science* eadk3451, <https://www.science.org/doi/10.1126/science.adk3451>

freedom of expression provisions set out in Article 10 of the European Convention on Human Rights and the Human Rights Act 1998, which enshrined them in UK law. Her proposals would amount to state-supervised regulation of political speech, with some kind of enforcement body determining who and what could be amplified and to what extent at precisely the time when political free speech matters most.

And like the claim that misinformation reduced Covid vaccine take-up, the data undermine her case. In January 2025, the UK's national body for data science and artificial intelligence, the Alan Turing Institute, submitted written evidence to an inquiry by the Commons Public Administration and Constitutional Affairs Committee into the general election the previous year. It said its research had found "no conclusive evidence" that dis- or misinformation, including deepfakes and misinformation generated by AI, had "meaningfully impacted election results". In a refrain that may by now sound familiar, it said those who accessed such material were "typically already ideologically aligned with the content, so that the material reinforced existing partisan divisions rather than moving undecided voters".⁸¹

81 Written evidence from The Alan Turing Institute (RGE 11) to the Public Administration and Constitutional Affairs Committee, Review of the 2024 General Election inquiry, <https://committees.parliament.uk/writtenevidence/135409/pdf/>

9. Promoting “trustworthy” sources: the case of the BBC

It could be said that this is a peculiar time for a government to be advocating the algorithmic amplification of trustworthy sources, notably the BBC, for its reputation is not at a high point. It may have editorial staff devoted to the cause of combating misinformation, and in 2023, it set up its own, lavishly-funded unit with more than 60 staff to fact-check information using open-source intelligence and to produce its own, gold standard reports, BBC Verify. But this got off to a rocky start. As I revealed in an article for *UnHerd*, it had been forced to withdraw or modify 6% of its output during its first 12 months of operation. One embarrassing example was a misinformed story about the disturbances that followed the Southport murders: having initially broadcast the claim that a racist mob had attacked a car occupied by Asian men on Humberside, BBC Verify had to issue a correction, for the police had confirmed that in fact, the victims were white.⁸² (For some reason, this false claim has not figured in politicians’ lamentations over the effect of misinformation on the Southport riots.) Another was a high-profile investigation claiming that car insurance firms were racially discriminating against ethnic minority drivers by charging them higher premiums. Seven months after it was broadcast and an article published on the BBC News website, the Corporation took it down, replacing it with a statement admitting there was no “ethnicity penalty” after all, and that the BBC’s reporter had failed to see “the limitations of the data we used”. The Corporation said the story “did not meet our normal editorial standards”, and as a result, “we have taken the decision to permanently remove [it]”.⁸³

Much worse was to come. In 2025 the veteran journalist Michael Prescott, who had served for three years as an adviser to the BBC’s Editorial Guidelines and Standards Committee, sent the Board a memorandum after leaving the role. His central charge was not that errors occurred, but when the BBC’s own advisers documented them, senior executives repeatedly resisted or minimised their findings and failed to act. This failure had become systemic.

⁸² <https://unherd.com/2024/11/who-verifies-bbc-verify/>

⁸³ <https://www.bbc.com/news/business-68349396>

One of the cases he cited made an immediate impact: his disclosure that a programme made by *Panorama*, the BBC's venerable current affairs flagship, had spliced together two passages of Trump's speech made before the storming of the Capitol on January 2021 that had in fact been delivered almost an hour apart. According to *Panorama*, Trump told the crowd: "We're going to walk down to the Capitol and I'll be there with you, and we fight. We fight like hell." The cut came after the words "I'll be there with you", and in other portions of the speech, Trump had stressed that any protest must be peaceful, but the edited version created the misleading impression that the once and future president had issued a call to violent action.⁸⁴ Immediately after the edited speech, the programme showed footage of the Proud Boys marching on Congress, although their march had begun before Trump started speaking: a misleading quotation followed by a false chronology, which suggested the violent Proud Boys were responding to Trump's command.

Early this century, I was a *Panorama* reporter, and I am certain that what to me, as a journalist with 45 years of experience, looks like a flagrant and deceptive breach of journalistic ethics, would not have passed muster when I worked there. *Panorama* edits were painstaking, exhausting experiences, in which my colleagues and I spent days poring over the nuances of every shot and every line in a script. But when an internal BBC review raised the Trump speech edit, the executives' response was that shortening speeches was normal practice. That was true, but it did not address the issue of making a US president appear to say what he didn't. Prescott's written warning to the Chairman that to accept such editing set a dangerous precedent went unanswered.

The leak of Prescott's memo in November 2025 had unprecedented consequences. Within days Director-General Tim Davie and the chief executive of BBC News, Deborah Turness, the creator of BBC Verify, had resigned. The Chairman, Samir Shah, apologised to the Culture, Media and Sport Select Committee for *Panorama*'s "error of judgement", and its Chair said the Corporation had serious questions to answer about both its editorial standards and its management's handling of internal concerns.⁸⁵ Meanwhile in America, Trump refused to accept the BBC's apology, launching a lawsuit in federal court that may yet inflict an enormous financial blow on Britain's "trustworthy" national broadcaster.

⁸⁴ <https://www.telegraph.co.uk/news/2025/11/03/bbc-report-reveals-bias-donald-trump/>

⁸⁵ <https://committees.parliament.uk/committee/378/culture-media-and-sport-committee/news/210064/bbc-editorial-standards-cms-committee-writes-to-corporation-chair-over-panorama-editing-concerns-and-middle-east-coverage/>

The Trump edit was not a unique incident, for in certain fields that generate strong and opposing political views, the BBC has long been mired in controversy.

One is climate. Since 2018, the BBC has had a formal policy of excluding so-called “deniers” from its coverage. According to a memo from its head of news, “to achieve impartiality, you do not need to include outright deniers of climate change in BBC coverage, in the same way you would not have someone denying that Manchester United won 2-0 last Saturday. The referee has spoken.”⁸⁶ Meanwhile, the climate-sceptic analyst Paul Homewood has for the past three years compiled a catalogue of BBC climate reporting checked against primary observational sources.⁸⁷ It shows a recurring pattern: the assertion of a worsening trend that the custodians of the relevant data say does not exist.

For example, the BBC has reported that Atlantic basin hurricanes have become more powerful, with a rising proportion reaching major-hurricane intensity, but the US National Oceanic and Atmospheric Administration has stated that although projections suggest this will occur in future, there is as yet “no strong evidence [that] century-scale increasing trends” have become established.⁸⁸ The BBC reported “more than 4,500 deaths” caused by a 2022 UK heatwave, but the Office for National Statistics report that it cited said that in this and other periods of hot weather excess deaths were merely a “short-term mortality displacement, where deaths among vulnerable individuals are ‘brought forward’ to within the heat-periods”.⁸⁹ Two recent examples of climate-related misinformation carry particular weight because the BBC itself has admitted they were wrong. One was a claim that half a year’s rain fell in 36 hours during a storm in Bologna: the BBC admitted, after a complaint, that this was untrue – six months after it was broadcast.⁹⁰ Its assertions that polar bears are declining in number and that this is attributed by scientists to global warming and sea-ice loss also had to be corrected by the Executive Complaints Unit, which acknowledged that the global polar bear population has risen since

86 <https://www.carbonbrief.org/exclusive-bbc-issues-internal-guidance-on-how-to-report-climate-change>

87 P. Homewood, ‘BBC Climate Misinformation’ and ‘BBC Misinformation 2025’ (2025–2026), compilations supplied to the FSU. The items cited in the text are those anchored by an external primary source or by the BBC’s own complaints findings.

88 <https://www.gfdl.noaa.gov/global-warming-and-hurricanes/>

89 <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/excessmortalityduringheatperiods/englandandwales1juneto31august2022#main-points>

90 <https://www.bbc.co.uk/contact/ecu/weather-for-the-week-ahead-bbccouk>

the 1960s and is now stable.⁹¹ Homewood's dossiers contain many more examples.

Another controversial area is BBC coverage of the Middle East, a subject that Prescott examined at length in his memorandum. An egregious example was BBC reporting of an explosion in the courtyard of the Al-Ahli Arab Hospital in Gaza City on 17 October 2023, shortly after the terrorist attacks that killed some 1,200 people in Israel. The Hamas-run Gaza Health Ministry announced the hospital had been destroyed by an Israeli airstrike, leaving 500 staff and patients dead. Within the hour, the BBC correspondent Jon Donnison told viewers it was hard to see what else the explosion could have been. Near-identical wording was disseminated to millions by BBC push notifications. The BBC's international editor, Jeremy Bowen, endorsed the claim on the 10pm BBC1 news, blaming Israel for the fact that the hospital had been "flattened".⁹²

The consequences of these reports were serious. When they were broadcast, President Biden had been about to meet the leaders of Jordan, Egypt and the Palestinian Authority in Amman: in protest at the supposed Israeli attack on the hospital, the talks were cancelled. The claim made in Britain by supporters of the Palestinian cause that Israel was perpetrating genocide received a significant boost. Yet the BBC was guilty of disseminating rank misinformation. Subsequent Western intelligence assessments attributed the explosion to a failed rocket launched from within Gaza, and it soon became clear that the hospital had not been destroyed at all – with the death toll being far lower than Hamas's initial figure. More than a month later, the Executive Complaints Unit accepted "it was not consistent with the BBC's standards of due accuracy to offer any view about responsibility for the incident at a point where so little reliable information was available".⁹³

Bowen was unrepentant. Asked on the BBC programme *Behind the Stories* why he had said the hospital had been flattened, he said: "Oh yeah, well I got that wrong because I was looking at the pictures and what I could see was a square that appeared to be flaming on all sides and there was, sort of, a void in the middle... I was wrong on that but I don't feel particularly

91 <https://www.bbc.co.uk/contact/ecu/two-polar-bears-kill-canadian-worker-in-rare-attack-bbccouk>

92 <https://www.thejc.com/news/bbc-international-editor-jeremy-bowen-admits-gaza-hospital-report-was-inaccurate-jtluww3n>

93 <https://www.bbc.co.uk/contact/ecu/the-context-bbc-news-channel-17-october-2023-and-related-items>

bad about that. It was just the conclusion I drew.” As for wrongly attributing the explosion to an Israel strike, he said: “I don’t regret one thing in my reporting because I think I was measured throughout. I didn’t race to judgement.”

Yet the BBC report of the Al-Ahli strike is not the only correction the BBC has had to issue over its coverage of the current Middle East war. For example, it had to apologise for a misinformed report that the Israeli Defence Forces were “targeting” doctors and nurses at al-Shifa, another hospital in Gaza.⁹⁴ Another case was especially serious. In February 2025, the Corporation broadcast *Gaza: How to Survive a War Zone*, a documentary narrated by a boy of 13. The BBC did not tell viewers that he happened to be the son of a Hamas minister, a member of a terrorist organisation proscribed in UK law. Worse, this fact was not established by the BBC’s editorial machinery but by the independent researcher David Collier. Ofcom ruled that the BBC had breached the Broadcasting Code, and that the failure to disclose the fact that the boy’s father was a Hamas official was “materially misleading”.⁹⁵ It ordered the BBC to broadcast a summary of its findings.

It has been observed that in war, truth is often the first casualty. All journalists will sometimes make inadvertent mistakes. Yet Ofcom’s comments on this case bear directly on the Green Paper’s proposal to force platforms to privilege “trustworthy” sources: “Trust is at the heart of the relationship between a broadcaster and its audience, particularly for a public service broadcaster such as the BBC. This failing had the potential to erode the significantly high levels of trust that audiences would have placed in a BBC factual programme about the Israel-Gaza war... this represents a serious breach of our rules.”

Indeed, there is already evidence that trust in the BBC is at an unimpressive level. As the campaign group Newswatch noted last month, Ofcom’s Media Nations 2026 report records that only 33% of adults surveyed associate the BBC with impartiality and 45% with trustworthiness.⁹⁶

None of this proves that the BBC is worthless. But it makes the Green Paper’s proposal to privilege its output over other sources look questionable.

⁹⁴ <https://www.politico.eu/article/bbc-says-sorry-to-israel-after-reporting-it-targeted-medical-staff/>

⁹⁵ <https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-standards/ofcom-investigation-finds-bbc-gaza-documentary-in-breach-of-broadcasting-code>

⁹⁶ Ofcom, Media Nations 2026: <https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-adults/media-nations-2026>

10. Back to the future: restoring Charles I's censorship regime

In July 1637, in the reign of the autocratic monarch Charles I, the Star Chamber issued a *Decree concerning Printing*. It said nothing could be published unless licensed by approved, government-appointed authorities, with enforcement delegated to the Stationers' Company, a private body whose members' commercial monopoly accorded with the Crown's interest in suppression. In the eyes of the King, the new regime was a means to ensure that material deemed "seditious", "scandalous" or "offensive" would be stifled. These categories, it would become clear, could be defined in whatever way the licensors wished.⁹⁷ In view of the events that unfolded shortly afterwards, culminating in the King's execution in 1649, it could be said that the Decree did not realise its underlying purpose, of promoting political cohesion by suppressing dissent. Five years before he was beheaded, against the backdrop of tumult and civil war that led to that event, John Milton had inaugurated the great English tradition of free speech advocacy with *Areopagitica*, a denunciation of these regulatory measures. Not the least of his arguments was that the state should not treat adults like children but trust their judgement, even if that meant allowing them to be exposed to heterodox views.

The proposals now being considered are not an exact echo of the Caroline regime's, which remained in force in a somewhat altered form during the interregnum and afterwards – the last vestiges of statutory press regulation did not finally vanish until the Licensing of the Press Act 1662 expired in 1695. But there are uncomfortable parallels. Confronted by the rapid growth of a new information technology in just a few decades, the British Government proposes it should give exalted, licensed status to approved publishers while denying it to everyone else, along with giving itself powers to censor content in times of emergency or during elections, and educational programmes designed to teach children and adults what they should trust. Charles I tried to certify truth and privileged distribution to those he deemed

⁹⁷ Decree of Star Chamber concerning Printing, 11 July 1637.

trustworthy, and delegated policing to the 17th century equivalent of a fact-checking agency, the Stationers' Company. He justified these measures by the need to maintain harmony. They failed, and their modern analogues are likely to meet the same fate.

They will, if enacted, likely result in a damaging strategic cost. In the United States, the Biden administration flirted with internet censorship, and for a time, Imran Ahmed of the CCDH enjoyed privileged access to the White House and other organs of the state. Under Trump 2.0 and the free expression advocate Under Secretary of State for Public Diplomacy Sarah Rogers, veneration of the First Amendment has come roaring back. The Republican Ohio Congressman Michael Rulli is set to introduce a federal bill that would torpedo Ofcom and its burgeoning powers of censorship, the Guaranteeing Rights Against Novel International Tyranny and Extortion (GRANITE) Act. This would prohibit any American court from recognising or enforcing a foreign judgment, fine, order or demand made under a "foreign censorship law" that restricted free expression, and allow any American served with such a demand to have it declared a nullity. Nor would a federal Act arrive in a vacuum: state versions of the GRANITE Act have advanced in Wyoming and West Virginia; in July 2026 the American Legislative Exchange Council adopted its core provisions as national model policy; and the State Department confirmed as early as December 2025 that a federal version was likely to be debated in the House.⁹⁸

Most of the platforms that the present British Government wants to regulate are American-owned. In March, the US platform 4chan responded to a £520,000 fine levied by Ofcom for its alleged failure "to block under-18s' access to pornography" and other alleged violations with open ridicule.⁹⁹ If Rulli's bill is passed, both the OSA and any new UK regulations would be in direct conflict with American law. They would also create an open invitation for Washington to treat British ministers and regulators as adversaries of the American Constitution: not a great environment for a UK government that wants to see Britain emerge as a world-leader in AI.

Against all this, this briefing has sought to marshal evidence. The British

98 Byrne, P., 'The GRANITE Act: How Congress Can Strike Back Against Foreign Censors', 18 October 2025, with updates to 31 July 2026, <https://prestonbyrne.com/2025/10/18/the-granite-act-how-congress-can-strike-back-against-foreign-censors/>; Sarah Rogers, Under Secretary of State for Public Diplomacy, interview, GB News, 4 December 2025.

99 <https://www.bbc.com/news/articles/c624330lq1ko>

public is being asked to accede to the most far-reaching peacetime apparatus of speech control in modern history: a regime in which state-approved news is pushed into every feed by law; electoral discourse subjected to state-sponsored tests of accuracy and balance; emergency powers to remove content held in readiness for the next supposed crisis; the classification of truth delegated to lobby groups and unverified AI machines; and the public schooled from childhood in which voices to trust – on the strength of evidence that is weak where it is not absent, contradicted where it is not merely contested, and in its official form, manufactured to order: disproportionate measures in response to an unchecked moral panic.

A free people should not accept this, and it should not accept the misinformed claims on which it is based. Milton, not Charles I, should supply our policy blueprint. If we believe in truth, “we do injuriously, by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter? Her confuting is the best and surest suppressing.”¹⁰⁰

100 Milton, J., *Areopagitica*, 1644, <https://www.gutenberg.org/cache/epub/608/pg608-images.html>

