BREAKING THE ECHO CHAMBER: ENHANCING DISINFORMATION RESILIENCE IN THE UK

By DR THEO ZENOU and DR HELENA IVANOV

HJS
Henry Jackson Society

CENTRE FOR RESILIENT SOCIETY

DEMOCRACY | FREEDOM | HUMAN RIGHTS

November 2025

Published in 2025 by The Henry Jackson Society

The Henry Jackson Society
Millbank Tower
21-24 Millbank
London SW1P 4QP

Registered charity no. 1140489
Tel: +44 (0)20 7340 4520

www.henryjacksonsociety.org

Title: “BREAKING THE ECHO CHAMBER:
ENHANCING DISINFORMATION RESILIENCE IN THE UK”
By Dr Theo Zenou and Dr Helena Ivanov

£9.95 where sold

Cover image: Social media icons displayed on a smartphone with abstract disinformation text in background by JRdes at Shutterstock (https://www.shutterstock.com/image-photo/social-media-icons-logo-displayed-on-2378698417).

BREAKING THE ECHO CHAMBER: ENHANCING DISINFORMATION RESILIENCE IN THE UK

By DR THEO ZENOU and DR HELENA IVANOV

HJS
Henry
Jackson
Society

DEMOCRACY | FREEDOM | HUMAN RIGHTS

CENTRE FOR RESILIENT SOCIETY

November 2025

About the Authors

Dr Helena Ivanov is an associate research fellow at the Henry Jackson Society. She holds a PhD in International Relations at the London School of Economics and Political Science. Her research focuses on the relationship between propaganda and violence against civilians. In her thesis, Helena examined the role propaganda played during the Yugoslav Wars and produced a model for studying propaganda which details the key phases, functions, discourses, and techniques of propaganda (the model itself is applicable to other contexts). Additionally, Helena also served as a Manager at the Centre for International Studies at the LSE.

Prior to her PhD, Helena completed an MPhil in Politics: Political Theory at the University of Oxford, and holds a BA in Politics from the University of Belgrade.

Dr Theo Zenou is a research fellow at the Henry Jackson Society. He recently completed a PhD in history at Cambridge University, where he also taught undergraduates. He has been a Theodore Sorensen Fellow at the John F. Kennedy Presidential Library and a Visiting Researcher at Boston University.

Theo has written features, essays, and reviews for the *The Washington Post*, *The Economist*, *FT*, *The Guardian*, *The Times* and *The Sunday Times*, *TLS*, *The Spectator* and others.

Prior to his doctorate, he worked in TV for several years.

Acknowledgements

We thank Andrew Fox and Tal Hagin for their thoughtful review and valuable insights. We are also grateful to Zachary Joseph at Dataquest Analytics for his support in conducting the small-scale survey of UK students.

Contents

About Us

DEMOCRACY | FREEDOM | HUMAN RIGHTS

About The Henry Jackson Society

The Henry Jackson Society is a think-tank and policy-shaping force that fights for the principles and alliances that keep societies free, working across borders and party lines to combat extremism, advance democracy and real human rights, and make a stand in an increasingly uncertain world. The Henry Jackson Society is a company limited by guarantee registered in England and Wales under company number 07465741 and a charity registered in England and Wales under registered charity number 1140489.

For more information, please see **www.henryjacksonsociety.org**.

CENTRE FOR RESILIENT SOCIETY

About the Centre for Resilient Society

The **Centre for Resilient Society (CRS)** is a citizen-focused, international research centre within the Henry Jackson Society, which seeks to identify, diagnose and propose solutions to threats to the social resilience of liberal Western democracies.

The centre's work includes addressing the twin challenges posed by radicalisation and terrorism. The centre is unique in addressing violent and non-violent extremism. By coupling high-quality, in-depth research with targeted and impactful policy recommendations, it aims to combat the threat of radicalisation and terrorism in our society.

The centre's work also includes broader challenges of democratic resilience – including threats from both foreign interference and domestic issues. This includes the potential harm that various forms of social, cultural and political insecurity, conflict and disengagement can pose to the long-term sustainability of democracies, including the resilience of their institutions, public policy outcomes, citizens' health and wellbeing, and economic growth and prosperity. It also explores the balance between free speech and hate speech, and encourages respectful debate between those of different views, rather than cancellation. Moreover, it underscores how social and political instability can make nations vulnerable to internal and external actors seeking to deepen cleavages, undermine consensus and, ultimately, to weaken democratic functioning.

Executive Summary

Disinformation remains one of the most significant threats facing Western liberal democracies. It can distort public understanding of political events, deepen societal divisions, and undermine social cohesion. Its influence on electoral outcomes has been extensively documented. Yet democratic societies face a unique challenge: they must curb the spread of false information while safeguarding the fundamental right to free speech – a balance that has proven exceptionally difficult to strike.

This report examines how disinformation affects university students in the United Kingdom. Drawing on three distinct case studies – vaccination, the Southport stabbings, and Candace Owens' claim that Brigitte Macron is a man – we assess the reach and impact of misleading narratives. To underpin our analysis, we conducted a small-scale poll of 500 British university students.

Our findings are clear: disinformation exerts a strong and worrying influence. For example, more than 60% of respondents reported that information they encountered online would make them hesitant to vaccinate their children. Should even half of them act on this hesitancy, the UK risks falling below the herd-immunity threshold, with serious public-health consequences.

In response, this report outlines a series of policy recommendations designed to limit the spread and impact of disinformation while upholding the principles of free expression.

Introduction

Disinformation is fast becoming one of the most dangerous weapons of our time. It has proven capable of undermining public health – leading to a resurgence of diseases once thought eradicated by lowering vaccination rates. It has had a tangible impact on election campaigns across democratic states, revealing that even mature democracies are vulnerable to its corrosive effects. And as the old saying goes, "the first casualty of war is the truth" – disinformation has played a pivotal role in justifying violence, or at the very least, propping up regimes that support or perpetrate it.

The emergence and rapid advancement of artificial intelligence (AI) have made the challenge even more acute. Disinformation is expected to become increasingly sophisticated and harder to detect in the coming years, exacerbating its dangers and complicating efforts to counter it.

As previous research by the Henry Jackson Society has shown,[1] the United Kingdom is currently ill-prepared to confront this form of information warfare waged by hostile states and their proxies. Many individuals lack the basic tools and knowledge to verify information – particularly when it comes to using modern technologies such as reverse image search or software designed to identify deepfakes. Furthermore, social media algorithms often reinforce users' existing beliefs, creating echo chambers that intensify polarisation. Crucially, people are frequently unaware that they are even inside these echo chambers, making it difficult to address the problem.

The consequences of this are serious. Social polarisation poses a direct threat to the UK's social cohesion and democratic stability. Worse still, such divisions create opportunities for exploitation by adversarial actors like China and Russia, who actively seek to capitalise on democratic weaknesses.[2]

To strengthen the UK's democratic resilience and prepare for future challenges, it is vital that British citizens are equipped with the tools to recognise, question and resist disinformation. But first, we must understand how disinformation spreads and why people are so susceptible to it.

By better understanding the sources, mechanisms and consequences of disinformation, this research will provide a foundation for building a more informed, resilient and cohesive Britain.

[1] Helena Ivanov, "Disinformation on Campus: The Rise of Antisemitism and the Failure to Respond", The Henry Jackson Society, 18 May 2025, https://henryjacksonsociety.org/publications/disinformation-on-campus-the-rise-of-antisemitism-and-the-failure-to-respond/.

[2] For further details see: Patrick Wintour, "China spends billions on pro-Russia disinformation, US special envoy says", *The Guardian*, 28 February 2023, https://www.theguardian.com/world/2023/feb/28/china-spends-billions-on-pro-russia-disinformation-us-special-envoy-says; Chris Kremidas-Courtney, "Hybrid storm rising: Russia and China's axis against democracy", European Policy Centre, 2 May 2025, https://www.epc.eu/publication/Hybrid-storm-rising-Russia-and-Chinas-axis-against-democracy-64b158/; Ryan C. Berg and Henry Ziemer, "Assessing the Impact of China-Russia Coordination in the Media and Information Space", CSIS, 5 August 2025, https://www.csis.org/analysis/assessing-impact-china-russia-coordination-media-and-information-space.

Methodology

This paper will draw on a combination of primary and secondary research. A vast body of data already exists regarding people's reliance on social media platforms for information, as well as the volume and nature of disinformation circulating online. We will begin by reviewing this literature to establish preliminary insights into the types of sources people use and the extent of their exposure to disinformation.

In addition to this desk-based research, we carried out our own preliminary fieldwork. Specifically, we conducted a small-scale, anonymous quantitative survey with 500 students from universities across the UK. Of the respondents, 91.6% are UK nationals, 4.8% are from EU member states and 3.6% are international students (non-UK/non-EU). Moreover, 41.8% are currently pursuing an undergraduate degree and 45.0% are on postgraduate studies programmes, with 12.8% currently working towards their PhD degrees. Finally, 56.2% identified as female and 43.6% as male.

We sought to ensure that our poll reflected diversity both in students' fields of study and in their political leanings. In terms of subject areas, 27.6% are studying business and economics, 25.4% natural sciences and engineering, 15.0% social sciences and law, 12.8% arts and humanities and 11.0% health and medical sciences, while 8.2% selected 'other'. Politically, 15.8% identified as left-wing, 16.4% as centre-left, 24.6% as centrist, 16.2% as centre-right and 19.2% as right-wing. A further 6.0% reported having no political leanings, while 1.8% preferred not to say.

Our focus on this demographic is grounded in several considerations.

First, younger people are significantly more likely to rely exclusively – or predominantly – on social media for news and information. As such, they offer a valuable lens through which to understand the scale and impact of the issue. Second, as university students, they may be more capable of recognising disinformation, and we aim to test this assumption to assess whether – and to what extent – higher education offers a degree of protection against it. Finally, current students have largely grown up immersed in the era of digital media. Examining their levels of media literacy and their ability to identify or challenge disinformation is therefore especially relevant.

While our findings illuminate trends among digitally literate youth, they cannot be generalised to the wider population without further sampling.

Findings

The findings from our polling are deeply concerning and demonstrate that even a generation raised in the digital era struggles to identify disinformation. More worryingly, the long-term implications of this are likely to be significant. As expected, social media platforms and television remain the most relied-upon sources for political information. In terms of time spent, 30.4% of respondents reported using social media for between one and two hours a day to stay politically informed, 11.8% for more than two hours, 36.0% for between thirty minutes and one hour and 19.4% for less than thirty minutes.

When it comes to platforms, X (formerly Twitter) dominates, with 38.2% of respondents identifying it as their primary source of news. Facebook followed at 28.2%, while Instagram ranked third with 10.0%. Other platforms, such as TikTok and YouTube, attracted only single-digit percentages.

However, the real challenge arises when comparing respondents' confidence in spotting disinformation with their actual experience. When asked how confident they felt in identifying false or misleading information online, 34.4% said they were very confident and 53.0% somewhat confident. Yet, when asked whether they had ever believed something online that later turned out to be false, 16.8% admitted this happened to them often, 47.2% said it had happened a few times and 24.6% said once or twice. Only 8.6% claimed this had never occurred.

The problem also extends beyond students themselves. When asked about their families and friends, 16.0% of respondents said that their close contacts very often share fake news online, 27.0% said often and 26.6% said sometimes. These preliminary findings suggest that the challenge of disinformation is not confined to academic institutions but permeates wider social circles.

And finally, as our case studies will demonstrate, believing disinformation carries tangible real-world consequences, most clearly illustrated in the case of vaccines. Strikingly, 65.2% of respondents reported feeling very or somewhat concerned about vaccinating their children as a result of information encountered online. Needless to say, if even half of these individuals were ultimately to decide against vaccination, the UK would fall dangerously below the threshold required for herd immunity.

PART I – Scope and Scale of the Problem

The problem of disinformation

Social media platforms are now a fact of life. To illustrate:

> As of 2025, according to datareportal.com there are over 5.3 billion active social media user identities globally, representing approximately 65% of the world's population. […] On average, a user spends around 2 hours and 26 minutes on social media per day and interacts with 6.7 different platforms monthly. […] While Facebook retains its massive user base, particularly for groups and events, platforms like TikTok and Instagram are particularly dominant among younger audiences, and their growth rates reflect this. [3]

Zooming in on the UK specifically:

> As of January 2025: there are 67.8 million internet users in the UK, representing a 97.8% internet penetration rate. The UK is home to 54.8 million social media user identities, 79.0% of the total population. […And] over half of UK adults (52%) now use social media for news, up from 47% in 2023, highlighting social media's role as a news source. [4]

The UK's youth is also very connected. In February 2024, Ofcom, the online safety regulator, reported that:

> 99% of children spend time online; nine in 10 children own a mobile phone by the time they reach the age of 11; three-quarters of social media users aged between eight and 17 have their own account or profile on at least one of the large platforms; […] almost three-quarters of teenagers between age 13 and 17 have encountered one or more potential harms online. [5]

Even though both the young and old remain connected in the UK, a generational gap has emerged when it comes to how social media platforms are used. For example, young people tend to rely on social media as political news sources far more than their older counterparts. Specifically, in 2025, "58% of those aged under 35 primarily use social media to access political news [whereas] only 8% of those aged 55 and over primarily use social media for this purpose, with 46% using television instead." [6] Likewise, as outlined in the introduction, our polling data also show that social media platforms are increasingly becoming the primary source of information for those seeking to stay politically informed.

Thus, it is more than reasonable to assume that information circulating online is likely to impact the political attitudes of younger generations who primarily rely on social media to get politically informed.

What has also become evident is that social media does not simply inform young people politically – it increasingly enables them to shape political outcomes themselves. The BBC recently highlighted a group of young Britons, many still below voting age, who have nonetheless established themselves as influential political commentators online. Their reach is far from trivial. According to the BBC, dozens of these so-called "accidental influencers",

[3] "2025 UK Social Media Statistics", Metricool, 30 June 2025, https://metricool.com/uk-social-media-statistics/.

[4] Ibid.

[5] John Woodhouse and Maria Lalic, "The impact of smartphones and social media on children", House of Commons Library, 13 May 2024, https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0103.

[6] John Curtice, Marta Mezzanzanica and Katy Bailey, "Scrolling politics: engagement, trust and polarisation in the digital age", National Centre for Social Research, 25 June 2025, https://natcen.ac.uk/publications/bsa-42-politics-and-social-media.

most aged between 16 and 31, attract on average around half a million views per post. Some individual videos have drawn nearly a million views, and together their content has amassed more than 15 million views in total. [7]

The rise of social media as a primary source of news has coincided with a marked decline in public trust in mainstream media outlets. In 2024, according to the Edelman Trust Barometer, "The UK saw the biggest drop in trust in the media and was the least-trusted out of 28 countries surveyed [...] with 31% of people saying they trusted the media [...] a drop of six percentage points since the 2023 Trust Barometer." [8]

The erosion of trust in traditional outlets, coupled with growing dependence on social media, has created fertile ground for disinformation. This is largely because social media platforms maintain far weaker fact-checking standards than professional news organisations, which not only have greater resources to verify information but can also be held legally accountable for publishing falsehoods. By contrast, individual social media users face far fewer constraints in terms of the content that they can publish. Moreover, in recent months, platforms themselves have rolled back their fact-checking efforts.

Following Donald Trump's re-election as US president, many major platforms abandoned the use of independent fact-checkers altogether as some tech executives sought to improve their relations with the president. Most notably, Meta's founder Mark Zuckerberg announced that Facebook and Instagram would replace professional fact-checkers with "X-style 'community notes'" where users themselves would judge the accuracy of posts. [9] He defended this move by claiming that fact-checkers were "too politically biased" and argued that it was time to "get back to our roots around free expression". [10] Nevertheless, despite the scaling back of fact-checking, trust and confidence in social media platforms remain relatively high. According to our polling, 21.0% of respondents expressed a great deal of trust in these companies to remove disinformation and regulate content appropriately, while a further 42.6% reported having a fair amount of trust.

While one might reasonably argue that fact-checkers have, at times, been overly restrictive, swinging the pendulum entirely in the opposite direction is far from prudent. It runs the risk of a whole new kind of disinformation, whereby community notes themselves become a source of disinformation. Besides, in an environment where actors such as China and Russia actively conduct disinformation warfare through sophisticated bot networks, dismantling professional oversight is a deeply risky move.

This risk is amplified by the rise of AI, which enables malign actors to spread falsehoods more effectively and at scale. While it is true that AI can also be programmed to curb disinformation, it nonetheless increases the risk of falsehoods going viral. [11] For example, in 2024, the Center for Strategic and International Studies (CSIS) revealed that "Russia employed AI to create over 1,000 fake American profiles on social media, then used those profiles to spread anti-Ukraine, pro-Russian narratives in the United States. In short, Russia has coopted AI to lie." [12]

7 Marianna Spring, "How social media is making young people accidental election influencers", *BBC News*, 28 June 2024, https://www.bbc.com/news/articles/cd1rl6p5p32o.

8 Charlotte Tobitt, "Trust in media: UK drops to last place in Edelman survey of 28 nations", *Press Gazette*, 18 January 2024, https://pressgazette.co.uk/media-audience-and-business-data/trust-in-media-uk-edelman-barometer-2024/.

9 Liv McMahon, Zoe Kleinman and Courtney Subramanian, "Facebook and Instagram get rid of fact checkers", *BBC News*, 7 January 2025, https://www.bbc.com/news/articles/cly74mpy8klo.

10 Ibid.

11 For an example of AI being used to fight disinformation, see "Disinformation – Hamas-Israel War", Cyabra, 24 October 2023, https://cyabra.com/reports/disinformation-hamas-israel-war/.

12 Emily Harding, "A Russian Bot Farm Used AI to Lie to Americans. What Now?", CSIS, 16 July 2024, https://www.csis.org/analysis/russian-bot-farm-used-ai-lie-americans-what-now.

Likewise, China – traditionally slower than Russia in conducting effective disinformation campaigns – appears to be catching up, aided by advances in AI. According to *The New York Times*:

> A new technology can track public debates of interest to the Chinese government, offering the ability to monitor individuals and their arguments as well as broader public sentiment. The technology also has the promise of mass-producing propaganda that can counter shifts in public opinion at home and overseas. [13]

Thus, leaving fact-checking solely to platform users creates several problems. First, ordinary users are often unable to distinguish between genuine accounts and fake or automated bot accounts. Second, as both our previous and current research demonstrate, people often struggle to recognise disinformation when they encounter it. According to our findings, 16.8% of respondents admitted that they frequently believed something that later turned out to be false, 47.2% said this had happened to them a few times and 24.6% acknowledged it had happened once or twice. This means that highly sophisticated propaganda promoted by hostile actors may go unnoticed by users – yet still carry significant political consequences.

Finally, another major problem with the way social media platforms operate is their tendency to create echo chambers, where users are primarily exposed to accounts and content that reinforce their existing ideological positions. [14] Our research confirms this trend. When asked to what extent they agreed with the statement, "The news and information I see on social media is primarily from people who share my views," 19.6% of respondents strongly agreed and 44.2% agreed. By contrast, only 10.0% disagreed or strongly disagreed – a clear indication of the strength of these echo chambers.

More worryingly, our findings suggest that these echo chambers are not only the product of algorithms but are also deliberately constructed by users themselves. When asked whether they purposefully follow accounts that reinforce their political or social beliefs while avoiding those that challenge them, 27.2% said they did so to a great extent and 44.6% to a fair extent. This means that more than 70% of respondents admitted to creating echo chambers by choice.

As a result, individuals may repeatedly encounter information that is false but nevertheless serves to confirm their pre-existing beliefs. Research shows this is particularly dangerous: when strong believers are exposed to misinformation or disinformation that aligns with their worldview, attempts to counter it are often far less effective. [15]

And there we have it – the perfect storm. A landscape in which trust in mainstream media has eroded, driving people towards social media platforms that are not only scaling back their fact-checking systems but also have strong financial incentives to foster echo chambers. Within these echo chambers, disinformation not only flourishes but becomes far harder to challenge. Added to this are malign foreign actors – most notably Russia – now harnessing sophisticated AI technologies to generate and disseminate falsehoods with unprecedented speed and reach. The result is a troubling political reality: Western societies are increasingly polarised, less cohesive than in the past and populated by voters making decisions based on distorted or outright false information, with profound real-world consequences.

13 Julian E. Barnes, "China Turns to A.I. in Information Warfare", *The New York Times*, 6 August 2025, https://www.nytimes.com/2025/08/06/us/politics/china-artificial-intelligence-information-warfare.html.

14 Sruthi Dhulipala, "The echo chamber effect: How algorithms shape our worldview", Campaign, 27 September 2023, https://www.campaignasia.com/article/the-echo-chamber-effect-how-algorithms-shape-our-worldview/491762.

15 James Devitt, "Online Misinformation Most Likely to be Believed by Ideological Extremists, New Study Shows", NYU, 30 September 2024, https://www.nyu.edu/about/news-publications/news/2024/september/online-misinformation-most-likely-to-be-believed-by-ideological-.html.

It is clear that something must be done. Yet determining what and how remains fraught, not least because any measures risk infringing upon free speech – one of the cornerstones of democratic life. Make no mistake: the right to freely debate ideas with others is precious. It must be preserved at all costs. None of us want to live in a society where the government – or a corporation – tells us what we can and can't believe. State-sponsored truth is the stuff of dystopias.

But we should not use this as an excuse to let disinformation run amok. For disinformation itself is a threat to free speech. It poisons – and sometimes prevents –the free exchange of ideas. What is the point of debating anything when no one can tell the difference between truth and lies? We must bear in mind, therefore, that limitations on disinformation are not infringements of free speech. Instead, they safeguard free speech. This is comparable to the status of hate speech in European countries. Holocaust denial, for instance, is a crime in France. It is expressly forbidden. Would anyone in their right mind say that France does not have free speech?

A careful, balanced strategy is therefore essential. But before exploring possible solutions, it is necessary to demonstrate the scale of the problem by examining three conceptually distinct case studies of disinformation.

The first is vaccine disinformation – an area where accurate information is widely available, yet false narratives have nevertheless persuaded large numbers of people. The second is the Southport stabbings – an incident in which limited police communication created fertile ground for false claims to spread unchecked. The third concerns Candace Owens' claim that the French president's wife, Brigitte Macron, is in fact a man – a wholly fabricated story that gained such traction that Mrs Macron has been forced to pursue a high-profile defamation case, even submitting scientific evidence of her sex to the court.

Through these three case studies, we will show how disinformation thrives, and just how deeply it can shape perceptions and behaviour.

PART II – Case Studies

Vaccinations

Vaccines have saved millions of lives and remain one of humanity's greatest medical achievements. They are the reason many deadly diseases have been eradicated altogether. Yet, despite their undeniable role in improving health and longevity, recent years have seen a troubling resurgence of anti-vaccine sentiment – bringing back illnesses once thought consigned to history, such as measles.

Vaccine conspiracy theories are not new, but their modern form was shaped by two pivotal global moments. The first was the measles, mumps and rubella (MMR) vaccine scare of the late 1990s, triggered by Andrew Wakefield's now-discredited 1998 *Lancet* paper, which falsely claimed a link between the MMR vaccine and autism.[16] The second was the Covid-19 pandemic of 2020, which unleashed a wave of misinformation and disinformation on a scale never seen before.

Both events severely undermined public trust in vaccines, health authorities and medical expertise. Vaccination rates fell and diseases once eliminated began to resurface, as more people turned to misleading online content – including that promoted by figures with medical credentials – rather than evidence from established scientific institutions.

Now, that erosion of trust risks deepening further with President Trump's decision to appoint Robert F. Kennedy Jr, a long-standing vaccine sceptic, as US Health Secretary – a move that threatens to lend new legitimacy and momentum to the anti-vaccination movement. Ultimately, the rise of vaccine scepticism illustrates a deeper problem: even in an age of unprecedented access to accurate information, conspiracy theories still prove more persuasive to many.

The UK and vaccination rates

While global trends are deeply troubling, the situation in the UK is deteriorating at an alarming pace. Specifically, "the UK is the worst-performing G7 country for coverage of measles vaccines, as rates lag behind Europe." [17] It was only in "2016, [that] the UK had a major public health win when the World Health Organization (WHO) declared it had eliminated measles. But two years later this status was revoked as cases doubled." [18]

The outbreaks are a result of insufficient immunisation levels. According to the WHO and UNICEF:

> … the UK has just 89 per cent coverage of the first measles vaccine in 2024 and 85 per cent coverage of the second dose. This is down from 93 per cent a decade earlier and well below WHO target of 95 per cent, which it said is needed to effectively eliminate the spread of the disease in the community. [19]

Consequently, the UK found itself "in the midst of a measles emergency with 140 cases reported so far in 2024 alone." [20] Things have not improved much in 2025, with vaccination

16 For further details on this publication see: T.S. Sathyanarayana Rao and Chittaranjan Andrade, "The MMR vaccine and autism: Sensation, refutation, retraction, and fraud", *Indian Journal of Psychiatry* 53(2) (April to June 2011), https://pmc.ncbi.nlm.nih.gov/articles/PMC3136032/.

17 Rebecca Thomas, "Why the UK is lagging behind other countries on measles vaccinations", *Independent*, 15 July 2025, https://www.independent.co.uk/news/health/measles-mumps-rubella-vaccine-symptoms-child-dies-uk-b2789272.html.

18 Gustavo Corrêa, "How misinformation is fuelling the current measles crisis", Gavi, 26 January 2024, https://www.gavi.org/vaccineswork/how-misinformation-fuelling-global-measles-crisis.

19 Thomas, "Why the UK is lagging behind other countries on measles vaccinations".

20 Corrêa, "How misinformation is fuelling the current measles crisis".

rates still far below the 95% herd immunity threshold: "According to NHS data, there have been 529 confirmed cases in England [by July 2025], with 68% occurring in children under the age of 10." [21]

While there is no single data point clearly proving direct causality between vaccine refusal and the rise of vaccine disinformation, the data clearly shows that there is a strong correlation between the two. It is therefore our hypothesis that disinformation has had a negative impact on public health. It leads us to ask a simple set of questions:

How did this all start? Why do people no longer trust the vaccines? And why are vaccination rates dropping so significantly in the UK?

For starters, the now-infamous research linking the MMR vaccine to autism refuses to fade away. Although Dr Wakefield has long since lost his licence to practise medicine, and *The Lancet* formally retracted his study in 2010, trust in the MMR vaccine – and, by extension, in vaccines more broadly – has steadily declined. The second major factor was the outbreak of the Covid-19 pandemic in 2020. That moment marked a dangerous surge in the spread of misinformation and disinformation online, not only about the virus itself but also about the vaccines that experts had worked to develop in record-breaking time.

It is tempting to dismiss this as a matter of people gullibly believing everything they read online, but some understanding of their perspective is important if solutions are to be found. Most parents genuinely want the best for their children and find the stories about vaccines deeply unsettling – particularly when they are presented by individuals who either have, or claim to have, medical expertise. What complicates this further is that many of today's parents did not grow up in a digital world; they only encountered the internet and social media as young adults, often without the digital literacy needed to distinguish credible information from falsehoods. Against that backdrop, exposure to persistent online misinformation can be especially persuasive. When parents are continuously confronted with claims – now even echoed by the US Health Secretary – that vaccines are harmful or that their risks outweigh their benefits, and when this occurs on such a scale that herd immunity can no longer be achieved, it becomes clear that we are dealing with a systemic failure.

States and responsible institutions have not managed to mount an effective response to the scale of this disinformation campaign. In the words of Dr Elizabeth Whittaker, clinical lead for paediatric infectious diseases at Imperial College Healthcare NHS Trust, "exposure to misinformation online isn't counteracted by contact with a trusted healthcare professional." [22]

What is particularly troubling is that even younger generations – who are often assumed to be more adept at spotting disinformation due to their higher levels of digital literacy – are not immune to vaccine-related falsehoods.

The sheer prevalence of vaccine disinformation is striking. When asked whether they had come across unusual or unexpected claims about the Covid-19 vaccine – essentially, negative or misleading information – 60.2% of respondents said yes. More concerning still, this disinformation clearly shaped people's willingness to be vaccinated. While 73.6% of respondents went ahead with the jab, 23% did not, preventing the achievement of herd immunity. Among those who refused, 44.3% cited concerns over vaccine safety – precisely the talking points most commonly pushed by online disinformation campaigns.

[21] Annie Kelly, "Thursday briefing: Misinformation, access and cuts – the UK's measles surge explained", *The Guardian*, 17 July 2025, https://www.theguardian.com/world/2025/jul/17/thursday-briefing-misinformation-access-and-cuts-the-uks-measles-surge-explained.

[22] Ibid.

The picture becomes even more alarming when we turn to routine childhood vaccinations such as MMR and Hepatitis B. We asked respondents whether, based on information they had seen online, they would have concerns about vaccinating their children. A striking 30.2% said they would be very concerned, and a further 35% somewhat concerned. In other words, as many as 65% might hesitate or refuse to vaccinate their children – placing the UK dangerously below the threshold required for herd immunity. Even if only half followed through on such hesitancy, uptake of the MMR jab would still fall below 70%, with serious consequences for public health.

The story of vaccine disinformation and declining vaccination rates shows that evidence alone is not enough to protect public trust. Even when a topic is well researched and overwhelming data exists to prove the safety and effectiveness of vaccines, if accurate information is not communicated clearly, accessibly and persuasively, it risks being drowned out by disinformation. Countering falsehoods therefore requires more than simply presenting the facts – it demands a carefully thought-through strategy for how, when and by whom accurate information is delivered, so that it resonates with the public and restores confidence.

Southport stabbings

The Southport case represents a stark example of how disinformation can escalate into violence and disorder. Unlike vaccine conspiracies – where accurate data and extensive scientific evidence were available but proved insufficient to counter the spread of falsehoods – the danger here lay in the absence of clear and credible information. Because the police were legally required to withhold the attacker's identity, a vacuum of knowledge emerged. That uncertainty provided fertile ground for malicious actors online, who quickly stepped in to fill the gap with false claims and inflammatory speculation. These narratives gained traction precisely because they offered apparent certainty in a moment of confusion, spreading rapidly across digital platforms and stoking public anger. In turn, this helped transform an already tense situation into one marked by violent riots. The case demonstrates how disinformation thrives not only when truth is poorly communicated but also when silence, however necessary, leaves space for lies to dominate.

On 29 July 2024:

> Three young girls [were] killed in a 'ferocious' knife attack in Southport at a Taylor Swift-themed dance and yoga event. […] Eight more children were injured, with five left in a critical condition […] Two adults also suffered critical injuries […] Armed officers detained a male and seized a knife, with police later saying that a 17-year-old boy had been arrested on suspicion of murder and attempted murder. The attack [was] not being treated as terror-related, police said. [23]

In the immediate aftermath of the attack, it became clear that only a limited amount of information about the attacker could be shared with the public. This was partly because he was underage, but also because:

> The UK has strict laws on what information can be published about suspects in criminal cases before they go on trial. This is so the jury in their future trial – who are 12 ordinary members of the public – do not read or see anything that would allow them to pre-judge the case before it has been heard in full at court. [24]

[23] Jeremy Culley and Hafsa Khalil, "Southport stabbings – what we know about attack", *BBC News*, 29 July 2024, https://www.bbc.com/news/articles/cy68z9dw9e7o.

[24] Josh Halliday, "Southport attack: why is so much information being published only now?", *The Guardian*, 21 January 2025, https://www.theguardian.com/uk-news/2025/jan/21/southport-attack-why-is-so-much-information-being-published-only-now.

Thus, according to a BBC report updated on 31 July 2024, the only information made public was that he was a "17-year-old boy from Banks [...] originally from Cardiff [...] [and that] his parents are from Rwanda and he has an older brother also born in Cardiff." [25]

This scarcity of official information created fertile ground for disinformation to flourish. In the immediate aftermath of the stabbing, "A flood of misinformation [...] has been spread on numerous social media platforms by sources ranging from far-right activists to fake news websites and conspiracy theorists." [26] Although government officials urged the public to refrain from circulating false claims about the incident, the disinformation quickly took on a life of its own, spreading like wildfire and ultimately provoking "a protest by hundreds of far-right activists, believed to be supporters of the English Defence League [during which] missiles [were] thrown at police and a local mosque [was] attacked." [27] In the coming days, protests and riots exploded across the country.

What this case illustrates is not only the capacity of disinformation to fuel violence, but also the extraordinary speed at which falsehoods can spread. A striking example was a LinkedIn post by local resident Eddie Murray, who inaccurately claimed that the perpetrator was a migrant. His post, initially attracting only modest attention, quickly became one of the most widely viewed social media posts about the stabbing. It was rapidly replicated across platforms, including X, where it was circulated by accounts promoting anti-immigration measures. [28] Within hours, it had garnered more than two million views and was even cited by international news outlets as though it were verified.

This episode highlights how a single unsubstantiated claim, originating from a local source, can be amplified through social media ecosystems into a widely accepted false narrative. Ultimately, according to BBC Verify, "a screenshot of Eddie Murray's original LinkedIn post was viewed more than three million times on X." [29] When asked why he shared the information, he responded that "he had only posted what he had been told, genuinely believing his posts were correct." [30] In the end, the riots continued for weeks, leading to numerous arrests and serving as a stark reminder of the power of disinformation to drive radicalisation and violence.

Our polling further underscores just how powerful disinformation proved to be in the Southport case. An overwhelming majority of our respondents reported exposure to false claims surrounding the incident. In particular, 60% said they had seen stories online alleging that the perpetrator was a Muslim asylum seeker who had arrived in the UK on a small boat – demonstrating how quickly disinformation can spread and go viral.

Crucially, these narratives did not merely circulate; they also shaped beliefs. When asked whether they thought the perpetrator was in fact a Muslim asylum seeker who had come to the UK by small boat, 21% of respondents said they believed it completely, while a further 36% said they believed it somewhat. In other words, more than half of respondents accepted the falsehood to at least some degree.

The data also reveal another key dimension of disinformation: even when it fails to fully convince, it can sow confusion. A striking 19.2% of respondents said they were unsure what

25 Culley and Khalil, "Southport stabbings – what we know about attack".

26 Ben Quinn, "Misinformation about Southport attack suspect spreads on social media", *The Guardian*, 30 July 2024, https://www.theguardian.com/uk-news/article/2024/jul/30/misinformation-southport-attack-suspect-social-media-conspiracy-theories.

27 Ibid.

28 Ed Thomas and Shayan Sardarizadeh, "How a deleted LinkedIn post was weaponised and seen by millions before the Southport riot", *BBC News*, 25 October 2024, https://www.bbc.com/news/articles/c99v90813j5o.

29 Ibid.

30 Ibid.

to believe, leaving only 23.8% who firmly rejected the claim. This matters, because when large sections of the public either accept or are inclined to accept falsehoods, disinformation gains the power to polarise, to fuel mistrust and – at its most dangerous – to spark unrest and violence.

The Southport case illustrates how dangerous an absence of information can be in moments of crisis. When only limited facts are available, uncertainty creates fertile ground for speculation and falsehoods. In such circumstances, disinformation not only spreads quickly but can also take on the appearance of authority, filling the gaps left by official silence. This dynamic makes it far easier for misleading narratives to gain traction and, as the events in Southport showed, those narratives can fuel anger, mobilise crowds and ultimately spill over into violence.

The Brigitte Macron Hoax

In our disinformation era, even simple truths are contested. The old adage was that you need to see it to believe it. But this is no longer the case. We no longer believe what we see. As a society, we have fallen down the rabbit hole and now roam in a wonderland of alternative facts. The very idea of reality is up for debate. [31]

Perhaps nothing better exemplifies this than the ludicrous – yet widely circulated – claim that Brigitte Macron, the First Lady of France, is secretly a transgender man. The Macron hoax shows how a single falsehood, born in obscure corners of the internet, can become a global conspiracy – in this case, to the point that Ms Macron herself and her husband, President Emmanuel Macron, have been forced to take legal action. [32]

As far as we know, the Macron hoax originated in a "small far-right newsletter". But it first went viral in December 2021. [33] France, like the rest of the world, was in the grip of the Covid pandemic. As we saw, the pandemic unleashed a torrent of disinformation online, including anti-vaccine disinformation. Trust in politicians was also low. Conspiratorial thinking thrived. And who better to spread lies about than the wife of the man running the country?

According to *The Guardian*'s Angelique Chrisafis, a self-described "spiritual medium" who went by the name Amandine Roy played a key role in spreading the Macron hoax. [34] In December 2021, Roy conducted a lengthy YouTube interview with Natacha Rey, a small businesswoman-turned-investigative journalist. Rey trumpeted, rather proudly, that she was "self-taught, and not from the inner circle of mainstream media".

For years, Rey went on, she had been investigating an explosive story. It was so big it threatened the French Republic itself. Brigitte Macron was born a man. But there was more. Ms Macron was actually her brother, Jean-Michel Trogneux. She had not only lied about her gender, but also about her identity. [35]

Since then, Roy and Rey have been sued for defamation by Ms Macron and by her brother, Jean-Michel Trogneux. They were found guilty but were acquitted by an appeals court. This

31 See the work of Jean Baudrillard for reflections and analysis on this issue. Christopher Bray, "Revelling in illusion: the French sociologist-cum-philosopher who hit peak absurdity back in 1991", *The Spectator*, https://www.spectator.co.uk/article/revelling-in-illusion-the-french-sociologist-cum-philosopher-who-hit-peak-absurdity-back-in-1991/.

32 For an overview of the case study, see Lucas Minisini, "Who is Candace Owens, the pro-Trump influencer spreading transphobic fake news about Brigitte Macron?", *Le Monde*, 13 March 2025, https://www.lemonde.fr/en/m-le-mag/article/2025/03/13/who-is-candace-owens-the-pro-trump-influencer-who-is-spreading-transphobic-fake-news-about-brigitte-macron_6739123_117.html.

33 Angelique Chrisafis, "The Macrons v Candace Owens: lawsuit marks new phase in battle against conspiracy theories", *The Guardian*, 26 July 2025, https://www.theguardian.com/world/2025/jul/26/emmanuel-brigitte-macron-candace-owens-conspiracy-theory-jean-michel-trogneux.

34 Ibid.

35 Ibid.

doesn't mean, of course, that Roy and Rey were right, instead, "the Paris Appeals Court [...] let them off, arguing they had made the mistake in 'good faith'." [36]

But the damage had been done. The YouTube video took off on social media. Hours after it was posted, it hit half a million views. The hashtag #JeanMichelTrogneux began trending on French social media. It was pushed by far-right accounts as well as anti-vaccine accounts – a reminder that conspiracists often assist each other.[37] In this case, this wasn't only true in France. In this case, this was not only true in France: conspiracy theorists around the world did not wait long to jump on the bandwagon.

Last year, the popular far-right American influencer Candace Owens – known for her conspiratorial rhetoric – released a series titled *Becoming Brigitte* on YouTube. Based on a book by Xavier Poussard, a self-described French journalist, it recycles the Macron hoax. Hosted by Owens and formatted like a Netflix series, with cliffhangers and twists, *Becoming Brigitte* is conspiracy entertainment. And it has reached millions of viewers across the world. The effect was to transform a niche French hoax into a global *cause célèbre*.

Owens reportedly benefits financially from the Brigitte hoax. In addition to any revenue from the show itself, she has also sold merchandise online. As far as Owens is concerned, however, she is a principled rumourmonger speaking truth to power. She is on a crusade against "the perverts that run the world".[38] In her telling, far from being a fake news merchant, she is the only sane journalist left in the world.

The Macron hoax, however, is not just a vehicle to promote Owens. It also has broader political and societal implications. For one, the hoax has been used to delegitimise Macron's presidency and centrist policies. It should come as no surprise, then, that the European and American far-right has eagerly spread the hoax. As has Russian media.[39] For another, the hoax is also misogynistic and transphobic. It weaponises gender and sexuality to discredit women in the public eye, drawing on the same conspiratorial vocabulary used against figures such as Michelle Obama or Jacinda Ardern. These narratives only increase prejudice.

By July 2025, the campaign had reached such intensity that the Macrons decided to act. They filed a defamation lawsuit in Delaware, where Owens' companies are incorporated. In a statement, the Macrons said Owens' "campaign of defamation was plainly designed to harass and cause pain to us and our families and to garner attention and notoriety". They added that they had given her "every opportunity to back away from these claims, but she refused." [40] In *Paris Match*, Emmanuel Macron explained why he was determined to pursue legal action: "This is about defending my honour! Because this is nonsense. This is someone who knew full well that she had false information and did so with the aim of causing harm, in the service of an ideology and with established connections to far-Right leaders." [41]

[36] "Court overturns libel convictions of two women who said French first lady was a man", *Le Monde*, 10 July 2025, https://www.lemonde.fr/en/france/article/2025/07/10/french-court-clears-defendants-in-french-first-lady-brigitte-macron-s-gender-rumours-trial_6743245_7.html.

[37] Chrisafis, "The Macrons v Candace Owens".

[38] Chad de Guzman, "Candace Owens Says She Won't 'Shut Up' After Defamation Lawsuit: 'Brigitte Macron Is Definitely a Man'", *Time*, 24 July 2025, https://time.com/7305131/candace-owens-brigitte-macron-man-emmanuel-wife-defamation-lawsuit-response/.

[39] Kaye Wiggins and Leila Abboud, "Macrons hired investigators to research US influencer Candace Owens", *Financial Times*, 11 August 2025, https://www.ft.com/content/986e658e-7b52-4d8e-ac44-2b6978e462d7.

[40] Mike Wendling, "Macrons file defamation suit against US influencer Candace Owens", *BBC News*, 23 July 2025, https://www.bbc.co.uk/news/articles/c8739w8py4jo.

[41] Anoushka Mutanda-Dougherty, Melanie Stewart-Smith and Victoria Farncombe, "Macrons to offer 'scientific evidence' to US court to prove Brigitte is a woman, lawyer says", *BBC News*, 18 September 2025, https://www.bbc.co.uk/news/articles/ckg3llj5nxdo.

Owens responded by ridiculing the lawsuit. She called it a "catastrophic PR strategy" and said the Macrons were "defaming me because you don't want people to continue to watch the series".[42] In other words, Owens held up the lawsuit as proof she was right in the first place. Only if she was speaking the truth would she ever be sued – or so her thinking went. The position, however, ignores the fact that defamation is a common ground for a lawsuit.

The case has now become a test of how democracies can respond to digital falsehoods that transcend borders. The Macrons' lawyers have stated they are prepared to submit photographic and scientific evidence to prove that Brigitte Macron is a woman. That such evidence is even deemed necessary shows how thoroughly the disinformation has distorted public discourse.[43]

Our polling confirms how widely the rumour has circulated. When asked, *"Have you heard about or seen the claim that Brigitte Macron is a man?"* 60.4% of respondents said yes. Barely two in five had not encountered it. For a claim targeting a foreign first lady, that level of visibility is extraordinary – a measure of how efficiently social media can propel even the most implausible conspiracies into global consciousness.

Belief in the hoax was also significant. When asked how much they believed the claim, 17.6% of respondents said they believed it *completely* and 28.6% said they believed it *somewhat*. Only 38.2% rejected it outright, while 15.6% said they were unsure. In other words, almost half of the respondents accepted the falsehood to some degree, and another one in six were uncertain.

The numbers matter because they show how disinformation can spread. Even when the truth is eventually established in a court of justice, the lies will remain in the court of public opinion.

[42] de Guzman, "Candace Owens Says She Won't 'Shut Up'".

[43] Mutanda-Dougherty, Stewart-Smith and Farncombe, "Macrons to offer 'scientific evidence' to US court".

Part III – Current Policy Landscape

Fighting Disinformation Online: Where We Are Today

Disinformation is one of the most complex policy challenges facing liberal democracies. Governments face a delicate balancing act. On the one hand, they must curb the spread of false or harmful content. On the other, they must do so without undermining free expression. Social media has become the main battleground in this struggle. And disinformation has an edge: falsehoods spread faster than institutions can correct them. The reason is simple. Algorithms are designed to increase user engagement by any means necessary. The accuracy of the content they boost to achieve that is, mostly, irrelevant. [44]

Clearly, more needs to be done in the fight to regulate disinformation. But before imagining what *could* be done, we must first examine what *is* being done.

United Kingdom

In the United Kingdom, disinformation falls under the 2023 Online Safety Act. Its goal is not strictly to fight disinformation, however, but to make the internet safer – especially for children.

According to the Government, the Act "puts a range of new duties on social media companies and search services, making them more responsible for their users' safety on their platforms." [45] It applies to any service that hosts user-generated content, from social media and video-sharing platforms to search engines and dating apps. Moreover, its scope extends to non-UK companies whose services are accessed by UK users. [46] This means that it applies to Silicon Valley platforms such as Facebook or Instagram and to Chinese-backed TikTok.

Responsibility for enforcing the Act lies with Ofcom, the UK's media and communications regulator. Ofcom is tasked by the Online Safety Act with writing detailed codes of practice, monitoring compliance and carrying out investigations. It has the power to fine companies up to £18 million – or 10% of its global revenue. In extreme cases, it can even block access to non-compliant platforms. Senior managers can also face personal criminal liability if their companies repeatedly fail to meet "statutory duties". [47] While the law's primary focus is illegal material, such as child exploitation, terrorism or fraud, it does address disinformation indirectly. Large "Category 1" services (e.g. Facebook, X, TikTok and YouTube) must assess the risk of their algorithms amplifying false or misleading information. They are legally required to remove any content prohibited by their own terms of service and to ensure that those rules are enforced consistently. [48]

To guide this process, Ofcom has established a dedicated Advisory Committee on Disinformation and Misinformation, which met for the first time in May 2025. The committee's role is to help Ofcom interpret how false or misleading information online interacts with the broader duties set out in the Act. [49]

[44] See Smitha Milli, et al., "Engagement, user satisfaction, and the amplification of divisive content on social media", *PNAS Nexus* 4 (March 2025), https://academic.oup.com/pnasnexus/article/4/3/pgaf062/8052060.

[45] "Online Safety Act: explainer", Department for Science, Innovation & Technology, 24 April 2025, https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer#how-the-online-safety-act-is-being-implemented.

[46] Ibid.

[47] "Ofcom's approach to implementing the Online Safety Act", Ofcom, 26 October 2023, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation.

[48] "Online Safety Act: explainer".

[49] "Ofcom establishes Online Information Advisory Committee", Ofcom, 28 April 2025, https://www.ofcom.org.uk/about-ofcom/structure-and-leadership/ofcom-establishes-online-information-advisory-committee.

Under the new framework, the largest online services must conduct and publish regular risk assessments. They must also provide users with tools to control what they see. This includes filters for harmful but legal content, such as self-harm or eating-disorder material. Ofcom will issue codes of practice specifying how these duties must be met. Once laid before Parliament, those codes will become legally binding. [50]

The Online Safety Act offers a new model of regulation. It rightly treats social media platforms not as neutral conduits but as publishers with responsibilities. But it is too soon to know whether it will succeed. Will it make a lasting difference – or will it be too slow to catch up with social media companies?

Other Democratic Countries

These countries all have different legal codes from one another. Policies adopted in one country couldn't, for instance, be copied and pasted in the UK. Nonetheless, they are worth examining. They show us that there are a wide range of approaches to fighting disinformation. It will be up to policymakers to be nimble and to pick the most effective ones.

Brazil

Brazil is among the democratic countries that most regulates disinformation. This is made possible by its constitution, which considers "anti-democratic speech", such as denying the results of a fair election, to be a form of hate speech. As a result, Brazil was able to prosecute supporters of Jair Bolsonaro who wrongly claim that the 2022 election was stolen from him. [51]

Moreover, earlier this year, the Brazilian Supreme Court revised the Civil Rights Framework for the internet, which was first signed more than a decade ago. Specifically, it ruled that social media platforms can be held liable for illegal or harmful material, including hate speech, child pornography and attacks on democratic institutions. [52] The decision reframes platforms as actors with legal obligations to prevent harm. Advocates argue that this will increase transparency and accountability; critics fear it will enable censorship. [53]

United States

The United States has a different legal culture and a historical aversion to speech limitations. The first amendment, which protects freedom of speech, is considered sacrosanct. Only threats and incitement to violence are not protected.

This explains why there has been no major legislation regulating social media platforms. Not that there has been no attempt. The Kids Off Social Media Act was proposed in the Senate but never passed. It would have banned social media use for under-13s and outlawed personalised recommendation algorithms for users under 17. [54] There have also been proposals at the state level. In New York, Attorney General Letitia James has proposed mandatory age verification, parental consent for minors and new parental controls to restrict time online. [55] Given the

[50] "Ofcom's approach to implementing the Online Safety Act".

[51] Theo Zenou, "Dear Biden, stop backing down from the fake news wars", *The Boston Globe*, 21 June 2024, https://www.bostonglobe.com/2024/06/21/opinion/brazil-united-states-disinformation/.

[52] "Brazil's Supreme Court to Strike Down Internet Governance Safe Harbor Clause", *Covington*, 20 June 2025, https://www.cov.com/en/news-and-insights/insights/2025/06/brazils-supreme-court-to-strike-down-internet-governance-safe-harbor-clause.

[53] Débora Prado, "Brazil explores ways to hold platforms accountable in court and beyond", Association for Progressive Communications, 20 August 2025, https://www.apc.org/en/news/brazil-explores-ways-hold-platforms-accountable-court-and-beyond.

[54] "S.4213 - 118th Congress (2023-2024): Kids Off Social Media Act", Congress.gov, 30 April 2024, https://www.congress.gov/bill/118th-congress/senate-bill/4213.

[55] Letitia James, "Protecting Children Online", New York State Attorney General, 15 September 2025, https://ag.ny.gov/resources/individuals/consumer-issues/technology/protecting-children-online.

First Amendment, the focus in the US is less on disinformation than on protecting children's mental health.

Australia

In 2024, Australia became the first country to ban social media for under-16s. The law, championed by Prime Minister Anthony Albanese, will take effect in December 2025. At an event during the UN General Assembly, Albanese said the ban was "a crucial step in the right direction," adding that the "challenge we face is constantly evolving". [56] The Government argues that the law will limit the mental health harms linked to teenage social media use. Again, this is much broader than disinformation, but it will in effect limit teen exposure to disinformation.

Rather than rely on blanket ID checks, the law will require platforms to use "artificial intelligence" and "behavioural data" to estimate users' ages. As Albanese sees it, this policy will give Australian teenagers "three more years of being shaped by real-life experience, not algorithms".

The move has attracted plaudits. European Commission President Ursula von der Leyen said she was "inspired by Australia's example" and that Europe was "watching and learning". It has also come under strong criticism. Commentators have pointed out that it limits free speech and cannot be enforced. Young users can avoid the ban by simply using VPNs, which can be legally downloaded in minutes. [57]

India

Since 2023, India has tightened its content moderation laws. The Government now has the power to order content to be taken down. "Fake news" creators are directly concerned. When X challenged the law, India's Karnataka High Court dismissed the case. "Every platform that seeks to operate within the jurisdiction of our nation, which they do, must accept that liberty is yoked with responsibility," said Judge M. Nagaprasanna. [58] The Government defended its stance by arguing that the measures "tackled a proliferation of unlawful content and ensured accountability online". Critics, however, accused it of enabling political censorship. [59]

[56] "Australia's social media ban for teens draws praise at UN", *Reuters*, 25 September 2025, https://www.reuters.com/business/media-telecom/australias-social-media-ban-teens-draws-praise-un-2025-09-25/.

[57] Hannah Ritchie, "Australia approves social media ban on under-16s", *BBC News*, 28 November 2024, https://www.bbc.com/news/articles/c89vjj0lxx9o; Alasdair Pal and Cordelia Hsu, "Australia's under-16 social media ban sparks anger and relief", *Reuters*, 29 November 2024, https://www.reuters.com/world/asia-pacific/australian-pm-albanese-says-social-media-firms-now-have-responsibility-protect-2024-11-28/.

[58] Arpan Chaturvedi and Munsif Vengattil, "Musk's X loses bid to quash Modi's new content removal system", *The Star*, 24 September 2025, https://www.thestar.com.my/tech/tech-news/2025/09/24/indian-court-dismisses-x039s-petition-about-content-removal-mechanism.

[59] Ibid.

Part IV – Recommendations

Fighting Disinformation Online: Where We Should Go

Prescribing policy recommendations for disinformation is no easy task. But it's precisely for this reason that we must try. Disinformation threatens the very foundation of our free societies. It undermines the democratic process and sows distrust between citizens and their governments. Of course, there are no silver bullets against disinformation. No single policy can solve the problem once and for all.

Our aim here is to propose practical policies that can make a meaningful difference. When it comes to combatting disinformation, a 'one size fits all' approach will not work. The balance between risks and benefits varies significantly between democratic societies and repressive regimes, where individuals may rely on online platforms as one of the few avenues available to share the truth.

For this reason, the policies we recommend are intended exclusively for the United Kingdom. Other countries will require tailor-made approaches that reflect their specific political, social and technological circumstances.

SHORT-TERM RECOMMENDATIONS

Issue Public "Disinformation Weather Reports". The Government should publish regular public briefings charting emerging disinformation trends – just as the UK's National Cyber Security Centre (NCSC) does for cyber-attacks. The model already works elsewhere: NATO's Strategic Communications Centre of Excellence in Riga and the EU Hybrid COE in Helsinki both track foreign information manipulation. [60] Making such data public would help citizens see which topics are being targeted and by whom. Disinformation should be treated like pollution. Obviously, some concerns remain about whether governments could possibly abuse this policy – to derail narrative-driven campaigns that do not align with their policies. That is exactly why, in the first instance, we only propose this policy for the UK – a well-established democracy, where multiple other checks and balances are likely to prevent a government abusing its power.

Default to "Following-Only" Feeds. The EU Digital Services Act (Article 38) forces platforms to offer users at least one non-profiling feed. [61] The UK Government could go further by making that "Following" feed the default instead of algorithmic "For You" pages. Engagement-based algorithms increase polarisation by amplifying emotional and misleading content. [62] A following-only default would not solve the problem of echo chambers, but it would help restore user control and transparency. Obviously, there are benefits to "For You" feeds – for one, our data clearly shows that people tend to create echo chambers by choice. Subsequently, the "For You" feed can, in a way, help burst the bubble by exposing the user to some other content as well. For that specific reason, we recommend keeping the "For You" feed, but not making it the default.

[60] Gundars Bergmanis-Korāts, Marija Isupova and Raitis Ralfs Vecmanis, "Virtual Manipulation Brief 2025: From War and Fear to Confusion and Uncertainty", NATO Strategic Communications Centre of Excellence, 2 June 2025, https://stratcomcoe.org/publications?tid[]=59.

[61] "The Digital Services Act package", European Commission, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.

[62] Milli, et al., "Engagement, user satisfaction, and the amplification of divisive content on social media".

Introduce Screen-Time and "Digital Wellbeing" Caps. The Liberal Democrats have proposed a two-hour "doomscrolling cap" for under-18s and cigarette-style health warnings.[63] This is a step in the right direction. Platforms should be required to include screen-time tools that trigger "pause and reflect" prompts after a set limit. Such design nudges will reduce exposure to disinformation by cutting endless scrolling.

Introduce a "Civic Boost" Algorithm. Platforms should prioritise verified, public-interest content during major events such as elections or political crises. Google's SOS Alerts during natural disasters already provide a template.

Create a Media Resilience Fund. Disinformation thrives where journalism collapses. Australia's News Media Bargaining Code (2021) and Canada's Online News Act (2023) require platforms to compensate publishers.[64] The UK should go further by introducing a Digital Ad Responsibility Levy, a small tax on tech advertising revenue to be used to fund public-interest journalism and local reporting. This would strengthen the information ecosystem that disinformation exploits.

MEDIUM-TERM RECOMMENDATIONS

Launch an International "Digital Geneva Convention". In 2017, Microsoft proposed a "Digital Geneva Convention" to set norms against state-backed cyber operations.[65] The concept should be revived and expanded to cover disinformation campaigns. Disinformation is a form of hybrid warfare; it demands an international legal response.

Institutionalise "Prebunking" Campaigns. Experiments by the University of Cambridge show that "prebunking" – teaching people to recognise manipulation tactics before they encounter them – limits the spread of disinformation.[66] The Government should fund permanent prebunking campaigns, taking inspiration from Google Jigsaw's Project Shield.[67] These would act like vaccines for the mind, creating early immunity to falsehoods.

Ban Social Media for Young Users. In Australia, Prime Minister Anthony Albanese's Government passed the world's first social media age ban for under-16s in 2024. The UK should follow suit. Beyond child safety, this would also limit early exposure to disinformation and extremist material.

Embed Digital and Media Literacy in Education. Media literacy should become a core subject at schools and universities. Pupils and students must learn how algorithms shape what they see and how to fact-check before sharing.

Phase Out Anonymous Accounts. Anonymity enables trolling and viral falsehoods. Individuals should no longer be allowed to create anonymous accounts and ID checks should be conducted before obtaining a social media account. While this will not solve the problem of disinformation, it will limit its spread. Of course, we recognise the important role that whistleblowers and others who rely on anonymity play in resisting repressive regimes. It is precisely for this reason that

63 Paul Seddon, "Lib Dems want health warnings on teens' social media", *BBC News*, 22 September 2025, https://www.bbc.co.uk/news/articles/cp3vdw7wv2go.

64 "News media bargaining code", Australian Competition & Consumer Commission, 19 May 2020, https://www.accc.gov.au/by-industry/digital-platforms-and-services/news-media-bargaining-code/news-media-bargaining-code; "The *Online News Act*", Government of Canada, https://www.canada.ca/en/canadian-heritage/services/online-news.html.

65 "The need for a Digital Geneva Convention", Microsoft, 14 February 2017, https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/.

66 Jon Roozenbeek, et al., "Psychological inoculation improves resilience against misinformation on social media", *Science Advances*, 24 August 2022, https://www.science.org/doi/10.1126/sciadv.abo6254.

67 "Google Jigsaw: Project Shield & Prebunking", Vision of Humanity, 15 September 2025, https://www.visionofhumanity.org/google-jigsaw-project-shield-prebunking/.

we recommend implementing this policy exclusively in the UK – a well-established democracy where whistleblowers do not face the same dangers they would in authoritarian states.

We further propose that anonymous accounts be gradually phased out, allowing sufficient time to monitor, assess and adjust should any unforeseen challenges arise during implementation.

We acknowledge that this policy is not without flaws. However, as is often the case, governments must find the middle ground between protecting individual rights and safeguarding the public good. While anonymity does have legitimate benefits, at present these are far outweighed by the risks – particularly as millions of anonymous and bot-operated accounts can make entirely fabricated stories go viral, as our case studies demonstrate, with potentially serious political consequences.

Similar measures are already being introduced elsewhere. For instance, individuals purchasing pay-as-you-go SIM cards in the EU are now required to provide proof of identity.

Therefore, we recommend that this policy be introduced in stages, with careful review and evaluation at each step.

Title: “BREAKING THE ECHO CHAMBER: ENHANCING DISINFORMATION RESILIENCE IN THE UK”
By Dr Theo Zenou and Dr Helena Ivanov

The Henry Jackson Society
Millbank Tower, 21-24 Millbank
London SW1P 4QP, UK

www.henryjacksonsociety.org

HJS
Henry
Jackson
Society

DEMOCRACY | FREEDOM | HUMAN RIGHTS

CENTRE FOR RESILIENT SOCIETY

November 2025